Communication

The Tetramer Structure of the Glycoside Hydrolase Family 27 α-Galactosidase I from Umbelopsis vinacea

Zui FUJIMOTO,1,1 Satoshi KANEKO,2 Wook-Dong KIM,2,* Gwi-Gun PARK,2,3 Mitsuru MOMMA,1 and Hideyuki KOBASHI,2,**

1 Protein Research Unit, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan
2 Food Biotechnology Division, National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan
3 Department of Food and Bioengineering, Kyungwon University, Seongnam, Gyeonggi, 461-701, Republic of Korea

Received August 13, 2009; Accepted September 8, 2009; Online Publication, October 7, 2009 [doi:10.1271/bbb.90604]

The crystal structure of Umbelopsis vinacea α-galactosidase I, which belongs to glycoside hydrolase family 27, was determined at 2.0 Å resolution. The monomer structure was well conserved with those of glycoside hydrolase family 27 enzymes. The biological tetramer structure of this enzyme was constructed by the crystallographic 4-fold symmetry, and tetramerization appeared to be caused by three inserted peptides that were involved in the tetramer interface. The quaternary structure indicated that the substrate specificity of this enzyme might be related to the tetramer formation. Three N-glycosylated sugar chains were observed, and their structures were found to be of the high-mannose type.

Key words: crystal structure; α-galactosidase; glycoprotein; Umbelopsis vinacea; tetramer

α-Galactosidase (E.C. 3.2.1.22) is an exoglycosidase that hydrolyzes an α-linked galactosyl residue from galactooligosaccharides, polymeric galactoglucomannans, and galactolipids. α-Galactosidases are widely distributed in animals, plants, and microorganisms, and in plants, α-galactosidase is one of the key enzymes in the degradation of the cell wall galactomannan during germination.11 α-Galactosidases can be classified into glycoside hydrolase (GH) 4, 27, 36, 57, 97, or 110, according to CAZY database.2,3 α-Galactosidases from eukaryotes have high amino acid sequence homologies and are generally classified into GH27, whereas prokaryotic α-galactosidases are mostly classified into GH36.

To date, the crystal structures of α-galactosidases from rice (Oryza sativa), human (Homo sapiens), and fungi (Hypocrea jecorina), which belong to GH27, have been determined.4,6 These enzymes consist mainly of two domains, an N-terminal domain consisting of a β(β/α)-barrel with a catalytic site and a C-terminal β-structural domain. They have been analyzed in complex with D-galactose, and details of the substrate-binding mechanism were identified. Besides the α-galactosidases, the crystal structures of two α-N-acetylgalactosaminidases from chicken (Gallus gallus) and Bacillus halodurans, and β-L-arabinopyranosidase from Streptomyces avermitilis are available for GH27.7,8

We have purified α-galactosidases from several species and have determined the substrate specificities of each of them.9–13 Umbelopsis vinacea (filamentous fungi, earlier we used the synonym Mortierella vinacea) possesses two types of α-galactosidases, and one of which, α-galactosidase I (UvGalI), has a molecular mass of 50–56 kDa and 240 kDa when analyzed by SDS–PAGE and gel filtration respectively. The broadness of the SDS–PAGE band suggests that UvGalI might be a heterogeneous glycoprotein.14 In addition, gel filtration data suggest that it might exist as a tetramer in solution. Enzymatic investigation indicated that it showed different substrate specificities toward galactomannan,14,15 as compared to the other α-galactosidases. UvGalI is specific only for the terminal α-galactosyl residue that is linked to the non-reducing end mannose of mannan, while U. vinacea α-galactosidase II and rice α-galactosidase are specific for the α-galactosyl side-chain of galactomannooligosaccharides, as well as for the terminal α-galactosyl residue.11,12 A crystallographic analysis has been started in order to obtain the structural basis of the catalytic mechanism and substrate specificity of the enzyme.

The strain of U. vinacea that we used was kindly donated by the Hokkaido Sugar (Tokyo). UvGalI was purified and crystallized as previously reported.16 Crystallization trials of UvGalI yielded two different crystal forms, type 1 and type 2, and structure analyses succeeded for both crystal forms. Their overall structures were similar, and hence we describe the structure analysis here for the type-1 crystal.

Diffraction data for native crystals were obtained at beamline BL41XU, Spring-8, Harima (λ = 1.0313 Å).
A UvGalI crystal was mounted in a quartz glass capillary 0.3 mm in diameter and then flash-frozen in a nitrogen stream at 95 K. Diffraction data were collected using the MarCCD X-ray detector (Rayonix, Evanston, IL). The native data set was processed and scaled using DSP/MOSFLM\(^{17}\) up to 2.0 Å resolution. The crystals belonged to the tetragonal space group I4\(_2\)2\(_2\), with cell dimensions of \(a = b = 142.5\,\text{Å}, c = 130.6\,\text{Å}\) (Table 1).

