Structure and Catalytic Mechanism of a Glycoside Hydrolase Family-127 β-L-Arabinofuranosidase (HypBA1)

Chun-Hsiang Huang1, Zhen Zhu1, Ya-Shan Cheng2,3, Hsiu-Chien Chan1, Tzu-Ping Ko4, Chun-Chi Chen1, Iren Wang4, Meng-Ru Ho4, Shang-Te Danny Hsu4, Yi-Fang Zeng4, Yu-Ning Huang1, Je-Ruei Liu1,2,4,5* and Rey-Ting Guo1*

1Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
2Genozyme Biotechnology Inc., Taipei 106, Taiwan
3Institute of Biochemical Sciences, Taiwan
4Institute of Biological Chemistry, Taiwan
5Institute of Biotechnology, Taiwan
6Institute of Biochemical Sciences, Taiwan
7Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan

Abstract

The β-L-arabinofuranosidase from Bifidobacterium longum JCM 1217 (HypBA1), a DUF1680 family member, was recently characterized and classified to the glycoside hydrolase family 127 (GH127) by CAZy. The HypBA1 exerts exo-glycosidase activity to hydrolyze β-1,2-linked arabinofuranose disaccharides from non-reducing end into individual L-arabinoses. In this study, the crystal structures of HypBA1 and its complex with L-arabinose and Zn2+ ion were determined at 2.23-2.78 Å resolution. HypBA1 consists of three domains, denoted N-, S- and C-domain. The N-domain (residues 1-5 and 434-538) and C-domain (residues 539-658) adopt α/α-barrel architectures, and the S-domain (residues 6-433) adopts an (α/α)6-barrel fold. HypBA1 utilizes the S- and C-domain to form a functional dimer. The complex structure suggests that the catalytic core lies in the S-domain where Cys417 and Glu422 serve as nucleophile and general acid/base, respectively, to cleave the glycosic bonds via a retaining mechanism. The enzyme contains a restricted carbohydrate-binding cleft, which accommodates shorter arabino oligosaccharides exclusively. In addition to the complex crystal structures, we have one more interesting crystal which contains the apo HypBA1 structure without Zn2+ ion. In this structure, the Cys417-containing loop is shifted away due to the disappearance of all coordinate bonds in the absence of Zn2+ ion. Cys417 is thus diverted from the attack position, and probably is also protonated, disabling its role as the nucleophile. Therefore, Zn2+ ion is indeed involved in the catalytic reaction through maintaining the proper configuration of active site. Thus the unique catalytic mechanism of GH127 enzymes is now well elucidated.

Keywords: Glycoside hydrolase; Arabinofuranose; β-L-Arabinofuranosidase; Crystal structure; Synchrotron radiation

Abbreviations

Araf : Arabinofuranose; BSA: Bovine Serum Albumin; DUF: Domain of Unknown Function; EDTA: Ethylenediaminetetraacetic Acid; GH: Glycoside Hydrolase; HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid); HRGPs: Hydroxypyrrole-Rich Glycoproteins; Hyp: Hydroxyproline; MIRAS: Multiple Isomorphous Replacement with Anomalous Scattering; MR: Molecular Replacement; RMD: Root-Mean-Square Deviations; SEC/MA/LS: Size Exclusion Chromatography-Multi-Angle Light Scattering; SLS: Static Light Scattering

Introduction

Arabinofuranose (Araf) is broadly distributed in nature as an important component of glycoconjugates, most existing as α-L-forms. However, certain β-L-arabinofuranosyl linked residues are also found in bacteria and plants, including mycobacterial cell wall arabinans, the hydroxyproline (Hyp)-rich glycoproteins (HRGPs), glycopeptide hormones and biopolymers [1-9]. In spite of the abundance of β-L-Araf-containing sugars (L-Araf) in bacterial and plant cells, degradation and metabolism of these polysaccharides remain largely undiscovered due to the lack of knowledge of corresponding degradative enzymes.

