Crystal Structure of Polygalacturonase from Erwinia carotovora ssp. carotovora*

(Received for publication, June 9, 1998, and in revised form, June 30, 1998)

Richard Pickersgill†, Drummond Smith, Kathryn Worboys, and John Jenkins
From the Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, United Kingdom

The crystal structure of the 40-kDa endo-polygalacturonase from Erwinia carotovora ssp. carotovora was solved by multiple isomorphous replacement and refined at 1.9 Å to a conventional crystallographic R-factor of 0.198 and Rfree of 0.239. This is the first structure of a polygalacturonase and comprises a 10 turn right-handed parallel β-helix domain with two loop regions forming a “tunnel like” substrate-binding cleft. Sequence conservation indicates that the active site of polygalacturonase is between these two loop regions, and comparison of the structure of polygalacturonase with that of rhamnogalacturonase A from Aspergillus aculeatus enables two conserved aspartates, presumed to be catalytic residues, to be identified. An adjacent histidine, in accord with biochemical results, is also seen. A similarity in overall electrostatic properties of the substrate-binding clefts of polygalacturonase and pectate lyase, which bind and cleave the same substrate, polygalacturonic acid, is also revealed.

Polygalacturonases are classified as family 28 glycosyl hydrolases (1); they hydrolyze the α-1,4 glycosidic bond in polygalacturonic acid or in the homogalacturonan regions of pectin (2–4). Plant polygalacturonases are important in fruit ripening and senescence, whereas the microbial enzymes are concerned with pathogen attack. Erwinia carotovora ssp. carotovora, which produces the endo-polygalacturonase PehA, is a major causal agent of soft rot in cultivated plants (5). The amino acid sequence of PehA has 20% identity with that of rhamnogalacturonase A from Aspergillus aculeatus (RGase A) whose structure has recently been solved (6). RGase A cleaves the α-1,2 glycosidic bond between alternating rhamnose and galactose in the rhamnogalacturonan regions of pectin. Glysidosidases either retain or invert the configuration of the anomeric carbon of the cleaved bond (7–9), and polygalacturonases and rhamnogalacturonases have been shown to be inverting enzymes (10, 11). The separation of the two acid residues involved in catalysis is on average 9.0–10.0 Å in inverting enzymes, a spacing argued to allow both substrate and a water molecule to fit between the general acid and the general base (12, 13). The architectures and mechanism of RGase A, PehA, and all family 28 glycosidases are expected to be conserved, whereas specificity has diverged. Comparison of the structures of PehA and RGase A is therefore expected to aid the identification of the catalytic amino acids and reveal the molecular details responsible for their differences in specificity. Several experiments suggest the presence of a histidine close to the active center of PehA (14, 15), but there is no histidine in the active site of RGase A. Phage P22 tailspike protein with endorhamnosidase activity may also be related to these enzymes (16), and active site residues have been proposed for tailspike protein based on structural studies (17, 18). There is also functional similarity between PehA and pectate lyase, which cleaves the same substrate but by a calcium-dependent β-elimination reaction. This functional similarity might be expected to result in similar size, shape, and surface charge of the substrate-binding clefts, differences might reflect their detailed specificity for degree or pattern of methylation or for branches in the pectin molecule. The structures of three pectate lyases (19–21) and two pectin lyases (22, 23) have been determined. No structure of polygalacturonase has been reported to date; however, three successful crystallizations have been reported (24–26). We now report the crystal structure of endo-polygalacturonase from E. carotovora ssp. carotovora refined at 1.9 Å resolution, suggest the identity of the catalytic amino acids, and reveal similarities in the binding clefts of polygalacturonase and pectate lyase.