Structural analysis was performed by the molecular replacement method with the program MolReps.\(^{18}\) A structural model of rice \(\alpha\)-galactosidases (Protein Data Bank code 1UAS, 39% identity with the UvGalI amino acid sequence) was used as a search model.\(^{19}\) The solution was subjected to a cycle of maximum likelihood simulated annealing refinement using the program Refmac5, and model building was done using the auto-modeling program ARP/wARP.\(^{19,20}\) Further model rebuilding was conducted manually with the program QUANTA2000 (Accelrys Software, San Diego, CA) and COOT.\(^{21}\) The model was refined for several cycles with the program Refmac5.\(^{22}\) During the course of refinement, \(F_{\text{obs}} - F_{\text{calc}}\) maps showed the presence of sugar chains attached to the three N-glycosylation sites. Similarly, bound waters, a Tris (tris(hydroxymethyl)aminomethane) molecule, and polyethylene glycol 400 (PEG400) molecules were identified. The structural refinement statistics are summarized in Table 1. The stereochemistry of the model was analyzed with the program PROCHECK.\(^{23}\) The figures of the ribbon structures of the UvGalI structure have been deposited in the Protein Data Bank through the PDB accession no. 3A5V.

The crystal structure of UvGalI was determined at 2.0 Å resolution. The asymmetric unit contained one UvGalI peptide chain of 397 amino acid residues, 22 sugar moieties of three \(N\)-linked carbohydrate chains, 556 water molecules, and one Tris and four PEG400 molecules. Since the enzyme was produced as a dimer, eight water molecules, and one Tris and four PEG400 molecules were identified. The structural refinement statistics are summarized in Table 1. The stereochemistry of the model was analyzed with the program PROCHECK.\(^{23}\) The figures of the ribbon structures of the UvGalI structure have been deposited in the Protein Data Bank through the Protein Data Bank Japan, Osaka University, under code name 3A5V.

### Data collection and Structure Refinement Statistics for UvGalI

| Table 1. Data Collection and Structure Refinement Statistics for UvGalI |
|-----------------|-----------------|-----------------|
| **Data collection** | **Structure refinement** | **Values in parentheses are for the highest resolution shell.** |
| **Cell parameter** | **Resolution (Å)** | **R-factor** |
| \(a, b (\text{Å})\) | 142.5 | 0.142 | 0.125 |
| \(c (\text{Å})\) | 130.6 | 0.158 | 0.204 |
| **No. of unique reflections** | 45,293 | 1.92 | 3.49 |
| **I/σ(I)** | 7.7 | 2.4 | 0.7 |
| **R_{merge} (%)** | 8.4 | 27.9 | 1.6 |
| **Completeness (%)** | 99.9 | 99.9 | 99.9 |
| **Resolution (Å)** | 50–2.0 | 2.11–2.00 | 2.05–2.00 |
| **R-factor** | 0.142 | 0.170 | 0.204 |
| **R_{merge} (%)** | 8.4 | 27.9 | 1.92 |
| **Bond angles (°)** | 0.014 | 0.291 | 1.129 |

*The \(R_{merge}\) factor was calculated using 5% of the unique reflections.*

After the initial phase was calculated, model building was started according to the amino acid sequence of \(U. \text{vinacea} \alpha\)-galactosidase deduced from the DNA sequence (DDBJ databases accession no. S79440),\(^{25}\) but in the course of refinement, there was some inconsistency between the electron density of the model and the amino acid sequence. Amino acids for the model were assumed according to the electron density, and 60 amino acids were finally modeled to be different from the deduced amino acid sequence. The cause of inconsistency is still unknown. We have made efforts to clone the gene encoding UvGalI, but they have not yet succeeded. The structural model, therefore, contains some uncertain residual models, due to the 2.0 Å resolution electron density. Figure 1 shows the differences between the sequences, aligned with the sequence of rice \(\alpha\)-galactosidase.

Figure 2 shows the structure of UvGalI. The monomer structure was conserved to those of known GH27 \(\alpha\)-galactosidases. The catalytic domain was comprised of a \(\beta/\alpha\)_\(8\)-barrel, and the active site pocket was found on the C-terminal side of the central \(\beta\)-strands of the catalytic domain. The C-terminal domain was made up of eight \(\beta\)-strands, forming two tandemly repeated Greek-key motives. Compared to rice \(\alpha\)-galactosidase,\(^{4}\) most of the basic secondary structures in the catalytic domain and the C-terminal domain were well conserved (Fig. 1). The root-mean square difference of Ca-atoms between UvGalI and rice \(\alpha\)-galactosidase was calculated to be 1.1 Å. These structures were found to be in good topological agreement and well superimposed, but some insertions and deletions were observed in the loop regions. Among the differences, UvGalI had three remarkable inserted peptides: peptides 137–141 (Ins1) and 196–215 (Ins2) were located around the active cleft of the \(\beta/\alpha\)_\(8\)-barrel, and peptide 346–351 (Ins3) was located in the C-terminal domain. The largest peptide, 196–215, which was located after the 6th \(\beta\)-sheet of the \(\beta/\alpha\)_\(8\)-barrel, was composed of 20 amino acids and contained two disulfide bridges within the kinked peptide (Cys201–Cys215 and Cys203–Cys208), forming a small spherical domain located in front of the catalytic pocket.

