

**Open Access** 

# Structure and Catalytic Mechanism of a Glycoside Hydrolase Family-127 $\beta$ -L-Arabinofuranosidase (HypBA1)

Chun-Hsiang Huang<sup>1</sup>, Zhen Zhu<sup>1</sup>, Ya-Shan Cheng<sup>2,3</sup>, Hsiu-Chien Chan<sup>1</sup>, Tzu-Ping Ko<sup>4</sup>, Chun-Chi Chen<sup>1</sup>, Iren Wang<sup>4</sup>, Meng-Ru Ho<sup>4</sup>, Shang-Te Danny Hsu<sup>4,6</sup>, Yi-Fang Zeng<sup>7</sup>, Yu-Ning Huang<sup>7</sup>, Je-Ruei Liu<sup>5,7,8\*</sup> and Rey-Ting Guo<sup>1\*</sup>

<sup>1</sup>Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

<sup>2</sup>Genozyme Biotechnology Inc., Taipei 106, Taiwan

<sup>3</sup>AsiaPac Biotechnology Co., Ltd., Dongguan 523808, China

<sup>4</sup>Institute of Biological Chemistry, Taiwan

<sup>5</sup>Agricultural Biotechnology Research Center, Academia Sinica, Taipei 115, Taiwan

<sup>6</sup>Institute of Biochemical Sciences, Taiwan

<sup>7</sup>Institute of Biotechnology, Taiwan

<sup>8</sup>Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan

# Abstract

The  $\beta$ -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  $\beta$ -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 Zn<sup>2+</sup> 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  $\beta$ -jellyroll architectures, and the S-domain (residues 6-433) adopts an ( $\alpha/\alpha$ )<sub>6</sub>-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 Cys<sup>417</sup> and Glu<sup>322</sup> serve as nucleophile and general acid/base, respectively, to cleave the glycosidic 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 Zn<sup>2+</sup> ion. In this structure, the Cys417-containing loop is shifted away due to the disappearance of all coordinate bonds in the absence of Zn<sup>2+</sup> ion. Cys417 is thus diverted from the attack position, and probably is also protonated, disabling its role as the nucleophile. Therefore, Zn<sup>2+</sup> 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)-1piperazineethanesulfonic acid); HRGPs: Hydroxyproline-Rich Glycoproteins; Hyp: Hydroxyproline; MIRAS: Multiple Isomorphous Replacement with Anomalous Scattering; MR: Molecular Replacement; RMSD: Root-Mean-Square Deviations; SEC/MALS: 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  $\alpha$ -L-forms. However, certain  $\beta$ -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  $\beta$ -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  $\beta$ -L-arabinofuranosidase (HypBA1, belonging to GH127 family) and  $\beta$ -L-arabinobiosidase (HypBA2, belonging to GH121 family) from *Bifidobacterium longum* JCM 1217 that participate in the  $\beta$ -L-arabinooligosaccharides metabolisms

have been characterized [10,11]. In the hydrolysis of Hyp-linked  $\beta$ -L-arabinooligosaccharides, HypBA2 releases the  $\beta$ -1,2-linked Araf disaccharide ( $\beta$ -Ara<sub>2</sub>) from Araf- $\beta$ -1,2-Araf- $\beta$ 1,2-Araf $\beta$ -Hyp (Ara<sub>3</sub>-Hyp). HypBA1 subsequently liberates L-arabinoses by hydrolyzing  $\beta$ -Ara<sub>2</sub> [10,11]. In addition, Ara-, Ara<sub>2</sub>- and Ara<sub>3</sub>-Hyp as well as Ara<sub>2</sub>- and Ara-Me (methanol) are also preferred substrates for HypBA1, but Ara<sub>4</sub>-Hyp, extensin, potato lectin, pNP-arabino-, galacto-, gluco- and xylo-pyranosides are not. Based on the substrate specificity and catalytic products, HypBA1 was proposed to hydrolyze  $\beta$ -L- arabinooligosaccharides in an exo-acting manner [11]. Moreover, Fujita mutated three glutamate residues which are strictly conserved among the HypBA1 homologous (Glutamates are commonly observed

\*Corresponding author: Rey-Ting Guo, Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China, Tel: 86-022-84861999; Fax 86-022-24828701; E-mail: guo\_rt@tib.cas.cn

Je-Ruei Liu, Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan, Tel: 886-2-3366-6011; Fax: 886-2-3366-6001; E-mail: irliu@ntu.edu.tw