In the previous studies, the β-L-arabinofuranosidase (HypBA1, belonging to GH127 family) and β-L-arabinobiosidase (HypBA2, belonging to GH121 family) from Bifidobacterium longum JCM 1217 that participate in the β-L-arabinooligosaccharides metabolisms have been characterized [10,11]. In the hydrolysis of Hyp-linked β-L-arabinooligosaccharides, HypBA2 releases the β-1,2-linked Araf disaccharide (β-Ara2) from Araf2-β1,2-Araf2-β1,2-Araf-Hyp (Ara2-Hyp). HypBA1 subsequently liberates L-arabinosides by hydrolyzing β-Ara2 [10,11]. In addition, Ara2-Hyp, extensin, potato lectin, pNP-arabin-, galacto-, gluco- and xylo-pyranosides are not. Based on the substrate specificity and catalytic products, HypBA1 was proposed to hydrolyze β-L-arabinooligosaccharides in an exo-acting manner [11]. Moreover, Fujiita mutated three glutamate residues which are strictly conserved among the HypBA1 homologous (Glutamates are commonly observed replacing to aspartate). Therefore, HypBA1 is associated with the formation and metabolism of β-L-arabinofuranose and β-L-arabinobioside residues in a variety of glycosylation.
catalytic important residues among all GH families) and proposed that Glu\textsuperscript{122}, Glu\textsuperscript{138} and Glu\textsuperscript{146} might be potential catalytic residues [11].

Very recently, the crystal structures of the ligand-free and L-Araf complex forms of HypBA1 (PDB ID: 3WKW and 3WKX) were determined [12]. Based on the structural analyses, biochemical experiments and quantum mechanical calculations, Ito showed that the nucleophile function is likely served by Cys\textsuperscript{417} rather than a glutamate [12]. In the meantime, we have determined the crystal structures of HypBA1 in native form, in apo form and in complex with its product L-Araf. Here, by analyzing these structures, the relationship between the glutamate residues, the catalytic cysteine, the Zn\textsuperscript{2+} ion, and the substrate in HypBA1 is elucidated, and implications on the retaining mechanism of GH127-family enzymes are discussed.

Materials and Methods

Protein expression, purification, crystallization and data collection

The expression and purification methods employed for the protein have been described before [13]. To obtain phase information by multiple isomorphous replacement (MIR), the apo HypBA1 crystals grown in 0.4 M ammonium acetate and 18% w/v polyethylene glycol 3350 were used for preparing heavy atom derived crystals by soaking the Heavy Atom Screen Hg kit (Hampton Research). The apo crystals (isomorphous to native crystal) were soaked with various mercury-containing reagents (final concentration 2 mM) in cryoprotectant solution (0.5 M ammonium acetate, 25% w/v polyethylene glycol 3350 and 5% w/v glycerol) for at least 1 hr.

To remove any metal ion from protein, the purified enzyme (both Se-Met protein and native protein are used here) was dialyzed against 100 mM EDTA for two times (at least 12 hrs for each time) before crystallization. The crystals of Se-Met protein without Zn\textsuperscript{2+} ion (apo crystal) diffracted X-rays better than the native crystals. The HypBA1-Araf complex crystal was obtained by soaking the native crystal switch a cryoprotectant solution containing 50 mM Araf.

The X-ray diffraction datasets were collected at the beam line BL13B1 and BL13C1 of the National Synchrotron Radiation Research Center (NSRRC, Taiwan). The data was processed using the program HKL2000 [14]. Prior to structural refinements, 5% randomly selected reflections were set aside for calculating R\textsubscript{free}, as a monitor [15].

Size exclusion chromatography-multi-angle light scattering (SEC/MALS)

Absolute molecular weight of the purified protein was determined by static light scattering (SLS) using a Wyatt Dawn Helos E2 multiangle light scattering detector (Wyatt Technology) coupled to an AKTA Purifier UPC10 FPLC protein purification system using a Superdex 200 10/300 GL size-exclusion column (GE Healthcare). HypBA1 protein with 0.75, 1.4, 1.6, 10, and 20 mg/ml concentration were applied to the size-exclusion column with a buffer containing 20 mM HEPES (pH 7.0) and 0.02% NaN\textsubscript{3} by a flow rate of 0.5 ml/min. A 1.8 mg/ml concentration of BSA was used for the system calibration to scan the crystal to see if there is any metal ion existed.