MATERIALS AND METHODS

Enzyme Expression, Purification, and Crystallization—The endo-polygalacturonase gene from E. carotovora ssp. carotovora (PehA) was expressed in Bacillus subtilis (27). The 26-amino acid signal sequence is cleaved, and the mature polygalacturonase produced (PehA) comprised 376 amino acids (28). Purification and crystallization of PehA have been described in detail elsewhere (26). Single crystals were grown by the hanging drop method using 18% polyethylene glycol 6000, 0.2 M magnesium acetate, 0.1 M sodium cacodylate at pH 6.5 as the reservoir solution, and the protein drop comprised 4 μl of protein at 5 mg/ml plus 2 μl of reservoir. Crystals were harvested into mother liquor of the same composition as reservoir except that the polyethylene glycol 6000 concentration was increased to 20%. The crystals are monoclinic, space group C2, with a = 81.3 Å, b = 53.0 Å, c = 103.1 Å, and β = 112.6°. There is a single molecule in the asymmetric unit.

Data Collection—Because the PehA crystals were susceptible to radiation damage, all data were collected using crystals cooled to 100 K in the nitrogen stream of an Oxford Cryosystems Cryocooler. The cryoprotectant was mother liquor supplemented with 20% glycerol and 10% 2-propanol. Native data were collected using the EMBL X31 beam line at the DORIS storage ring, DESY (λ = 1.000 Å) equipped with a MAR300 image plate. Data from the chloroplatinum derivative prepared at pH 6.5 were collected using beam line DW21b at LURE (Orsay, France) (λ = 1.016 Å) equipped with a MAR300. Lutetium chloride data were also collected at LURE but using DW32 (λ = 0.970 Å) with MAR345. Data from the chloroplatinum derivative prepared at pH 7.5 and from the platinum orange (chloro2, 2′, 2′terpyridine)platinum(II)chloride) derivative were collected using a MacScience (Siemens) rotating anode generator with double mirrors and nickel filter (λ = 1.5418 Å) equipped with a rotating image plate.

* This work was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Tel.: 44-118-9357143; Fax: 44-118-9257917; E-mail: richard.pickersgill@bbsrc.ac.uk.

‡ This research was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1bhe) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.
Crystal Structure of Polygalacturonase

The various crystallographic parameters are defined as follows: \( R_{merge}(I) = \frac{\Sigma I_i - \langle I \rangle}{\Sigma I_i} \), where \( I_i \) is the intensity of the \( i \)-th observation, \( \langle I \rangle \) is the mean intensity of the reflection, and the summation extends over all data. \( R_{merge}(I) = \frac{\Sigma |I_{obs} - I_{calc}|}{\Sigma I_{obs}} \), the mean relative isomorphous difference between the native protein (\( F_{obs} \)) and the derivative (\( F_{calc} \)) data; \( R_{merge}(I) = \frac{\Sigma |F_{obs} - F_{calc}|}{\Sigma F_{obs}} \), where \( F_{obs} \) is the calculated heavy atom structure factor contribution; phasing power, \( \frac{F_{obs}}{\Sigma F_{calc}} \), where \( F_{calc} \) is the root mean square lack of closure; \( R \) is the root mean square deviation from ideal stereochemistry; phasing power, \( \frac{F_{obs}}{\Sigma F_{calc}} \), where \( F_{calc} \) is the root mean square lack of closure; \( R \) is the root mean square deviation from ideal stereochemistry; phasing power, \( \frac{F_{obs}}{\Sigma F_{calc}} \), where \( F_{calc} \) is the root mean square lack of closure.

