# Active-Site Architecture of Endopolygalacturonase I from *Stereum purpureum* Revealed by Crystal Structures in Native and Ligand-Bound Forms at Atomic Resolution[†,‡]

Tetsuya Shimizu,[§,||] Toru Nakatsu,[§] Kazuo Miyairi,[||] Toshikatsu Okuno,[||] and Hiroaki Kato*[,§]

*Kinetic Crystallography Research Team, Membrane Dynamics Research Group, RIKEN, Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148 Japan, and Department of Biochemistry and Biotechnology, Faculty of Agriculture and Life Sciences, Hirosaki University, 3 Bunkyo-cho, Hirosaki, Aomori 036-8561, Japan*

*Received January 11, 2002; Revised Manuscript Received April 9, 2002*

ABSTRACT: Crystal structures of endopolygalacturonase from *Stereum purpureum* were solved in native and two galacturonic acid complex states at atomic resolution. Endopolygalacturonase catalyzes the hydrolysis of α-1,4-glycosidic linkage of polygalacturonate in pectin. The native structure was determined by the multiple wavelength anomalous dispersion method and was refined anisotropically with SHELXL-97, with an *R* factor of 11.4% and an $R_{free}$ factor of 14.0% at 0.96 Å resolution. The enzyme folds into a right-handed parallel β-helix with 10 complete turns. The crystal structures of its binary complex with one D-galacturonate and its ternary complex with two D-galacturonates were also determined to identify the substrate binding site at 1.0 and 1.15 Å resolutions, respectively. In the binary complex, one β-D-galactopyranuronate was found in the +1 subsite, thus proving the strong affinity of the +1 subsite expected from the bond cleavage frequency on oligogalacturonates. In the ternary complex, an additional β-D-galactofuranuronate was found in the −1 subsite. In both subsites, the recognition of the galacturonate carboxy group is important in galacturonate binding. In the +1 subsite, the carboxy group interacts with three basic residues, His195, Arg226, and Lys228, which were conserved in all endopolygalacturonases. In the −1 subsite, the unique nonprolyl *cis*-peptide bond is believed to be involved in binding the carboxy group of the substrate. The active site architecture of the complexes provides insight into the mechanism of inverting glycosyl hydrolases and also sheds light on the basis of the differences between the family 28 and the other inverting glycosyl hydrolases.

Pectin is a major component of the plant cell wall, and its degradation is an important process in infection by plant pathogenic organisms, plant senescence, and fruit ripening (*1*). It consists of so-called "smooth" and "hairy" regions. The smooth region, or the homogalacturonan component, consists of a backbone of α-1,4-linked D-galacturonic acid residues, which are partly methylated and acetylated. The hairy region, known as rhamnogalacturonan, is characterized by stretches of 1,2-α-L-rhamnose-1,4-α-D-galacturonic acid dimers. L-Arabinose and D-galactose may be attached to the rhamnose residues.

Endopolygalacturonases (EC 3.2.1.15) are involved in pectin degradation. The enzymes belong to the family 28 glycosyl hydrolases and catalyze the hydrolysis of the α-1,4-glycosidic bonds between adjacent α-D-galacturonic acid residues within the pectin main chain (*2*). The endopolygalacturonases are inverting glycosidases that invert the anomeric configuration of the products during the reaction (*3*). In this mechanism, the hydrolysis proceeds by a general acid catalyst donating a proton to the glycosidic oxygen and a catalytic base guiding the nucleophilic attack of a water molecule on the anomeric carbon of the galacturonate moiety bound at the −1 subsite. Studies of the bond cleavage frequency with oligo-galacturonates have indicated that many endopolygalacturonases to hydrolyze the first glycosidic bond from the reducing end of tri- and tetragalacturonate (*4*−*6*).

*Stereum purpureum* is a pathogenic fungus that causes silver-leaf disease in apple trees, and it secretes large amounts of several endopolygalacturonases into the culture medium (*7*, *8*). These endopolygalacturonases degrade the pectin in leaves and cause the silver leaf symptoms. Endopolygalacturonase I (endoPG I)[1], a glycoprotein with 335 amino acid residues, is one of these endopolygalacturonases (*9*). EndoPG I consists of three glycoforms, which differ in the number of glycosylation sites. The glycoforms of endoPG I can be separated by cation exchange column chromatography *(4)*, and their glycosylation sites and sugar chain structures have been well defined in a previous study (*10*). We obtained

[†] This work was supported, in part, by the Ministry of Education, Culture, Sports, Science, and Technology (13460046 and 13033044). T.S. is a Junior Research Associate Fellow of RIKEN.

[‡] Coordinates and observed structure factor amplitudes for the structure described in this paper have been deposited with the Protein Data Bank (accession codes 1K5C, 1KCC, and 1KCD).

* Corresponding author. Telephone: +81-791-58-1825. FAX: +81-791-58-1826. E-mail: katohiro@spring8.or.jp.

[§] RIKEN.

[||] Hirosaki University.

[1] Abbreviations: endoPG, endopolygalacturonase from *Stereum purpureum*; MAD, multiple wavelength anomalous dispersion; Gal*p*A, galactopyranuronic acid; Gal*f*A, galactofuranuronic acid; $F_o$ and $F_c$, observed and calculated X-ray structure factor amplitudes; *R*-factor, X-ray residual factor.

10.1021/bi025541a CCC: $22.00 
Published on Web 05/04/2002

single crystals for our atomic resolution (0.96 Å) study by using deglycosylated endoPG Ia, so as to avoid the heterogeneity of the sugar chains (*11*).

Recently, the crystal structures of the endopolygalacturonases from the bacterium *Erwinia carotovora* (*12*) and the fungi *Aspergillus niger* (*13*) and *A. aculeatus* (*14*) were solved at 1.9, 1.68, and 2.0 Å resolutions, respectively. These structures fold into a right-handed parallel $\beta$-helix, which is also found in other pectin degrading enzymes, such as rhamnogalacturonase A (*15*) and pectate lyase (*16*). Three conserved aspartate residues were proposed to be critical for the catalysis (*17*) in those structures. These residues were implicated as the acid−base catalysts in the reaction of the endopolygalacturonase. The distances between the pairs of the three aspartate residues are 4.1 ∼ 6.5 Å. In contrast, in the common inverting enzymes, the average distance between the catalytic residues is 10 Å (*18*). This finding indicated that the active site architecture of the endopolygalacturonases could differ from that of most other inverting glycosyl hydrolases. Site-directed mutagenesis experiments with the *A. niger* enzyme revealed that Asp153, Asp173, and Asp174 in endoPG I are involved in its catalytic function, whereas His195, Arg226, Lys228, and Tyr262 function primarily in substrate recognition (*19*, *20*). However, there is little structural evidence regarding the substrate and/or inhibitor binding site or the architecture of the active site.

In this paper, we describe the crystal structures of the native and product complexes of endoPG I from *S. purpureum* at an atomic resolution. In these complexes, one molecule of galacturonate was observed in the +1 subsite, and two molecules of galacturonate were observed in the +1 and −1 subsites, respectively. These structures provide experimental evidence of the substrate binding mechanism and insight into the active site architecture as well as the reaction mechanism of endoPG I.