Received June 28, 2014; Accepted July 23, 2014; Published August 03, 2014

Citation: Huang CH, Zhu Z, Cheng YS, Chan HC, Ko TP, et al. (2014) Structure and Catalytic Mechanism of a Glycoside Hydrolase Family-127  $\beta$ -L-Arabinofuranosidase (HypBA1). J Bioprocess Biotech 4: 171 doi: 10.4172/2155-9821.1000171

**Copyright:** © 2014 Huang CH, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

catalytic important residues among all GH families) and proposed that Glu<sup>322</sup>, Glu<sup>338</sup> and Glu<sup>366</sup> 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<sup>417</sup> 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<sup>2+</sup> 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 [13] were used for preparing heavy atom derived crystals by using the Heavy Atom Screen Hg kit (Hampton Research). The apo crystals (isomorphous to native crystal) were soaked with various mercurycontaining 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  $Zn^{2+}$  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 of HKL2000 [14]. Prior to structural refinements, 5% randomly selected reflections were set aside for calculating  $R_{\rm free}$  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 Heleos II 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<sub>2</sub> by a flow rate of 0.5 ml/min. A 1.8 mg/ml concentration of BSA was used for the system calibration as a control. The absolute molecular weights of individual peaks in the size-exclusion chromatogram were determined by the SLS data in conjunction with the refractive index measurements (Wyatt Optilab rEX, connected downstream of the LS detector). A standard value of refractive index, dn/dc=0.185 ml/g, was used for proteins and the buffer viscosity h=1.0164 cP at 25°C was calculated using SEDNTERP. The value of reference refractive index, 1.3441 RIU, was taken directly from the measurement of the Wyatt Optilab rEX when buffer only passing through the reference cell.

Page 2 of 7

#### Sedimentation velocity

Sedimentation velocity was performed using an XL-A analytical ultracentrifuge (Beckman Coulter, Fullerton, CA) using absorption optics. HypBA1 and control reference buffer (approximately 400 µL) were added to 12 mm thick epon double-sector centerpieces in an AN60-Ti rotor and spun at 20°C and 45,000 rpm after an initial 90min temperature equilibration period. Detection of concentrations as a function of radial position and time was performed by optical density measurements at wavelength of 280 nm and absorbance profiles were recorded every 3 min. The protein samples were in the buffer containing 20 mM HEPES, pH 7 with a concentration of ca. 6 µM. The buffer density and viscosity were calculated by SEDNTERP [16] and the sorted data were analyzed with the standard c(s) model [17] in SEDFIT version 14.3. The plot of analytical ultracentrifugation concentration profiles and their fits were made with the software GUSSI, provided by Dr. Chad Brautigam in Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, USA).

#### Structural determination and refinement

The HypBA1 structure was solved by MIR with anomalous scattering (MIRAS) using one native and two mercury datasets  $[(C_2H_5HgO)HPO_2 \text{ and } C(HgOOCCH_3)_4]$ . In our previous studies, mercury can bind to free Cys residues very easily and the phase problem can be easily solved by using at least two mercury datasets by MIR [18-21]. The mercury atom binding sites were located using AutoSol wizard of PHENIX using 2.4 Å resolution cutoff [22]. Two mercury atom binding sites were identified and following density modification for phasing improvement was performed (figure of merit = 0.33). Automatic initial model building was carried out by AutoBuild of PHENIX. A 94% completeness model (620 residues with side chains) was built. Subsequent model building and structural refinement were carried out by using the programs COOT [23] and CNS [24].

The apo (treated with EDTA; no Zn<sup>2+</sup> ion observed) and productcomplex (HypBA1-L-Araf) structures were determined by using the molecular replacement (MR) method with PHASER [25] using refined native structure as a search model. The  $2F_o$ - $F_c$  difference Fourier map showed clear electron densities for most amino acid residues. Subsequent refinements by incorporating ligands and water molecules were according to 1.0  $\sigma$  map level. The refined structures were validated by using RAMPAGE [26]. Data collection and refinement statistics of these crystals are summarized in Table 1 and Supplementary Table S1. All Figures were prepared by using PyMOL [27].

#### X-ray fluorescence scan analysis

To investigate the metal ion presented in the HypBA1 crystals, fluorescence scan analysis of the protein crystal was performed at beam line BL13B1 of the National Synchrotron Radiation Research Center (NSRRC, Taiwan). The wavelength from 12 eV to 25588 eV was used to scan the crystal to see if there is any metal ion existed.