| X-ray data collection | Native | Chloroplatinite, pH 6.5 | Chloroplatinite, pH 7.5 | Platinum orange | Lutetium chloride |
|-----------------------|--------|------------------------|------------------------|-----------------|------------------|
| X-ray source          | X31    | DW21b                  | MacScience             | MacScience      | DW32             |
| Detector              | Hamburg| LURE                   | in house               | in house        | LURE             |
| Wavelength (Å)        | 1.000  | 1.016                  | 1.5418                 | 1.5418          | 0.970            |
| Temperature (K)       | MAR300 | MAR300                 | DIP1030                | DIP1030         | MAR345           |
| Resolution (Å)        | 1.9    | 2.0                    | 1.9                    | 2.0             | 2.0              |
| Completeness (%)      | 96.8   | 99.7                   | 96.7                   | 98.1            | 98.2             |
| Highest shell         | 95.2   | 98.9                   | 81.8                   | 82.6            | 96.8             |
| \( R_{merge}(I) \) (%)| Overall| 3.8                    | 4.5                    | 7.7             | 4.0              |
|                       | Highest shell | 11.4               | 7.1                    | 26.8            | 14.4             |
|                      | \( <I/I_0(\bar{I})> \) | Overall | 21.4                  | 36.1            | 20.8             |
|                      |          | Highest shell         | 8.7                    | 25.1            | 10.1             |
|                      |          | \( R_{free} \) (%)    | 19.8                   | 14.4            | 22.7             |
| Overall phasing statistics | Resolution for phase calculation (Å) | 2.9 | 2.9 | 2.9 | 4.0 |
|                       | Number of sites | 2 | 1 | 1 | 2 |
| Phasing power         | Acentric | 1.70                   | 1.82                   | 0.94            | 0.93             |
|                       | Centric  | 1.28                   | 1.24                   | 0.78            | 0.72             |
|                       | \( R_{cal} \) (%) | Acentric | 0.69             | 0.66            | 0.87             |
|                       |          | Centric              | 0.71                   | 0.74            | 0.81             |
|                       |          | Mean figure of merit  | 0.67                   |                 |                  |
| Refinement and model stereochemistry | X-ray data used in refinement | Resolution limits (Å) | 12.5–1.9 | 29430 |
|                       | Number of reflections used in refinement | 1567 |
|                       | Number of reflections used for \( R_{free} \) calculation | 2813 |
|                       | Number of protein atoms | 295 |
|                       | Number of solvent atoms | 19.8 |
|                       | \( R \)-factor (%) | 23.9 |
|                       | \( R_{free} \) (%) | 0.005 (0.020) |
| Root mean square deviation from ideal stereochemistry, target \( \sigma \) in parentheses | 1–1 distances (Å) | 0.021 (0.040) |
|                       | 1–2 distances (Å) | 0.019 (0.050) |
|                       | 1–4 distances (Å) | 0.071 (0.150) |
|                       | Chiral volumes (Å³) | 0.017 (0.030) |
|                       | Peptide planes (Å) | 0.005 (0.020) |
|                       | Aromatic planes (Å) | 16.4 |
| Analysis according to PROCHECK (38) | Ramachandran plot | 86.7 |
|                       | Residues in the most favoured regions (%) | 13.0 |
|                       | Residues in additionally allowed regions (%) | 0.0 |
|                       | Non-glycine residues in disallowed regions (%) | +0.10 |

with DIP1030 image plate. The five data sets were reduced using DENZO and SCALEPACK\(^1\) and the CCP4 program suite (29).

Structure Solution—Three platinum derivatives were prepared by soaking native crystals under the following conditions: 25 mM chloroplatinite for 16 h at pH 6.5, 25 mM chloroplatinite for 1.5 h at pH 7.5, using 0.1 M HEPES buffer in place of cacodylate, and 3 mM platinum orange for 65 h at pH 6.5. The fourth derivative was prepared by soaking a native crystal in 20 mM lutetium chloride for 9 h at pH 6.5.

Model Building and Refinement—Model building using \("O" (32) running on a Silicon Graphics Indigo\(^2\) Extreme was facilitated by the availability of the coordinates of RGase A. A complete model of the protein was built using iterative rounds of crystallographic refinement with XPLOR (33, 34) and ARP (35) with intervening manual revisions of the structure. 5% of the diffraction data were used for cross-validation purposes (36). ARP was not always successful in reducing the \( R \)-factor and \( R_{free} \) values but nevertheless gave the map with clearest indication of how to rebuild. The final refinement used the XPLOR solvent correction terms input to REFMAC (37). Atomic positions and isotropic B-factors were refined. The model stereochemistry was evaluated using the programs PROCHECK (38) and WHATCHECK (39).