## EXPERIMENTAL PROCEDURES

*Purification, Crystallization, and Preparation of the Pt Derivative*. EndoPG I was purified from a culture filtrate of *Stereum purpureum* and was crystallized as previously described (*11*). The enzyme has three isoforms with respect to differences in the sugar chain. EndoPG Ia, in which Asn92 and Asn161 are glycosylated with the high mannose type of sugar chain was used in this study after deglycosylation with endo-$\beta$-*N*-acetylglucosaminidase H (*10*). For the phase determination, a $K_2PtCl_4$ derivative was prepared by soaking the native crystals in a heavy atom solution containing 2.4 mM $K_2PtCl_4$, 25% (w/v) PEG 4000, and 50 mM MES buffer (pH 6.2) for a week.

*Data Collection*. The native crystal was flash-frozen in liquid nitrogen after soaking in a cryoprotectant solution containing 20% (w/v) PEG 4000, 20% (v/v) glycerol, 10% (v/v) 2-propanol, and 200 mM NaCl in 50 mM sodium acetate buffer (pH 5.0). The native crystal data were collected on beam line BL44B2 at SPring-8, using a marCCD165 detector (*11*). The crystal diffracted to 0.96 Å resolution using a 0.7 Å wavelength X-ray beam. In the case of the Pt derivative, the Pt-soaked crystal was back-soaked and flash-frozen in liquid nitrogen using the cryoprotectant solution at pH 6.2. The X-ray fluorescence spectrum was measured directly from the crystal on the BL44B2 beam line, using a Si diode X-ray fluorescence detector. Selection of the wavelengths of the peak and the edge for the multiple anomalous dispersion (MAD) data collection was made on the basis of the $f'$ and $f''$ anomalous scattering factors, determined from the fluorescence spectra using the program CHOOCH (*21*). This gave the peak at 1.07178 Å and the edge at 1.07216 Å. The remote was chosen at 1.03 Å to obtain the dispersive differences. Each wavelength data set was collected from 0° to 360° using 1° oscillation and a 5 s exposure time. The three data sets were collected in the order of peak, edge, and remote. They were collected at each wavelength to a resolution of at least 2.19 Å. The native and derivative data sets were integrated with MOSFLM (*22*) and were scaled with SCALA (*23*).

*MAD Phasing, Automatic Model Building, and Refinement*. A resolution range between 20 and 2.2 Å of the MAD data set was used for the phase determination. Four Pt sites were found, refined, and used for the phase calculation in the program SOLVE (*24*). The initial phase was modified by solvent flattening and histogram matching using DM (*25*). The initial model was constructed automatically by ARP/wARP (*26*). The native data set and the phase output from DM were combined, and 5% of the reflections were selected for cross-validation. The phase was further modified and extended to 0.96 Å with the mode warp of ARP/wARP. Then, the mode warpNtrace of ARP/wARP automatically built a model that had 321 of the 335 amino acid residues of endoPG I. The remaining parts of the structure were built manually using the program O (*27*), and the resulting model was refined by Refmac (*28*). Water molecules were added with the mode solvent of ARP/wARP. SHELXL-97 (*29*) was used for further refinement of the model, using the anisotropic temperature factor with stereochemical restraints. In this step, some multiple conformation structural parts were added to the model.

*Preparation of Galacturonate Complexes*. The binary complex of the enzyme with monogalacturonic acid was obtained by cocrystallization of the nondeglycosylated endoPG Ia with trigalacturonic acid. In the cocrystallization, 2 $\mu$L of a 2% (w/v) (37 mM) trigalacturonate solution were added to the hanging-drops, and the other conditions were the same as those used for the native crystallization (*11*). The crystals thus obtained were soaked in cryoprotectant solution containing 2% (w/v) trigalacturonate. The ternary complex, consisting of the enzyme and two molecules of monogalacturonic acid, was obtained by soaking the native crystals in a different cryoprotectant solution containing galacturonate [10% (w/v) (515 mM) galacturonate, 10% (w/v) PEG 10000, 20% (v/v) glycerol, and 200 mM NaCl in 50 mM sodium acetate buffer, pH 5.0].

The diffraction data of the complexes were also collected on beam line BL44B2 at SPring-8, to 1.0 and 1.15 Å resolutions for the binary and the ternary complexes, respectively, at 90 K. Both crystals were isomorphous with the native one. The data were processed and reduced in the same manner as for the native crystal. The initial models were determined by the rigid body and restrained refinements with Refmac from the native structure, and were rebuilt by ARP/wARP. Further refinements were performed as that used for the same manner in the native structure. The stereochemical dictionary for the refinement of the galacturonate moieties was derived from the structure of galactose in the

Table 1: Data Collection and Refinement Statistics

| | Pt derivative | | | | | |
|---|---|---|---|---|---|---|
| | remote | peak | edge | native | binary complex | ternary complex |
| **Data Collection** | | | | | | |
| X-ray source | | | | SPring-8 BL44B2 | | |
| wavelength (Å) | 1.03000 | 1.07178 | 1.07216 | 0.70000 | 0.70000 | 0.60000 |
| resolution (Å) | | | | | | |
| overall | 31.2−2.10 | 32.4−2.19 | 32.4−2.19 | 41.5−0.96 | 25.3−1.00 | 12.1−1.15 |
| outer shell | 2.21−2.10 | 2.31−2.19 | 2.30−2.19 | 1.01−0.96 | 1.05−1.00 | 1.21−1.15 |
| reflections | | | | | | |
| observed | 65,489 | 57,904 | 58,361 | 1,186,116 | 649,682 | 230,114 |
| unique | 16,941 | 15,065 | 15,078 | 171,534 | 152,404 | 104,270 |
| redundancy | 3.9 | 3.8 | 3.9 | 6.9 | 4.3 | 2.2 |
| completeness (%) | | | | | | |
| overall | 95.8 | 95.2 | 95.2 | 94.0 | 93.8 | 97.3 |
| outer shell | 91.4 | 88.8 | 91.6 | 89.4 | 89.9 | 96.2 |
| mean $I/\sigma I$ | | | | | | |
| overall | 25.6 | 21.6 | 23.7 | 5.6 | 6.1 | 6.5 |
| outer shell | 14.2 | 12.3 | 13.7 | 3.5 | 3.8 | 2.8 |
| $R_{sym}$ (%)[a] | | | | | | |
| overall | 1.8 | 2.1 | 2.0 | 3.7 | 5.4 | 7.1 |
| outer shell | 4.0 | 4.7 | 4.3 | 20.6 | 17.2 | 25.1 |
| **MAD Phasing** | | | | | | |
| resolution (Å) | 20.0−2.20 | | | | | |
| $R_\lambda$ (%)[b] | | | | | | |
| to remote | | 1.7 | 1.9 | | | |
| to peak | | | 0.9 | | | |
| $R_{ano}$ (%)[c] | 3.8 | 3.9 | 3.9 | | | |
| no. of site | | | 4 | | | |
| mean figure of merit | | | 0.72 | | | |
| **Structure Refinement and Stereochemistry** | | | | | | |
| resolution (Å) | | | | 10.0−0.96 | 10.0−1.00 | 10.0−1.15 |
| no. of unique reflections | | | | 171,380 | 152,299 | 104,207 |
| no. of reflections in test set | | | | 8,572 | 7,627 | 5,171 |
| $R_{work}$ (%)[d] | | | | 11.4 | 10.9 | 12.6 |
| $R_{free}$ (%)[c] | | | | 14.0 | 13.4 | 16.4 |
| no. of atoms | | | | | | |
| protein | | | | 2,466 | 2,502 | 2,437 |
| sugar chain | | | | 42 | 28 | 28 |
| ligand | | | | 0 | 13 | 26 |
| glycerol | | | | 12 | 0 | 0 |
| $Cl^-$ | | | | 1 | 1 | 1 |
| $Na^+$ | | | | 0 | 1 | 0 |
| solvent | | | | 504 | 585 | 510 |
| root-mean-square deviations | | | | | | |
| bond distances (Å) | | | | 0.018 | 0.016 | 0.014 |
| angle distances (Å)[f] | | | | 0.032 | 0.031 | 0.030 |
| chiral volumes (Å$^3$) | | | | 0.117 | 0.101 | 0.090 |