# **Results and Discussion**

# The overall structure

In this study, the crystal structures of native (with a Zn<sup>2+</sup> ion) and apo (no Zn<sup>2+</sup> 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  $\alpha$ -domain;

Data collection	Native	Apo (SeMet)	HypBA1-L-Araf
Resolution (Å)	25.00 - 2.78(2.88 - 2.78)	25.00 - 2.25(2.33 - 2.25)	25.00 - 2.23(2.31 - 2.23)
Space group	<i>P</i> 3 <sub>2</sub> 21	<i>P</i> 3,21	<i>P</i> 3 <sub>2</sub> 21
Unit-cell a, b, c (Å)	75.9, 75.9, 254.0	75.9, 75.9, 254.6	77.6, 77.6, 254.2
No. of Unique reflections	21797 (2178)	41517 (4078)	43124 (4226)
Redundancy	7.6 (9.0)	7.1 (7.1)	5.6 (5.7)
Completeness (%)	97.9 (100.0)	100.0 (100.0)	97.9 (98.3)
Mean I/o(I)	36.9 (6.1)	23.3 (2.6)	36.2 (4.4)
R <sub>merae</sub> (%)	6.8 (48.8)	8.1 (49.2)	5.0 (44.7)
Refinement			
No. of reflections	20120 (1778)	40022 (3803)	40524 (3318)
R <sub>work</sub> (95% of data)	22.3 (0.34)	21.5 (30.0)	20.4 (28.9)
R <sub>free</sub> (5% of data)	27.2 (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
No. of atoms / mean <i>B</i> -values (Å <sup>2</sup> )			
Protein atoms	4971 / 66.8	5018 / 44.2	5152 / 52.2
Water molecules	58 / 57.8	279 / 43.6	312 / 57.8
zinc ion	1 / 78.4	-	1 / 44.4
Ligand atoms	-	-	10 / 43.7
PDB ID	3WRE	3WRF	3WRG

Values in parentheses are for the outermost resolution shells

substrate-binding domain) and C-domain (C-terminal domain) (Figure 1A). Both adopting  $\beta$ -jellyroll folds, the N- (residues 1-5 and 434-538) and C-domain (residues 539-658) consist of 10  $\beta$ -strands, and 7  $\beta$ -strands and 3 small  $\alpha$ -helices, respectively. The S-domain (residues 6-433) comprises an ( $\alpha/\alpha$ )<sub>6</sub>-barrel and accommodates an L-Arafbinding 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 C-terminal domains of xyloglucanase (GH44, colored in cyan), human α-galactosidase (GH27, colored in magenta), β-xylosidase (GH39, colored in yellow) and a-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]. On the other hand, C-domain (colored in green, broken square) of HypBA1 shows structural similarity to N-terminal domain of ErbB4 kinase (colored in cyan) and LKB1 (colored in salmon) with RMSD of 2.6 Å (274 Ca atoms with 4% sequence identity) and 3.5 Å (311 Ca atoms with 15% sequence identity), respectively [33,34] (SI Figure S1). However, their functional roles remain to be investigated. A possibility is that these domains might associate with substrate to facilitate catalytic reaction, as a carbohydrate-binding motif does.

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-Arafbinding 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, Gln<sup>45</sup>, is located on the capping loop and a similar case has been reported [35].

Page 3 of 7

On the other hand, although not supplemented in the crystallization solution, a  $Zn^{2+}$  ion was observed adjacent to the active site as a strong



**Figure 1:** The overall structure of HypBA1-L-Araf. (A) A cartoon model of the HypBA1-L-Araf complex structure. Each domain is presented in different color: N-, S- and C-domain are in orange, cyan and green colors, respectively. The Zn<sup>2+</sup> ion is shown as a magenta sphere, the bound L-Araf molecule and Glu<sup>366</sup> is shown as yellow stick model. (B) The obvious difference between the native, apo and ligand-bound HypBA1 is the capping loop (colored in magenta). The surface representation of the complex structure by omitting the capping loop clearly shows that the loop covers the active site cleft upon ligand binding.

Table 1: Data collection and refinement statistics for HypBA1.