To remove any metal ion from protein, the purified enzyme (both Se-Met protein and native protein are used here) was dialyzed against 100 mM EDTA for two times (at least 12 hrs for each time) before crystallization. The crystals of Se-Met protein without Zn\textsuperscript{2+} ion (apo crystal) diffracted X-rays better than the native crystals. The HypBA1-Araf complex crystal was obtained by soaking the native crystal switch a cryoprotectant solution containing 50 mM Araf.

The X-ray diffraction datasets were collected at the beam line BL13B1 and BL13C1 of the National Synchrotron Radiation Research Center (NSRRC, Taiwan). The data was processed using the program HKL2000 [14]. Prior to structural refinements, 5% randomly selected reflections were set aside for calculating R\textsubscript{free}, as a monitor [15].

Absolute molecular weight of the purified protein was determined by static light scattering (SLS) using a Wyatt Dawn Helos E2 multiangle light scattering detector (Wyatt Technology) coupled to an AKTA Purifier UPC10 FPLC protein purification system using a Superdex 200 10/300 GL size-exclusion column (GE Healthcare). HypBA1 protein with 0.75, 1.4, 1.6, 10, and 20 mg/ml concentration were applied to the size-exclusion column with a buffer containing 20 mM HEPES (pH 7.0) and 0.02% NaN\textsubscript{3} by a flow rate of 0.5 ml/min. A 1.8 mg/ml concentration of BSA was used for the system calibration to scan the crystal to see if there is any metal ion existed.

The overall structure

In this study, the crystal structures of native (with a Zn\textsuperscript{2+} ion) and apo (no Zn\textsuperscript{2+} ion observed) and in complex with L-arabinose (HypBA1-L-Araf) were determined. HypBA1 folds into three domains, designated N-domain (N-terminal domain), S-domain (all α-domain,
substrate-binding domain) and C-domain (C-terminal domain) (Figure 1A). Both adopting β-jellyroll folds, the N- (residues 1-5 and 434-538) and C-domain (residues 539-658) consist of 10 β-strands and 3 small α-helices, respectively. The S-domain (residues 6-433) comprises an (α/α) 6-barrel and accommodates an L-Ara-binding cavity, so it is considered as the catalytic domain (Figure 1A).

In order to assess the unique structural feature of the N- and C-domains, a structural homology searches with DALI [28] were carried out (Figure S1). It reveals that N-domain (colored in green, broken square) of HypBA1 displays structural similarity to N-terminal domain of ErbB4 kinase (colored in orange) with RMSD of 3.0 Å (274 Ca atoms with 10% sequence identity), 3.5 Å (390 Ca atoms with 10% sequence identity), 2.6 Å (501 Ca atoms with 10% sequence identity) and 2.8 Å (1029 Ca atoms with 9% sequence identity), respectively [29-32]. On the other hand, C-domain (colored in green, broken square) of HypBA1 shows structural similarity to C-terminal domain of GH39, colored in yellow) and α-L-rhamnosidase (GH78, colored in orange) with RMSD of 3.0 Å (512 Ca atoms with 14% sequence identity), 3.5 Å (390 Ca atoms with 10% sequence identity), 2.6 Å (501 Ca atoms with 10% sequence identity) and 2.8 Å (1029 Ca atoms with 9% sequence identity), respectively [29-32]. The S-domain (residues 6-433) comprises an (α/α) 6-barrel and accommodates an L-Araf-binding cavity, so it is considered as the catalytic domain (Figure 1A).