[a] $R_{sym} = \Sigma |I - I_{av}|/\Sigma$ I. [b] $R_\lambda = \Sigma|I_{\lambda i} - I_{\lambda j}|/\Sigma|I_{av\lambda}|$. [c] $R_{ano} = \Sigma |I(+) - I(-)|/\Sigma|I_{av}|$. [d] $R_{work} = \Sigma |F_{obs} - F_{calc}|/\Sigma|F_{obs}|$ for the reflections of work set. [e] $R_{free} = \Sigma |F_{obs} - F_{calc}|/\Sigma|F_{obs}|$ for the reflections of test set. [f] Distances between two atoms that are bonded to same atom.

HIC-up database (*30*). The enzyme−ligand interactions were analyzed by HBPLUS (*31*) and LIGPLOT (*32*).

## RESULTS

*Structure Determination of Endopolygalacturonase I.* Details of the structural determination and the crystallographic refinement are given in Table 1. The structure of the native crystal was solved by the MAD method with a Pt derivative and was refined at 0.96 Å resolution. The program ARP/wARP constructed a structural model containing 321 amino acid residues automatically, and provided a high quality electron density map. The final structural model consisted of amino acid residues 3−335, two *N*-acetylglucosamine residues attached to Asn92 and Asn161, and 504 water molecules. The $R$ and $R_{free}$ factors were 11.4 and 14.0%, respectively. The Ramachandran plot for 277 nonglycine and nonproline residues showed 238 (85.9%) residues to be in the most favored region, 36 (13.0%) residues in the generously allowed region, 2 (0.7%) residues in the additional allowed region, and 1 (0.4%) residue in the disallowed region.

Two kinds of complex structures containing monogalacturonic acids bound at the proposed active site cleft were also determined and refined. The binary complex containing a single molecule of monogalacturonic acid as the ligand was obtained by cocrystallization of the enzyme with trigalacturonic acid. The final refined model had $R$ and $R_{free}$ factors of 10.9 and 13.4%, respectively, at 1.0 Å resolution. The ternary complex, prepared by soaking contained two molecules of monogalacturonic acid, had pyranose and furanose ring configurations, respectively. The ternary complex structure was refined at 1.15 Å resolution with an $R$ factor of 12.6% and a corresponding $R_{free}$ factor of 16.4%. The three structures met the criteria of atomic resolution defined by Sheldrick (*33*).

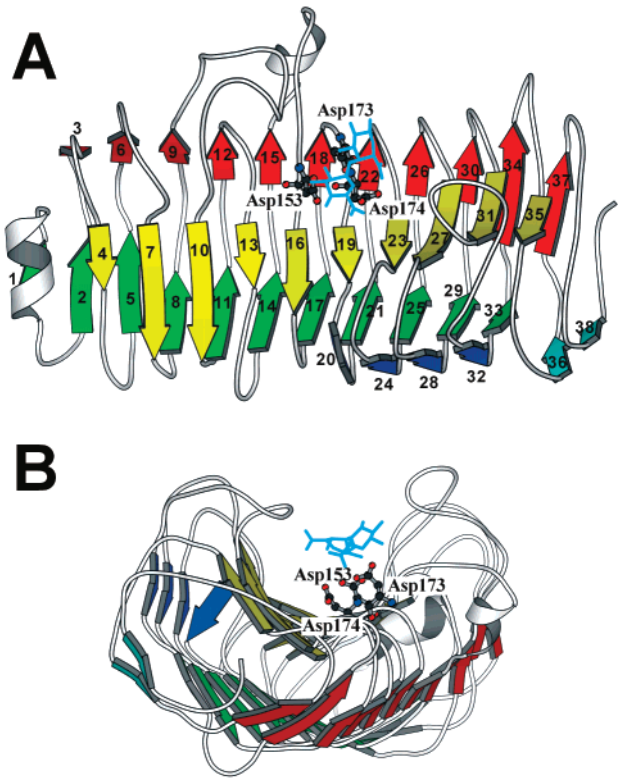


FIGURE 1: (A) Schematic representation of the structure of *S. purpureum* endoPG I with the N-terminus on the left and the C terminus on the right, viewed on $\beta$-sheet PB1 (yellow). PB2a, PB2b, PB2c, and PB3 are shown in blue, green, cyan, and red, respectively. The three aspartate residues, Asp153, Asp173, and Asp174, are shown in ball-and-stick models. The putative binding state of the digalacturonate part of the substrate is shown in a stick model (sky blue), and is discussed in a later section. (B) The structure viewed from the C-terminal side. A large cleft is formed by the parallel $\beta$-sheet, PB1, and the loops of both sides of PB1.

*Overall Structure*. EndoPG I has a right-handed $\beta$-helical structure, which is common to the pectinases, such as pectate lyases (*16*), rhamnogalacturonase (*15*), and endopolygalacturonases from *E. carotovora* (*12*), *A. niger* (*13*), and *A. aculeatus* (*14*) (Figure 1). The $\beta$-helical structure consists of five $\beta$-sheets, PB1, PB2b, PB2a, PB2c, and PB3 and comprises 10 complete turns from $\beta$-strand 4 in PB1 to 37 in PB3 (Figure 1A). The amino acid sequence of endoPG I has 25.8 and 40.6% identity with those of the *E. carotovora* endopolygalacturonase, PehA, and the *A. niger* endopolygalacturonase II, respectively (Figure 2). Structural superpositions of endoPG I with PehA and endopolygalacturonase II showed that 252 and 301 Cα atoms are in structurally equivalent positions, with rms differences of 1.6 and 1.4 Å, respectively. A comparison of endoPG I with *A. niger* endopolygalacturonase II revealed that most of the secondary structural elements are in the same positions. Differences in the two structures are seen only in their C-terminal regions. The C-terminus of the *A. niger* endopolygalacturonase II is fixed by a disulfide bridge, whereas that of endoPG I is fixed by an additional $\beta$-sheet structure, PB2c, composed of two $\beta$-strands, 36 and 38.