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<sup>2+</sup> ion forms a typical tetrahedral complex with four Zn<sup>2+</sup> ion-chelating residues, including Glu<sup>338</sup>, Cys<sup>340</sup>, Cys<sup>417</sup> and Cys<sup>418</sup> 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<sup>2+</sup> binding, the two Zn<sup>2+</sup> ion-chelating residues Cys<sup>417</sup> and Cys<sup>418</sup> in the apo structure were shifted away from the original positions (Figure 2B). Therefore, when the Zn<sup>2+</sup> ion is absent only a broken loop can be observed because of the absence of stabilizing interactions (Figure 2C).



**Figure 2:** The Zn<sup>2+</sup> ion binding site. (A)  $|F_0| - |F_0|$  omit map (contoured at 4.50) for Zn<sup>2+</sup>-chelating residues, including Glu<sup>338</sup>, Cys<sup>340</sup>, Cys<sup>417</sup> and Cys<sup>418</sup>, and L-Araf (product) and anomalous difference Fourier map (contoured at 12 $\sigma$ ) for identifying the location of the Zn<sup>2+</sup> ion are shown as green and red color, respectively. (B) The superimposition of the native (colored in green), apo (colored in cyan) and complex structure (colored in magenta) shows that the two Zn<sup>2+</sup> ion-binding residues (Cys<sup>417</sup> and Cys<sup>418</sup>) in the apo structure are flipped over in the absence of Zn<sup>2+</sup> ion. (C) The superimposition of the apo (colored in cyan) and complex structure (colored in magenta) reveals that the Cys<sup>417</sup> - and Cys<sup>418</sup>-containing loop is broken because of lack of Zn<sup>2+</sup> ion interaction.



**Figure 3:** Dimerization of HypBA1. (A) Dimeric HypBA1 structure with one bound L-Araf molecule and a Zn<sup>2+</sup> ion in each monomer (Figure 1). Because there is only one molecule in an asymmetric unit, the dimer is constructed by generating a symmetry-related partner. For clarity, the contact domains are presented in different colors (green versus wheat and cyan versus grey). (B) Front view of the HypBA1 dimer. (C) Close-up view of the interaction network of salt bridges at the dimer interface.

# 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<sup>2</sup> 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 multiangle 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).



Figure 4: SEC/MALS and AUC-SV analysis. (A) SEC/MALS profiles of 0.75 (black), 1.20 (red) and 1.65 mg/ml (blue) of HypBA1 proteins (100  $\mu$ l per injection). The thin line segments represent the calculated molar masses and the corresponding molecular weights of each peak (left ordinate axis) are given in number with the unit of kDa. The solid lines represent normalized UV absorbance (280 nm, right ordinate axis) and dashed lines are signals of light scattering. (B) A typical tracing of absorbance at 280 nm of the data of HypBA1 during the sedimentation velocity experiment. The symbols (open circles) are experimental data and the lines are computer-generated results by fitting the experimental data to the Lamm equation using the SEDFIT program. The residuals of the model fitting were shown in the lower panel. (C) Sedimentation coefficient distributions resulting from the fitting in (B). The major peak is compatible with a dimeric form of HypBA1 with its hydrodynamic properties.

Page 5 of 7

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). Taken 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  $|F_o| - |F_c|$  omit map and anomalous difference Fourier map of the bound L-Araf and Zn<sup>2+</sup> 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 Gln<sup>45</sup>, His<sup>142</sup>, His<sup>194</sup>, His<sup>270</sup>, Glu<sup>322</sup>, Glu<sup>338</sup>, Tyr<sup>386</sup> and Cys<sup>415</sup> (Figure 5A), but not Glu<sup>366</sup>. According to the relative spatial positions of L-Araf and Zn<sup>2+</sup> ion in the complex structure, Zn<sup>2+</sup> ion might not directly involve in the catalytic reaction because the Zn<sup>2+</sup> ion is distantly located to the C1 of product (5.0 Å) (data not shown). Moreover, the configuration of Zn<sup>2+</sup> ion also makes it unlikely to activate a water molecule for catalysis (which would turn out an inverted  $\alpha$ -sugar). This is because the Zn<sup>2+</sup> ion has formed an almost perfect tetrahedral coordination with Glu<sup>338</sup>, Cys<sup>340</sup>, Cys<sup>417</sup> and Cys<sup>418</sup> (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 twosugar-units substrate (Araf-\beta1,2-Araf; \beta-Ara,) 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 Glu<sup>322</sup> also binds to the O1 of the +1 sugar, which is in turn bound to Tyr<sup>386</sup>, and two other residues Gln44 and Tyr250 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 Ara<sub>2</sub>- and Ara<sub>3</sub>-Hyp remains to be elucidated.