\(^1\) Z. Otwinowski, unpublished programs.
RESULTS AND DISCUSSION

The results of the data collection, phasing, refinement, and the stereochemistry of the final model are presented in Table I. The final model comprises all 376 amino acids of mature PehA plus 295 water molecules. The stereochemistry as judged by PROCHECK is good, with an overall G-factor of +0.1, compared with a typical value of −0.3 for refinement at this resolution. Additionally the water molecules have good hydrogen bonding patterns. The overall fold of PehA is related to that of RGase A but with different loop structures and extensions (Fig. 1, A and B). However, the structure of PehA could not be solved by molecular replacement using RGase A as the search model. The structure comprises 10 complete coils of right-handed parallel \( \beta \)-helix architecture plus an \( \alpha \)-helix in an additional coil at the N-terminal end and an incomplete coil at the C-terminal end of the protein. Surprisingly, there is no polypeptide covering the C-terminal end of the parallel \( \beta \)-helix and no C-terminal extension. The last strand of parallel \( \beta \)-strand two (PB2) is important in stabilizing the PehA structure, which is necessary for secretion via the Out pathway (40). The B-factors, a measure of flexibility, are lowest for the central turns and increase toward the N-terminal and C-terminal ends of the parallel \( \beta \)-helix, the increase is most dramatic toward the uncapped C-terminal end. Superimposed on this general trend in B-factors are spikes correlating with the long loops protruding from the parallel \( \beta \)-helix domain (Fig. 2). In contrast to the PehA structure, RGase A has an additional \( \alpha \)-helix in the N-terminal extension and additional polypeptide chain at the C-terminal end of the parallel \( \beta \)-helix forming an antiparallel \( \beta \)-strand at the C-terminal end of parallel \( \beta \)-sheet 1 (PB1) followed by an extension that partially covers parallel \( \beta \)-sheet 3 (PB3). The loop structures are also different. Although the pectate and pectin lyases have three parallel \( \beta \)-sheets (Fig. 1C), PB1, PB2, and PB3, PehA and RGase A have an additional parallel \( \beta \)-sheet, PB1a (18). PehA has two disulfides, one be-

---

**Fig. 1.** Schematic representation of the structures of *E. carotovora* ssp. *carotovora* polygalacturonase (PehA) (A), *A. aculeatus* rhamnogalacturonase A (RGase A) (B), and *B. subtilis* pectate lyase (BsPel) (C). The parallel \( \beta \)-sheets PB1, PB1a, PB2, and PB3 are yellow, blue, green, and red, respectively. PehA and RGase A have 10 complete turns of right-handed parallel \( \beta \)-helix architecture. BsPel has seven turns and lacks PB1a. D, the active site of PehA is on the surface of parallel \( \beta \)-sheet one, on \( \beta \)-strands four, five, six, seven, and eight, and on the loops that precede these \( \beta \)-strands. Aspartates 202 and 223 are conserved in all polygalacturonases and in RGase A (see also Fig. 4). This figure was prepared using MOLSCRIPT (41).
between Cys15 just prior to the first β-strand of PB2 and Cys35 after the C-terminal end of the first α-helix. The second disulfide, Cys69-Cys196, is in the long loop preceding strand two of PB1. There are two cis-peptides, Pro116 and Ser257. The first is a perturbation of the usually regular αi-conformation turn between PB2 and PB3, the second is in the loop preceding β-strand seven of PB1.

Alignment of 36 polygalacturonase sequences reveals four conserved regions: Asn201-Thr202-Asp203, Gly222-Asp223-Asp224, Gly250-His251-Gly252, and Arg3-Ile3-Lys3. These sequences cluster on or before β-strands 5, 6, 7, and 8 of PB1. Asn-Thr-Asp forms strand 5, and Arg-Ile-Lys forms strand 8 of PB1. Gly-Asp-Asp forms the turn before β-strand 6, and Gly-His-Gly forms the turn before β-strand 7 of PB1. This clustering is a clear indication of functional conservation and is strongly suggestive that the region on the surface of β-strands 5–8 of PB1 and the adjacent loops form the catalytic site (Fig. 1D). The location of this site in a pronounced cleft further strengthens this proposal.