A large cleft is formed along the surface of the $\beta$-sheet PB1 and is surrounded by the three loops connecting $\beta$-strands 9 and 10, 12 and 13, and 15 and 16, as well as by the three loops on the other side, connecting $\beta$-strand 23 and 24, 27 and 28, and 31 and 32 (Figure 1). The proposed catalytic residues, Asp153, Asp173, and Asp174, are on the bottom of the cleft (Figure 1B). Eight residues that are completely conserved in all endo- and exopolygalacturonases (Asn151, Asp153, Asp173, Asp174, His195, Gly196, Arg226, and Lys228 for endoPG I) (*17*) have equivalent positions in the cleft among the four structures that have been solved. In addition, Asn91 and Tyr262 are also conserved in the four structures, and His150 is conserved in the structures of the fungal endopolygalacturonases.

*Galacturonic Acid Complex Structures*. The experimental evidence regarding the substrate binding site and the active site constitution was obtained by the analysis of two kinds of monogalacturonic acid complex structures. The first complex, the binary complex, was prepared by cocrystallization with 2% trigalacturonate; however, an unambiguous electron density was observed only for a single galacturonic acid at the large cleft (Figure 3A). The structure of the bound galacturonic acid was identified as $\beta$-D-galactopyranuronic acid (Gal*p*A) with the conventional $^4C_1$ chair form confor-

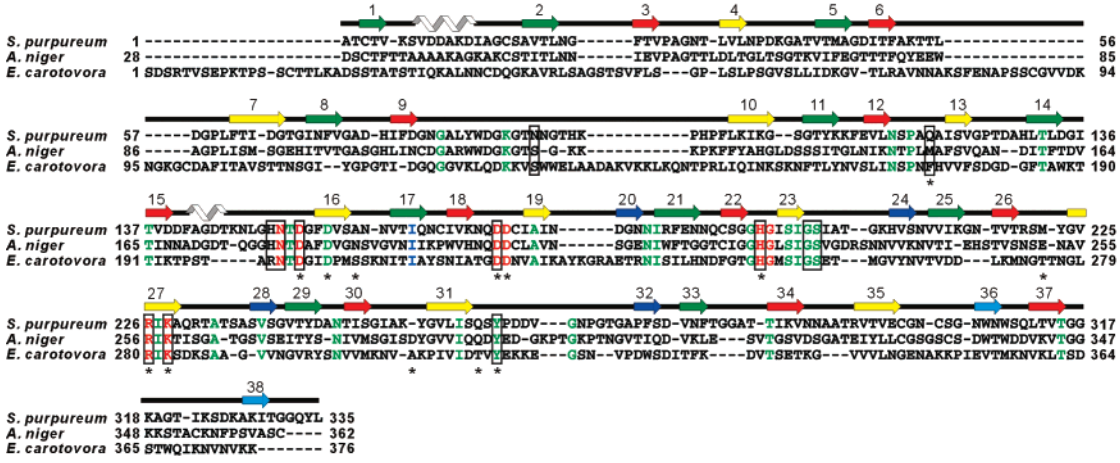


FIGURE 2: An amino acid sequence alignment of the three endopolygalacturonases, endoPG I of *S. purpureum*, endopolygalacturonase II of *A niger*, and PehA of *E. carotovora*, based on three-dimensional structure. Spirals, arrows, and lines above the amino acid sequences show α-helices, $\beta$-strands, and loops of the *S. purpureum* endoPG I, respectively. The colors of the $\beta$-sheets are the same as in Figure 1. The eight residues in red are invariant residues in the available endo- and exopolygalacturonases, as indicated by Kester (*17*). The residues in green are common residues in the three endopolygalacturonases in this alignment, except for the residues in red. The boxed residues are the residues involved in the binding of galacturonate molecules in the complex structures.

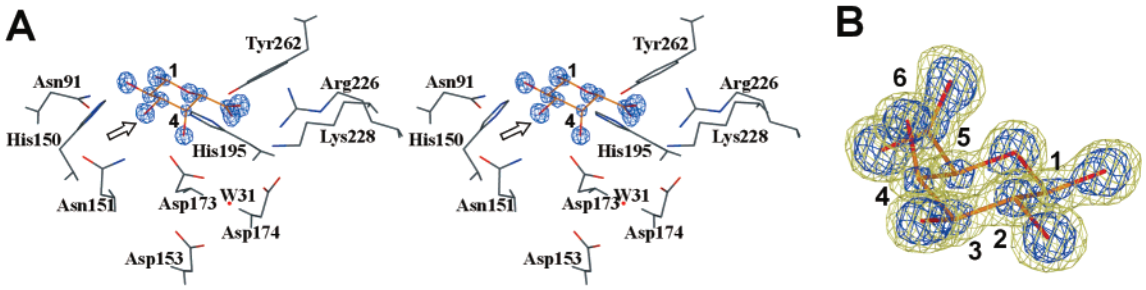


FIGURE 3: (A) Stereoview of the binding site of Gal*p*A in the binary complex. The Gal*p*A is shown as a stick model, and its carbon atoms are colored yellow. The omit $F_o - F_c$ electron density map for the bound Gal*p*A is contoured at 9.0 $\sigma$. Residues interacting with Gal*p*A and putative catalytic residues are included. A water molecule, W31 (red dot), that bound to Asp153 and Asp174 is considered as a nucleophilic water. The viewpoint of B is indicated by an arrow. (B) A close-up view of the bound Gal*p*A with the $F_o - F_c$ electron density map. The map is contoured at 3.0 (yellow) and 9.0 $\sigma$ (blue).