#### Proposed catalytic mechanism

As previously mentioned, three potential catalytic residues (Glu<sup>322</sup>, Glu<sup>338</sup> and Glu<sup>366</sup>) have been proposed. Among them, Glu<sup>366</sup> 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, Glu<sup>366</sup> might play a role in structural stability, although not participating directly in the catalytic reaction. By contrast, the residues Glu<sup>322</sup> and Glu<sup>338</sup> are more reasonable catalytic amino acids due to their proximal



**Figure 5:** The HypBA1 active-site. (A) Stereo view of the detailed interaction networks of active-site residues with  $Zn^{2+}$  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.

locations. In the crystal structure of HypBA1-L-Araf complex, Glu<sup>322</sup> is hydrogen bonded to the O1 atom of the sugar and it is in a good position for the general acid/base role. In the absence of L-Araf, Glu<sup>322</sup> is probably hydrogen bonded to His<sup>270</sup>, which is supposed to be protonated. When His<sup>270</sup> turns to interact with the sugar, the proton may remain bound to Glu<sup>322</sup>, making it a good general acid catalyst. On the other side Glu<sup>322</sup> is hydrogen bonded to Tyr<sup>336</sup>, which may also serve as a proton donor. Both His<sup>270</sup> and Tyr<sup>336</sup> are strictly conserved among the GH127 enzymes (Figure 3).

Regarding the nucleophile, Glu<sup>338</sup> binds to the O2 atom but it is nearly 4 Å away from C1, too far and unlikely to undertake this role, although it can be fully ionized by binding to  $Zn^{2+}$ . However, the finding that the mutant E338A only had 0.0013% activity left in a previous study [11] clearly shows that Glu<sup>338</sup> plays an important role in ligand recognition and  $Zn^{2+}$  binding. Very recently, Ito solved the  $Zn^{2+}$  ioncontaining 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  $Cys^{417}$  rather than  $Glu^{338}$  [12]. In our apo structure, the  $Cys^{417}$ -containing loop is shifted away due to the disappearance of all coordinate bonds in the absence of  $Zn^{2+}$  ion (Figure 2C).  $Cys^{417}$  is thus diverted from the attack position, and probably is also protonated, disabling its role as the nucleophile. Therefore,  $Zn^{2+}$  ion is involved in the catalytic reaction through maintaining the proper configuration of active site.

Page 6 of 7

As said by Ito, however, we cannot rule out other possibility of catalytic reaction mechanism, such as the utilization of two carboxylate residues (Glu<sup>322</sup> and Glu<sup>338</sup>) 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 substrateassisted mechanisms, proton transferring network, utilization of noncarboxylate residues and utilization of an exogenous base/nucleophile [38]. Interestingly, apart from Glu<sup>322</sup>, Glu<sup>338</sup>, Cys<sup>340</sup>, Cys<sup>417</sup>and Cys<sup>418</sup>, Tyr<sup>386</sup> is also strictly conserved among several GH127 members (Figure 3). The side chain of Tyr<sup>386</sup> is equally close to the C1 of L-Araf (3.2 Å) as is that of Cys<sup>417</sup>, and it may correspond to the non-carboxylate residue in an alternative mechanism. However, the lack of a base to subtract its proton renders Tyr<sup>386</sup> 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 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 (12ZCZDSY12500) 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).



Chun-Hsiang Huang and Zhen Zhu are the co-authors & have contributed equally.