The pronounced substrate-binding cleft is formed by two long loops that precede strands two and three of PB1 and three loops that follow strands seven, eight, and nine. Ser257, in cis-conformation in the loop preceding the seventh β-strand of PB1 may be important in maintaining the geometry of the substrate-binding cleft. The cleft is at an angle to the axis of the parallel β-helix as a result of the long loops toward the N-terminal end preceding PB1 and long loops toward the C-terminal end of the parallel β-helix occurring after PB1. This cleft is more “tunnel-like” than that of the pectate lyases, pectin lyases, or RGase A and is characterized by the presence of conserved aspartates and lysines at its center. Arginines and lysines also line the sides of the substrate-binding cleft, making the overall electrostatic potential in the substrate-binding cleft positive (Fig. 3).

In this respect PehA is similar to B. subtilis pectate lyase with calcium bound and different from Aspergillus niger pectin lyase and RGase A.

RGase A may be superimposed on PehA with a root mean square displacement of 1.7 Å for 280 equivalent α-carbon atoms. Comparison of the superimposed PehA and RGase A structures reveals that in the active site region a number of aspartates and lysines are conserved (Fig. 4). Additionally, Phe175 (151) in the loop before β-strand 4 and cis-Ser257 (226) in the loop following β-strand 7 of PB1 are conserved in both structures. The equivalent amino acid number in RGase A is given in parentheses. The conserved aspartates between PehA and RGase A in this region are clear candidates for the base and proton donor in the inverting mechanism: Asp202 (177), Asp205 (180), Asp223 (197), Asp224 (Glu198), and Asp311 (281). One of these water molecules may be the “nucleophilic water.”
second conserved water molecule forms hydrogen bonds to Asp202, (Asp/Glu)224 (198) and Lys229 (203). Conserved lysines in the substrate-binding cleft are at positions 229 (203) and 282 (253). There are no conserved arginines and there is no equivalent to His251 on the loop before strand 7 of PB1 of PehA in Rgase A. In PehA, Phe175, Asp202, and Asp224 are at equivalent positions on adjacent turns. Asp224 and His251 are also at equivalent positions. Phe175 stacks against Asp202, and Asp223 stacks against His251 (Fig. 4).

Three carboxylates have been proposed to be involved in the activity of P22 tailspike protein (17, 18); Asp395 and Glu359 are proposed to be general bases, and Asp392 is proposed to be the general acid. Steinbacker et al. also argue that tailspike is an inverting glycosidase. It has previously been argued that to accommodate a water molecule and substrate between the catalytic carboxylates the distance between the carboxylates must be 9.0–10.0 Å (12, 13). However, the proposed general bases Glu359 and Asp395 are 8.2 and 5.8 Å, respectively, from the proposed general acid, Asp392. Protonation must occur in the plane defined by the nucleophilic water, the anomic carbon, and the glycosidic oxygen. However, it may be possible for protonation of the glycosidic oxygen and nucleophilic attack at the anomic carbon to be from the same side of the bond in α-linked polysaccharides rather than opposite sides with a resulting shorter separation of carboxylates. If so, this would help to explain the short spacing between the conserved carboxylates in PehA and Rgase A.

The α-carbon positions of Asp392 (proposed general acid) and Asp202 of tailspike protein can be superimposed on Asp202 and Asp205 of PehA and the overall structural superposition is then 2.0 Å for 178 equivalent α-carbons. The side chain of Asp202 superimposes rather well but not that of 205, which is in any case not conserved in all polygalacturonases. It is clear that the active site geometry and distribution of charge in tailspike protein is different from that of PehA and Rgase A.

The structural results therefore suggest that aspartates 202 and 223 are directly involved in the mechanism of PehA and that the carboxylate at position 224 is probably also involved. One of the conserved two water molecules bound to these carboxylates may be the nucleophilic water. A more detailed assignment of the role of these carboxylates and water molecules awaits determination of the structure of a nonproductive enzyme substrate complex.

Acknowledgments—The B. subtilis strain expressing E. carotovora spp. carotovora polygalacturonase was kindly donated by Dr. Ilka Palva of the Food Research Institute (Jokioinen, Finland). We thank Thomas Petersen and Sine Larsen for making the Rgase A coordinates available prior to release in the protein data bank. We also acknowledge use of the EMBL X31 beam line at the DORIS storage ring (DESY, Hamburg, Germany) and the use of beamlines DW21b and DW32 at LURE (Orsay, France). We also thank EU AIR2 project CT94-1345 for support.