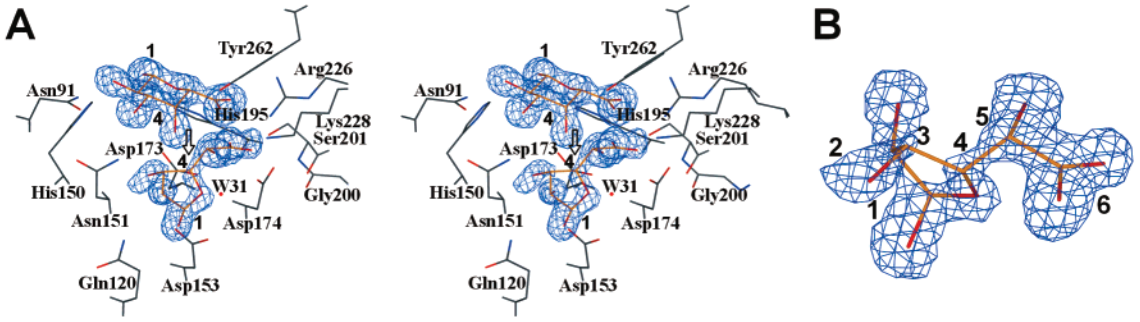


FIGURE 4: (A) Stereoview of the binding sites of Gal*f*A and Gal*p*A in the ternary complex. The Gal*f*A and Gal*p*A are shown as stick model, and their carbon atoms are colored yellow. The omit $F_o - F_c$ electron density map for the bound Gal*f*A and Gal*p*A is contoured at 3.0 $\sigma$. Residues interacting with the ligands and putative catalytic residues are included. A water molecule, W31 (red dot), that bound to Asp153 and Asp174 is considered to be a nucleophilic water. The viewpoint of B is indicated by an arrow. (B) A close up view of the bound Gal*f*A with the $F_o - F_c$ electron density map. The map is contoured at 3.0 $\sigma$.

mation. Although its C1 carbon was in the $\beta$-anomeric configuration, the C1 of the Gal*p*A unit was linked to C4 of the adjacent Gal*p*A (reducing end side) unit through a glycosidic bond with in the substrate, polygalacturonate, in the α-anomeric configuration. The same structure of the bound Gal*p*A was observed in the soaking experiment into a 94 mM monogalacturonic acid solution. Thus, the Gal*p*A molecule observed in the cocrystallization with trigalacturonate seems to be a hydrolytic product by the enzymatic reaction. Indeed, the considerable amount of monogalacturonic acid was detected in the mother liquor of the cocrystallization batch by use of an ion spray mass spectroscopy (data not shown). The second complex, the ternary complex, was prepared by soaking the native crystals in a solution containing a high concentration (515 mM) of D-galacturonic acid. In this complex, two lobes of electron density were found for the ligand (Figure 4A). One of the two lobes was located at the identical position of the Gal*p*A site in the binary complex (Figure 4A). The other lobe of the density was identified as $\beta$-D-galactofuranuronic acid (Gal*f*A) (Figure 4B). The electron density of Gal*f*A was weaker than that of Gal*p*A, because of the low occupancy of the bound Gal*f*A. The occupancy of Gal*f*A was estimated to be about 60% that of Gal*p*A by the SHELXL-97 occupancy refinement. The conformation of the furanose ring was classified into the $^1E$ envelope conformation. The Gal*f*A molecule was also in the $\beta$-anomeric configuration. Ramos et al. demonstrated that the ratio between the pyranose and the furanose isomers is about 9:1 in aqueous solution (*34*). Thus, appropriate amount of the furanose isomer may exist in the soaking solution.

The interactions between galacturonic acids and endoPG I are shown in Figure 5. The O4 of Gal*p*A hydrogen bonds (2.68 Å) with OD2 of Asp 173. This bonding is important, since it mimics the interaction between the most favorable candidate for the catalytic acid and the glycosidic oxygen during the reaction. Thus, the Gal*p*A binding site is believed to be the +1 subsite, because its location is at the reducing end side of the proposed catalytic residues, Asp173, Asp174, and Asp153. The other two of the four hydroxy groups of Gal*p*A interact directly with the enzyme. Thus, the O2 hydroxy group forms a hydrogen bond with NE2 of His150. The O3 hydroxy group makes two hydrogen bonds, one between ND2 of Asn91 and the other between ND2 of Asn151. The O1 hydroxy group does not interact with the enzyme. The carboxy group of Gal*p*A engages in electrostatic interactions and/or hydrogen bonds with four of the enzyme's residues. O61 of the carboxy group interacts with NH2 of Arg226 and ND1 of His195. The other oxygen of the carboxy group, O62, interacts with NZ of Lys228 and OH of Tyr262.

On the other hand, the binding site for Gal*f*A appears to overlap with the −1 subsite, because it is located on the opposite side of the +1 subsite, across from the catalytic aspartate residues. The distance between the Gal*p*A O4 hydroxy group and the Gal*f*A O5 hydroxy group is only 2.80 Å having no space for an additional galacturonate molecule between the Gal*p*A and Gal*f*A molecules. Six residues of the enzyme participate in the Gal*f*A binding. The O2 atom of Gal*f*A participates in hydrogen bonds with OD1 of Asp153 and ND2 of Asn151. O1 of Gal*f*A interacts with NE2 of Gln120. The binding of the carboxy group of Gal*f*A involves a unique nonprolyl *cis*-peptide bond between Gly200 and

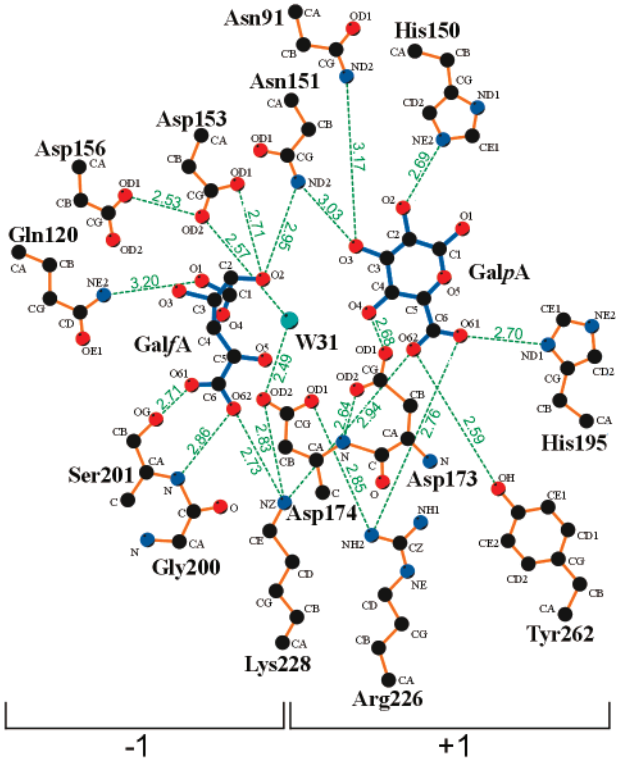


FIGURE 5: Schematic representation of the interactions between the enzyme and the ligands in the ternary complex, drawn by LIGPLOT (*32*). The interactions of Gal*p*A in the binary complex are practically the same as those in the ternary complex. The dotted lines show hydrogen bonds and electrostatic interactions, and their distances are indicated in angstroms.

Ser201 in which O61 interacts with OG of Ser201, and O62 interacts with the amide group of Ser201. Gly200 is necessary to maintain the *cis*-peptide bond, because it has unusual $\phi/\psi$ angles that are not allowed for the other amino acid residues ($\phi$ is 102.8° and $\psi$ is −171.7°). The carboxy group (O62) of Gal*f*A also interacts with NZ of Lys228. In the vicinity of the O4 atom of Gal*p*A, there is a water molecule "W31" that interacts with the carboxy groups of both Asp153 (2.57 Å) and Asp174 (2.49 Å). The distance between O4 of Gal*p*A and W31 in the ternary complex is 4.15 Å. In the *A. niger* endopolygalacturonase II, the water molecule at the equivalent position is considered to be the nucleophilic water of the hydrolytic reaction (*13*).