The catalytic pocket of the enzyme was located on the C-terminal side of the central \(\beta\)-barrel of the catalytic domain. The Tris molecule occupied the active site of UvGalI (Fig. 2). Two catalytic residues, Asp129 and Asp189, which act as a nucleophile and an acid/base catalyst respectively, were located on the upper and the lower side of the catalytic pocket with a distance of 6.3 Å in between. Most amino acids in the catalytic pocket were conserved with those of rice \(\alpha\)-galactosi-
dase. Based on the structure of rice α-galactosidase, Trp16 and disulfide bridge Cys101–Cys131 appeared to form hydrophobic surfaces of the pocket, and Cys166, Trp168, and Arg185 appeared to recognize the C-2 hydroxyl group of the galactose molecule.

Previous biochemical analyses have indicated that UvGalI appears to form a tetramer structure, \(^{11}\) and preliminary crystallographic studies of it also supported this hypothesis.\(^ {16}\) The crystal of UvGalI belonging to space group \(I\) contains one molecule in the asymmetric unit, but the tetramer can be constructed by four crystallographic 4-fold symmetry-related molecules of itself around the 4-fold axis (Fig. 2B). The tetramer structure is flat-cylindrical with dimensions of 100 Å along two edges and is 75 Å in thickness. The tetramer interfaces are located mainly between the catalytic domains of the adjacent molecules, and are made up by the van der Waals interactions, nine direct hydrogen bond interactions, and some water-mediated interactions. Among the hydrogen bonds, three bonds were observed between the carbohydrate chain N-linked to Asn155 and the adjacent protein molecule. The other N-linked sugar chains were also located on the surface of the tetramer, and they extended to the solvent. These three sugar chains were located around the tetramer interface.

The catalytic pocket was buried in the tetramer interface. The side-chain of Trp168, contributing to make the catalytic pocket, was involved in tetramer formation, being in the proximal of 4.5 Å to Lys230 of the adjacent molecule. Near the catalytic pocket, there was an inserted peptide, 196–215 (Ins2), from which nine residues participated in tetramer interface formation. As shown in Fig. 2B, this insertion gathered around the 4-fold axis and interacted with the self of the adjacent molecules, forming the tetramer interface. Similarly, the other two insertions were also involved in the tetramer interface. Thus, these insertions appear to play an important role in forming a stable tetramer for this enzyme.

There was a concave surface in the tetramer interface, and this concave maintained the accessibility of the catalytic pocket. Thus, the enzyme appears to have catalytic activity as a tetramer form, but its substrate specificity might be limited because the catalytic pocket is located at the bottom of the concave surface. Since the structure of the catalytic pocket is well conserved with the other α-galactosidas, the unique substrate specificity of UvGalI, that this enzyme is active only for the terminal α-galactosyl residue of the galactomannan, can be attributed to the structure of the concave surface. Further structural studies of the enzyme should reveal its catalytic mechanism in detail.

Three carbohydrate chains were observed in the UvGalI molecule to be N-linked carbohydrates. N-Glycosylated asparagine residues were Asn155, Asn276,

---

**Fig. 1.** Sequence Alignment of UvGalI with Rice α-Galactosidase (OsGal).

Inconsistent amino acids of the deduced sequence from DNA no. S79440 are shown. α-Helices and β-strands are indicated by coils. The catalytic residues and N-glycosylated asparagine residues are indicated by ** and * respectively. Three inserted peptides of UvGalI as compared to rice α-galactosidase are shown as Ins1, Ins2, and Ins3.
and Asn371 (Figs. 1 and 2). Two of these were located in the catalytic domain, and Asn371 was in the C-terminal domain. The electron density of the carbohydrate chain linked to Asn155 is shown in Fig. 3A, and the derived sugar chain structures are shown in Fig. 3B. The carbohydrate moieties in the sugar chain attached to Asn155 were observed best among the three chains, and the sugar chain structure was determined to be of the high-mannose type. The sugar chain attached to Asn155 was located on the reverse side of the (β/α)_{8}-barrel to the catalytic site, extending to the adjacent symmetry-related molecule, and was involved in the tetramer formation through three hydrogen bond interactions. Sugar chains attached to Asn276 and Asn371 extended to solvent regions, and only two or three carbohydrate moieties were identified.

Glycosylation sometimes interferes with crystallization of the protein because of its heterogeneity and
flexibility. In this study, however, UvGalI could be crystallized, probably because UvGalII has a stable tetramer structure having a 4-fold symmetry that is applicable to crystal packing. Furthermore, the tetramerization of UvGalI appears to change its substrate specificity and enhance stability. It remains to determine the biological implications of the tetramerization of UvGalI.

Acknowledgments

This study was supported in part by a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science. We also thank the beamline staff members of SPring-8, Harima, and the beamline staff members of the Photon Factory, Tsukuba, for help with diffraction experiments.

References