Table 1: Data collection and refinement statistics for HypBA1.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Native</th>
<th>Apo (SeMet)</th>
<th>HypBA1-L-Araf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>25.00 - 2.78(2.88)</td>
<td>25.00 - 2.25(2.33)</td>
<td>25.00 - 2.23(2.31)</td>
</tr>
<tr>
<td>Space group</td>
<td>P3_2 1</td>
<td>P3_2 1</td>
<td>P3_2 1</td>
</tr>
<tr>
<td>Unit-cell a, b, c (Å)</td>
<td>75.9, 75.9, 254.0</td>
<td>75.9, 75.9, 254.6</td>
<td>77.6, 77.6, 254.2</td>
</tr>
<tr>
<td>No. of Unique reflections</td>
<td>21797 (2178)</td>
<td>41517 (4078)</td>
<td>43124 (4226)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.6 (9.0)</td>
<td>7.1 (7.1)</td>
<td>5.6 (5.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.9 (100.0)</td>
<td>100.0 (100.0)</td>
<td>97.9 (98.3)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>36.9 (6.1)</td>
<td>23.3 (2.6)</td>
<td>36.2 (4.4)</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>8.6 (48.8)</td>
<td>8.1 (49.2)</td>
<td>5.0 (44.7)</td>
</tr>
</tbody>
</table>

Refinement

| No. of reflections | 20120 (1778) | 40022 (3803) | 40524 (3318) |
| Rprotein (%)      | 22.3 (0.34) | 21.5 (30.0) | 20.4 (28.9) |
| Rfree (%)          | 27.0 (0.40) | 25.5 (30.7) | 24.4 (32.8) |

Geometry deviations

| Bond lengths (Å) | 0.005 | 0.004 | 0.008 |
| Bond angles (°)  | 1.15  | 1.06  | 1.38  |

Ramachandran plot (%)

| Most favored       | 90.9  | 95.9  | 95.9  |
| Allowed            | 8.6   | 3.9   | 3.7   |
| Disallowed         | 0.5   | 0.2   | 0.4   |

Values in parentheses are for the outermost resolution shells

In order to investigate whether the L-Araf (product) could cause conformational change, the native, apo and complex structures were superimposed and no obvious conformational change was observed (data not shown). However, a few regions which are close to the L-Araf-binding cavity were missing in both the native and apo structures (residues 31-50 and 247-250 for native, residues 35-50 and 413-415 for apo), but all these loop regions can be clearly seen in the complex structure. Interestingly, the longest loop consisting of residue 31-50 was found stretching across the surface of the S-domain to cover, or cap, the ligand-binding cavity (Figure 1B). From the complex structure, the capping loop undergoes induced fit and is then stabilized through interacting with the L-Araf. Accordingly, the capping loop might play an important role in the catalytic reaction by regulating the substrate binding because one of the ligand binding residues, Gln⁴⁵⁵, is located on the capping loop and a similar case has been reported [35].

On the other hand, although not supplemented in the crystallization solution, a Zn²⁺ ion was observed adjacent to the active site as a strong

![Figure 1](image_url)
electron density. It was further validated by using X-ray fluorescence scan analysis and anomalous difference Fourier map (Figure 2A). Based on the $|F_o| - |F_c|$ omit map, it is clear that the Zn$^{2+}$ ion forms a typical tetrahedral complex with four Zn$^{2+}$-ion-chelating residues, including Glu$^{338}$, Cys$^{340}$, Cys$^{417}$ and Cys$^{418}$ and all the metal-to-ligand distances are about 2.3 Å (Figure 2A). Interestingly, even though no significant conformational change was caused by Zn$^{2+}$ binding, the two Zn$^{2+}$-ion-chelating residues Cys$^{417}$ and Cys$^{418}$ in the apo structure were shifted away from the original positions (Figure 2B). Therefore, when the Zn$^{2+}$ ion is absent only a broken loop can be observed because of the absence of stabilizing interactions (Figure 2C).