*Comparison of the Native and Two Complex Structures.* A superposition of the native and two complex structures around the ligand binding site is shown in Figure 6. Overall, while the native and two complex structures are almost the same, some structural changes in the side chains of certain amino acids were observed due to the ligand binding. Comparing the native and binary complex structures, positional changes of the side chains of Asn91, His150, and Arg226 were observed with Gal*p*A binding. In the native crystal structure, the side chain of Arg226 was disordered and was observed to split into two conformations; however, in the binary complex it converged into a single conformation. The convergence of the side chain conformation of Arg226 may have been caused by the electrostatic interaction with the carboxy group of Gal*p*A. The side chains of Asn91 and His150 shifted their positions slightly. These small differences are due to the exchange of the hydrogen bonding network around the bound Gal*p*A molecule. The structural

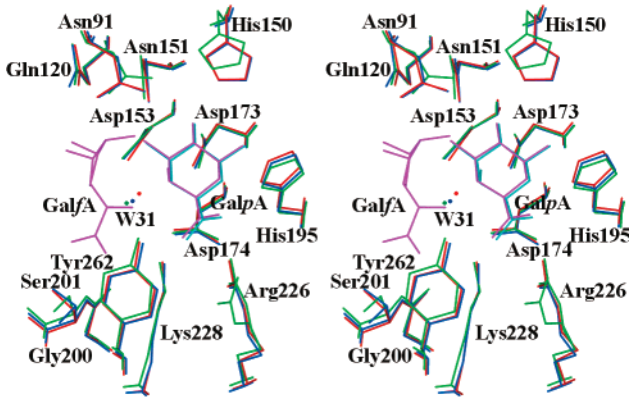


FIGURE 6: A stereoview of the superposed models of the native crystal and the two complexes around the active site. The enzyme models of the native, binary, and ternary complexes are shown in green, blue, and red, respectively. The nucleophilic waters of the native and the two complexes are in the same colors. The ligands of the binary and ternary complexes are in cyan and magenta, respectively.

changes between the binary and ternary complexes were observed in the vicinity of Ser 201. Due to the interaction between the OG atom and the carboxy group of Gal*f*A, the $\chi$1 angle of Ser 201 was shifted from 171° in the native structure to 70° in the ternary complex.

## DISCUSSION

*Active Site Architecture.* The crystal structures of endoPG I complexed with galacturonic acid allow us to identify some structural features responsible for their substrate recognition. Monogalacturonic acid is a product of the endopolygalacturonase reaction. Bond cleavage frequency studies have demonstrated that some endopolygalacturonases hydrolyze only the first glycosidic bond from the reducing end of tri- and tetragalacturonate (*4*−*6*). This property of the reaction indicates that the enzymes form only one productive enzyme−substrate complex with these substrates and most likely recognize the reducing end Gal*p*A residue in the +1 subsite. The +1 subsite is believed to have the highest affinity among the subsites for the polygalacturonate substrate in the endopolygalacturonase reaction. In binary complex, an electron density for one Gal*p*A molecule was observed in the +1 subsite. The tight binding of the substrate to subsite +1 is due to the electrostatic interactions between the carboxy group and the basic residues and the precise recognition of the galactose epimer. In site-directed mutagenesis studies of *A. niger* endopolygalacturonase II, the replacement of the corresponding residues of His195, Arg226, Lys228, and Tyr262 led to 10-fold or greater increases in the $K_m$ value (*19*, *20*). These residues are within hydrogen bonding distances to the carboxy group of the Gal*p*A, and the results of their mutations verified the importance of the carboxy group recognition in subsite +1 for productive substrate binding. In contrast, the replacement of the corresponding residue, Asp173, caused only a 2-fold increase in the $K_m$ value, but greatly decreased the $K_{cat}$ value (*19*). Asp173 is at a distance to make a hydrogen bond with O4 of the bound Gal*p*A. From this structure, we postulated that Asp173 serves as a general acid catalyst that donates a proton to the glycosidic oxygen, as described below in detail.

In the case of Gal*f*A binding of ternary complex, carboxy group of Gal*f*A is also recognized by the characteristic

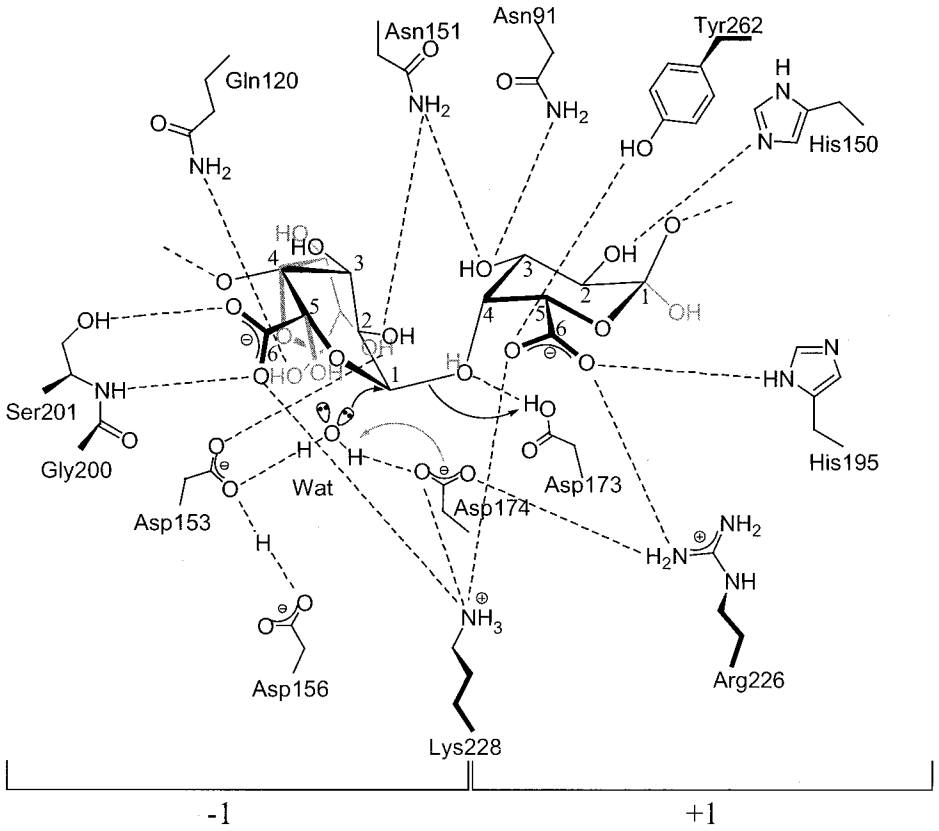


FIGURE 7: Schematic drawing of the proposed structure for substrate binding in the −1 and +1 subsites. The proposed substrate structure (black) was modeled based on the structures of the Gal*p*A and Gal*f*A molecules (gray). The Gal*p*A unit in the −1 subsite of the substrate model is distorted to a half-chair form. The nucleophilic water is in the preferred position to attack the C1 atom of the Gal*p*A unit in the −1 subsite.