Dimerization

Interestingly, although there is only one HypBA1 monomer in an asymmetric unit, HypBA1 forms a dimer with a crystallographic symmetry-related molecule (Figure 3A and 3B). An analysis with PDBePISA shows that the contact interface encompasses 77 residues that bury a total surface area of about 2781 Å$^2$ on the S- and C-domain [36]. The intermolecular forces include hydrogen bonds (not shown) and salt bridges (Figure 3C). To confirm that HypBA1 also forms a dimer in solution, size exclusion chromatography coupled with multi-angle light scattering (SEC/MALS) was conducted. SEC/MALS offers an estimate of the absolute molecular weights in solution based on the angular dependence of scattered light intensity, which is less dependent on the molecular shapes. At protein concentrations of 0.75-1.60 mg/ml, the SEC/MALS analysis using a Superdex 200 10/300 GL column indicates that its molecular mass is 134.9-138.8 kDa, corresponding to a dimeric form of HypBA1 (Figure 4A). As the elution peak of the
SEC/MALS is relatively symmetric, the calculated molecular weight distribution indicates that the sample is monodispersed. Accordingly, our SEC/MALS data suggests that the HypBA1 protein exists as a very stable dimer in solution under low protein concentrations or at increased protein concentrations of 10 mg/ml or 20 mg/ml (data not shown).

Furthermore, attempting to determine the molecular weight of HypBA1 at an even lower concentration in solution, we applied sedimentation velocity experiment by using analytical ultracentrifugation (AUC-SV). At a concentration of 6 μM, the polypeptide was detected as a single species with a sedimentation coefficient of 7.77 S, which corresponds to a molecular mass of 143.0 kDa, the mass of a HypBA1 dimer (Figure 4B and 4C). Together, the results, by employing two independent biophysical methods elucidate the oligomeric state of the recombinant HypBA1 in solution. At a much higher protein concentration (ca. 10 mg/ml), a symmetric elution peak was still observed and the corresponding molecular weight also coincided with the dimeric forms of HypBA1. The results are consistent with the AUC-SV data, which showed that HypBA1 was dimeric at a low protein concentration and only a small population of higher oligomers emerged when the protein concentration was increased (Figure 4). Both the results of the SEC/MALS and AUC-SV analyses are in a good agreement with the crystallographic findings that HypBA1 exhibits as a very stable dimerization propensity.

Ligand binding and substrate modeling

The \(|F| - |\bar{F}|\) omit map and anomalous difference Fourier map of the bound L-Araf and Zn²⁺ ion are both very clear, respectively (Figure 2A). Based on the binding mode of L-Araf, there are ten hydrogen bonds between the sugar unit and eight amino acid residues including Glu⁶, His⁴¹, His⁸⁹, His¹³⁰, Glu³⁻², Glu⁵⁸, Tyr²⁰⁶ and Cys¹³⁶ (Figure 5A), but not Glu⁴⁶⁶. According to the relative spatial positions of L-Araf and Zn²⁺ ion in the complex structure, Zn²⁺ ion might not directly involve in the catalytic reaction because the Zn²⁺ ion is distantly located to the C1 of product (5.0 Å) (data not shown). Moreover, the configuration of Zn²⁺ ion also makes it unlikely to activate a water molecule for catalysis (which would turn out an inverted α-sugar). This is because the Zn²⁺ ion has formed an almost perfect tetrahedral coordination with Glu⁴¹⁸, Cys³⁹⁸, Cys⁴¹⁷ and Cys⁴¹⁶ (Figure 2A).

By analyzing the HypBA1-L-Araf complex structure, we believe that the bound L-Araf (product) corresponds to the -1 subsite. The space adjacent to the O2 atom is too small to accommodate a sugar residue. To further elucidate possible substrate binding mode, a two-sugar-units substrate (Araf-β1,2-Araf-β-Araφ) was manually modeled into the potential subsites from -1 to +1 in the substrate-binding cavity (Figure 5B). The model was subsequently subjected to several cycles of energy minimization with CNS [24]. Interestingly, the location of the simulated model seems almost fit the size and shape of calculated cavity map (Figure 5B), which is generated by using the web server POCASA (POCKET-CAvity Search Application) [37]. Therefore, the accuracy of the simulated model is justified. In this model, the side chain of Glu³² also binds to the O1 of the +1 sugar, which is in turn bound to Tyr²⁰⁶, and two other residues Glu⁴³ and Tyr²⁰⁶ can interact with the O5 of the same +1 sugar. Beyond the +1 sugar, there is no room to accommodate additional L-Araf and Hyp units unless the capping loop is opened. How the enzyme binds to Ara₂⁻ and Ara₃⁻Hyp remains to be elucidated.