structural motifs consists of the side chain of Ser201 and Lys228, and *cis*-peptide bond between Gly200 and Ser201. In the other three known polygalacturonase structures, the nonprolyl *cis*-peptide bond between Gly200 and Ser201 and the Lys228 residue are conserved (*12*−*14*). The *cis*-peptide bond and Lys228 are also conserved in the structure of rhamnogalacturonase A from *Aspergillus aculeatus* (*15*). Rhamnogalacturonase A also belongs to the glycosidase family 28, and it cleaves the α-1,2 glycosidic linkage between galacturonate and rhamnose. The structural conservation of the three residues reveals that the *cis*-peptide bond and Lys228 probably form a carboxy group recognition motif in the −1 subsite of both enzymes. In both the −1 and +1 subsites, the binding of the carboxy group was considered as an important mechanism of substrate recognition. This is probably the reason endopolygalacturonases are able to cleave only free polygalacturonate and not the methylesterified substrate.

On the basis of the structures of Gal*f*A and Gal*p*A bound in the ternary complex, we attempted to construct a structural model of digalacturonic acid with a substrate molecule bound in both the −1 and +1 subsites across from the catalytic residues (Figure 1, Figure 7 and Supporting Information 1). The model of the bound digalacturonate substrate was designed to overlap with the Gal*p*A and Gal*f*A molecules in the ternary complex as much as possible, with special consideration of two restrictions: first, the structure of the Gal*p*A unit bound in the +1 subsite must completely overlap with that of the bound Gal*p*A molecule in the ternary complex, and second, the carboxy group of the second Gal*p*A unit bound in the −1 subsite must overlap with that of the Gal*f*A molecule bound in the ternary complex. These restrictions are likely to be appropriate determinants for the productive substrate binding, as described above, and allowed us to construct a reasonable model. To satisfy the second restriction, the modeled Gal*p*A unit in the −1 subsite should have the distorted half-chair conformation, in which C1 shifts into a more equatorial position from the ordinary chair conformation. Thus, the ring structure of Gal*p*A in the −1 subsite was built based on the distorted conformation of the valienamine pseudosaccharide moiety in acarbose, a transition-state analogue inhibitor of glucoamylase (*35*). The ordinary chair conformation for Gal*p*A in the −1 subsite led to severe steric conflict between the pyranose ring in the −1 subsite and the H4 hydrogen in the +1 subsite, and between the O2 hydroxy group in the −1 subsite and the O3 hydroxy group in the +1 subsite. The resulting model of the digalacturonate substrate shows the following characteristics in the −1 subsite: (i) the pyranose ring of the Gal*p*A residue has been distorted into the $^4H_3$ half-chair conformation. In that half-chair conformation, the C2, C1, O5, and C5 atoms are coplanar, a very important characteristic of the postulated oxocarbenium ion-like transition state in the glycosyl-hydrolase reaction; (ii) the positions of the carboxy group, C4, C5, and O5 in the model are overlapped with those of the observed Gal*f*A structure; and (iii) the observed furanose ring of Gal*f*A is perpendicular to that of the model, but the Gal*p*A unit in the −1 subsite caused no serious conflict with the enzyme and the Gal*p*A unit in the +1 subsite.

*Implications for Catalytic Mechanism*. The two major catalytic mechanisms for glycosyl hydrolases involve two

carboxy residues (*36*). In the case of the retaining enzymes, these residues are involved in a double-displacement mechanism via a covalent glycosyl-enzyme intermediate, and their carboxy groups are separated by approximately 5.5 Å. In the inverting enzymes, on the other hand, the distance between the catalytic carboxy groups is approximately 10 Å, and the reaction proceeds via a single-displacement mechanism. In the latter mechanism, one of the acidic residues acts as a general acid, donating a proton to the glycosidic oxygen of the scissile bond, while the second acidic residue acts as a general base, which activates the nucleophilic water molecule that attacks the anomeric carbon. Endopolygalacturonase is an inverting enzyme (*3*). As indicated previously (*12, 13*), the distances between the absolutely conserved three aspartates (average of the OD1−OD1, the OD1−OD2, the OD2−OD1, and the OD2−OD2 distances) are 4.4 Å (Asp153−Asp173), 5.5 Å (Asp153−Asp174), and 4.8 Å (Asp173−Asp174), respectively. No conserved acidic residues are found at a distance of about 10 Å from each other.

The digalacturonate model structure also sheds light on the catalytic mechanism. All three candidate Asp residues are located on the $\beta$ face of the −1 and +1 subsite pyranoside rings. As described above, Asp173 is at the appropriate position to be a proton donor to the fissile glycosidic bond. Asp153 or Asp174 seems to act as a general base to abstract a proton from the nucleophilic water. Both Asp153 and Asp174 were considered to be accepting hydrogen bonds from a candidate for the nucleophilic water. In this position, one of the lone pairs of tetrahedral oxygen atom of this water molecule is directed toward anomeric carbon in the −1 subsite Gal*p*A unit in the digalacturonate model. This placement is ideal for a nucleophilic attack on the anomeric carbon of the scissile glycosidic bond. Moreover, in this placement, this water molecule, the anomeric carbon in the −1 subsite Gal*p*A unit in the digalacturonate model, and the oxygen of the scissile glycosidic bond are very close and in the in-line coordination necessary for hydrolysis by the inverting mechanism. Consequently, the reaction mechanism of endoPG I is considered to occur as follows. First, the substrate is bound into the cleft, and the Gal*p*A unit in the −1 subsite is distorted into the half-chair conformation. Next, the nucleophilic water is activated by Asp153 or Asp174, and attacks the anomeric carbon of the scissile bond in an $S_N2$ manner. At the same time, Asp 173 provides a proton to the fissile glycosidic oxygen. Last, the glycosidic bond is broken and the anomeric configuration of the Gal*p*A unit in the −1 subsite is inverted to the $\beta$-configuration.

The structure of the digalacturonate model at the active site of endoPG I allows us to compare the active site structures at the subsite −1 and +1 moieties with the other α-glycosidases. A superposition of endoPG I with the diglacturonate model and acarbose complexes of two amylases (*Aspergillus awamori* glucoamylase (*37*), an inverting enzyme, and *Aspergillus oryzae* TAKA-amylase (*38*), a retaining enzyme) based on the structure of the ring of the substrate/inhibitor in the −1 subsite (Supporting Information 2) shows following features: (i) the oxygen atoms which attack the anomeric carbon (nucleophilic waters of endoPG I and glucoamylase, and OD1 of Asp206 of TAKA-amylase) locate approximately the same position in the $\beta$ face of the −1 subsite ring. Moreover, Asp174 of endoPG I, one of the two candidates for base catalyst, overlaps Glu400 of glucoamylase, its base catalyst; (ii) On the other hand, positions of the proton donor are divided into two groups; the acid catalyst of glucoamylase (Glu179) locates in the α face of the −1 subsite ring, whereas the acid or acid/base catalyst of endoPG I (Asp173) and TAKA-amylase (Glu230) locate in $\beta$ face of the ring. Consequently, in endoPG I, protonation of the glycosidic oxygen and nucleophilic attack at the anomeric carbon occur from the same side of the scissile bond as proposed by Pickersgill (*12*). This feature is a major determinant of the unusual distance between the catalytic residues of endoPG I.

*Catalytic Machinery*. Lys228 is likely to be a key player in the carboxy group recognition mechanisms of both the −1 and +1 subsites. It is located at a special position where NZ of Lys228 interacts with three carboxy groups, in both Gal*p*A in the −1 and +1 subsites, and in Asp174. The distances between the NZ atom of Lys228 and the coordinated oxygen atom in the three carboxy groups are almost the same (2.94 Å to O62 of Gal*p*A in the −1 subsite, 2.73 Å to O62 of Gal*f*A in the +1 subsite, and 2.83 Å to OD2 of Asp174). The angles of those three interactions also permit efficient hydrogen bonding. In the view from NZ of Lys228, the three oxygen atoms are placed like the tips of the legs of a tripod. This arrangement would have an advantage in positioning the catalytic residues and the Gal*p*A residues of the substrate in the −1 and +1 subsites through the modulation of those three carboxy residues.

There are two candidates for a general base residue, Asp153 and Asp174. Asp174 interacts with the basic residues, Lys228 and Arg226. This interaction would decrease the p$K_a$ of Asp174, though making it possible for Asp174 to act as a general base. On the other hand, Asp153 interacts with an acidic residue, Asp156, which would make the deprotonation of Asp153 unfavorable. However, the distance between Asp153 and Asp156 would facilitate low-barrier hydrogen bonding (2.53 Å). The low-barrier hydrogen bonding may affect the p$K_a$ of Asp153 to act as a general base. Mutagenesis experiments of the corresponding residues, Asp153 and Asp174, in the *A. niger* endopolygalacturonase II have demonstrated that the Ala mutation of each residue caused a more than 2000-fold decrease in the $k_{cat}$ value (*19*). Further experiments will be necessary to determine which residue is the general base.

## ACKNOWLEDGMENT

The authors thank Drs. Masaki Yamamoto, Takashi Kumasaka, and Shinichi Adachi for assistance in the X-ray data collection on the RIKEN beamlines at SPring-8, Dr. Jun Hiratake of Kyoto University for valuable discussions and critical reading of the manuscript, and the Supercomputer Laboratory at Institute for Chemical Research, Kyoto University, for providing CPU time.

## SUPPORTING INFORMATION AVAILABLE

This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES

1. Collmer, A. (1986) *Annu. Rev. Phytopathol. 24*, 385−409.
2. Henrissat, B. (1991) *Biochem. J. 280*, 309−316.

3. Biely, P., Benen, J., Heinrichova, K., Kester, H. C., and Visser, J. (1996) *FEBS Lett. 382*, 249−255.
4. Hasui, Y., Fukui, Y., Kikuchi, J., Kato, N., Miyairi, K., and Okuno, T. (1998) *Biosci. Biotech. Biochem. 62*, 852−857.
5. Benen, J. A., Kester, H. C., and Visser, J. (1999) *Eur. J. Biochem. 259*, 577−585.
6. Bonnin, E., Le Goff, A., Korner, R., Van Alebeek, G. W., Christensen, T. M., Voragen, A. G., Roepstorff, P., Caprari, C., and Thibault, J. (2001) *Biochim. Biophys. Acta 1526*, 301−309.
7. Miyairi, K., Fujita, T., Okuno, T., and Sawai, K. (1977) *Agric. Biol. Chem. 41*, 1897−1902.
8. Miyairi, K., Okuno, T., and Sawai, K. (1985) *Agric. Biol. Chem. 49*, 1111−1118.
9. Miyairi, K., Senda, M., Watanabe, M., Hasui, Y., and Okuno, T. (1997) *Biosci. Biotech. Biochem. 61*, 655−659.
10. Shimizu, T., Miyairi, K., and Okuno, T. (2000) *Eur. J. Biochem. 267*, 2380−2389.
11. Shimizu, T., Nakatsu, T., Miyairi, K., Okuno, T., and Kato, H. (2001) *Acta Crystallogr. D57*, 1171−1173.
12. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998) *J. Biol. Chem. 273*, 24660−24664.
13. van Santen, Y., Benen, J. A., Schroter, K. H., Kalk, K. H., Armand, S., Visser, J., and Dijkstra, B. W. (1999) *J. Biol. Chem. 274*, 30474−30480.
14. Cho, S. W., Lee, S., and Shin, W. (2001) *J. Mol. Biol. 311*, 863−878.
15. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) *Structure 5*, 533−544.
16. Yoder, M. D., Keen, N. T., and Jurnak, F. (1993) *Science 260*, 1503−1507.
17. Kester, H. C., Kusters-van Someren, M. A., Muller, Y., and Visser, J. (1996) *Eur. J. Biochem. 240*, 738−746.
18. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
19. Armand, S., Wagemaker, M. J., Sanchez-Torres, P., Kester, H. C., van Santen, Y., Dijkstra, B. W., Visser, J., and Benen, J. A. (2000) *J. Biol. Chem. 275*, 691−696.
20. Pages, S., Heijne, W. H., Kester, H. C., Visser, J., and Benen, J. A. (2000) *J. Biol. Chem. 275*, 29348−29353.
21. Evans, G., and Pettifer, R. F. (2001) *J. Appl. Crystallogr. 34*, 82−86.
22. Leslie, A. G. W. (1999) *Acta Crystallogr. D55*, 1696−1702.
23. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
24. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D55*, 849−861.
25. Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol. 72*, 245−270.
26. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol. 6*, 458−463.
27. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr. A47*, 110−119.
28. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
29. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol. 277*, 319−343.
30. Kleywegt, G. J., and Jones, T. A. (1998) *Acta Crystallogr. D54*, 1119−1131.
31. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.
32. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng. 8*, 127−134.
33. Sheldrick, G. M. (1990) *Acta Crystallogr. A46*, 467−473.
34. Ramos, M. L. D., Caldeira, M. M. M., and Gil, V. M. S. (1996) *Carbohydr. Res. 286*, 1−15.
35. Przylas, I., Terada, Y., Fujii, K., Takaha, T., Saenger, W., and Strater, N. (2000) *Eur. J. Biochem. 267*, 6903−6913.
36. Rye, C. S., and Withers, S. G. (2000) *Curr. Opin. Chem. Biol. 4*, 573−580.
37. Aleshin, A. E., Stoffer, B., Firsov, L. M., Svensson, B., and Honzatko, R. B. (1996) *Biochemistry 35*, 8319−8328.
38. Brzozowski, A. M., and Davies, G. J. (1997) *Biochemistry 36*, 10837−10845.

BI025541A