Proposed catalytic mechanism

As previously mentioned, three potential catalytic residues (Glu³⁻², Glu⁴⁶⁶ and Glu⁴⁶⁶) have been proposed. Among them, Glu⁴⁶⁶ is too far away from the substrate-binding cavity and unlikely to participate directly in the catalytic reaction (Figure 1A). However, the mutant E366A had 16% activity left in a previous study [11]. Consequently, Glu⁴⁶⁶ might play a role in structural stability, although not participating directly in the catalytic reaction. By contrast, the residues Glu³⁻² and Glu⁴⁶⁶ are more reasonable catalytic amino acids due to their proximal
containing HypBA1 structure in ligand-free and complex forms [12]. The subsequent structure-based mutagenesis and biochemical analysis, in conjunction with quantum mechanical calculations, allowed Ito to make a clear proposal that the nucleophile should be Cys417 rather than Glu338 [12]. In our apo structure, the Cys417-containing loop is shifted away due to the disappearance of all coordinate bonds in the absence of Zn2+ ion (Figure 2C). Cys417 is thus diverted from the attack position, and probably is also protonated, disabling its role as the nucleophile. Therefore, Zn2+ ion is involved in the catalytic reaction through maintaining the proper configuration of active site.

As said by Ito, however, we cannot rule out other possibility of catalytic reaction mechanism, such as the utilization of two carboxylate residues (Glu322 and Glu338) separated by a suitable distance (5.4 Å) for retaining mechanism. In this case, the bound L-Araf should represent the +1 sugar and the -1 sugar would be severely skewed to fit into the limited space, which is not likely. On the other hand, a recent review suggests that some GH families employ novel mechanisms instead of typical carboxylate/base/nucleophile, including substrate-assisted mechanisms, proton transferring network, utilization of non-carboxylate residues and utilization of an exogenous base/nucleophile [38]. Interestingly, apart from Glu322, Glu338, Cys340, Cys417 and Cys418, Tyr386 is also strictly conserved among several GH127 members (Figure 3). The side chain of Tyr386 is equally close to the C1 of L-Araf (3.2 Å) as is that of Cys417, and it may correspond to the non-carboxylate residue in an alternative mechanism. However, the lack of a base to subtract its proton renders Tyr386 a weak nucleophile. Consequently, the most reasonable catalytic mechanism may involve a Cys417-sugar intermediate, as shown in Scheme 1. Besides the use of a different nucleophile (Cys417 rather than an Asp or Glu) in the first step, the remaining steps are almost the same as those of classic retaining mechanism.

In summary, the native, apo and complex crystal structures of HypBA1 give us a first glimpse of the GH127 family with respect to protein folding and catalytic mechanism. The results presented here shall provide a critical starting point and a firm basis for further studies of the GH127 family. In addition to the catalytic S-domain, HypBA1 also contains N-domain and C-domain, the latter participating in protein folding and catalytic mechanism. The results presented here shall provide a critical starting point and a firm basis for further studies of the GH127 family. In addition to the catalytic S-domain, HypBA1 also contains N-domain and C-domain, the latter participating in dimer formation. To investigate the functions of this novel multi-domain protein, further experiments with mutagenesis and truncation are required.

Acknowledgement

This work was supported by National High Technology Research and Development Program of China (2012AA022200), Academy of Sciences (2011B09030042), Tianjin Municipal Science and Technology Commission (12CCZDSY12500) and National Science Council of Taiwan (NSC 102-2628-B-002-007-MY2). The synchrotron data collection was conducted at beam line BL13B1 and BL13C1 of NSRRC (National Synchrotron Radiation Research Center, Taiwan).

Figure 5: The HypBA1 active-site. (A) Stereo view of the detailed interaction networks of active-site residues with Zn2+ ion and L-Araf. There are four coordinate bonds (green dash lines) and ten hydrogen bonds (black dash lines) between the active-site residues and the bound ligands. (B) The size and shape of the potential carbohydrate-binding cavity are calculated from the web server POCASA and shown as grey map with PyMOL.
References