#### References

- Besra GS, Khoo KH, McNeil MR, Dell A, Morris HR, et al. (1995) A new interpretation of the structure of the mycolyl-arabinogalactan complex of Mycobacterium tuberculosis as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and 1H nuclear magnetic resonance spectroscopy. Biochemistry 34: 4257-4266.
- Tam PM, Lowary TL (2009) Recent advances in mycobacterial cell wall glycan biosynthesis. Curr Opin Chem Biol 13: 618-625.
- Amano Y, Tsubouchi H, Shinohara H, Ogawa M, Matsubayashi Y (2007) Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in Arabidopsis. Proc Natl Acad Sci USA 104: 18333-18338.
- Ohyama K, Shinohara H, Ogawa-Ohnishi M, Matsubayashi Y (2009) A glycopeptide regulating stem cell fate in Arabidopsis thaliana. Nat Chem Biol 5: 578-580.
- Matsubayashi Y (2011) Post-translational modifications in secreted peptide hormones in plants. Plant Cell Physiol 52: 5-13.
- Matsubayashi Y (2012) Recent progress in research on small posttranslationally modified peptide signals in plants. Genes Cells 17: 1-10.
- Mohnen D (2008) Pectin structure and biosynthesis. Curr Opin Plant Biol 11: 266-277.
- Okamoto S, Shinohara H, Mori T, Matsubayashi Y, Kawaguchi M (2013) Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. Nat Commun 4: 2191.
- 9. Kieliszewski MJ, Lamport DTA, Tan L, Cannon MC (2011) Hydroxyproline-Rich Glycoproteins: Form and Function. Annual Plant Reviews: Plant Polysaccharides, Biosynthesis and Bioengineering 41: 321.
- Fujita K, Sakamoto S, Ono Y, Wakao M, Suda Y, et al. (2011) Molecular cloning and characterization of a beta-L-Arabinobiosidase in Bifidobacterium longum that belongs to a novel glycoside hydrolase family. J Biol Chem 286: 5143-5150.
- Fujita K, Takashi Y, Obuchi E, Kitahara K, Suganuma T (2011) Characterization of a novel beta-L-arabinofuranosidase in Bifidobacterium longum: functional elucidation of a DUF1680 protein family member. J Biol Chem 286: 38079-38085.
- 12. Ito K, Saikawa K, Kim S, Fujita K, Ishiwata A, et al. (2014) Crystal structure of glycoside hydrolase family 127  $\beta$ -L-arabinofuranosidase from Bifidobacterium longum. Biochem Biophys Res Commun 447: 32-37.
- Zhu Z, He M, Huang CH, Ko TP, Zeng YF, et al. (2014) Crystallization and preliminary X-ray diffraction analysis of a novel β-L-arabinofuranosidase (HypBA1) from Bifidobacterium longum. Biochem Biophys Res Commun 447: 32-37.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 276: 307-326.
- Brunger AT (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. Acta Crystallogr D Biol Crystallogr 49: 24-36.
- 16. Hayes D, Laue T, Philo J (1995) Program Sednterp: sedimentation interpretation program, Durham. NH: University of New Hampshire, UK.
- Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78: 1606-1619.
- Guo RT, Kuo CJ, Chou CC, Ko TP, Shr HL, et al. (2004) Crystal structure of octaprenyl pyrophosphate synthase from hyperthermophilic Thermotoga maritima and mechanism of product chain length determination. J Biol Chem 279: 4903-4912.
- Sun HY, Ko TP, Kuo CJ, Guo RT, Chou CC, et al. (2005) Homodimeric hexaprenyl pyrophosphate synthase from the thermoacidophilic crenarchaeon Sulfolobus solfataricus displays asymmetric subunit structures. J Bacteriol 187: 8137-8148.
- Cheng YS, Ko TP, Wu TH, Ma Y, Huang CH, et al. (2011) Crystal structure and substrate-binding mode of cellulase 12A from Thermotoga maritime. Proteins

79: 1193-1204.

- Ren F, Ko TP, Feng X, Huang CH, Chan HC, et al. (2012) Insights into the mechanism of the antibiotic-synthesizing enzyme MoeO5 from crystal structures of different complexes. Angew Chem Int Ed Engl 51: 4157-4160.
- 22. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126-2132.
- 24. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc 2: 2728-2733.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658-674.
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67: 235-242.
- DeLano WL (2001) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
- Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545-W549.
- Fujimoto Z, Jackson A, Michikawa M, Maehara T, Momma M, et al. (2013) The structure of a Streptomyces avermitilis alpha-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. J Biol Chem 288: 12376-12385.
- Guce AI, Clark NE, Rogich JJ, Garman SC (2011) The molecular basis of pharmacological chaperoning in human alpha-galactosidase. Chem Biol 18: 1521-1526.
- Ariza A, Eklof JM, Spadiut O, Offen WA, Roberts SM, et al. (2011) Structure and activity of Paenibacillus polymyxa xyloglucanase from glycoside hydrolase family 44. J Biol Chem 286: 33890-33900.
- 32. Czjzek M, Ben David A, Bravman T, Shoham G, Henrissat B, et al. (2005) Enzyme-substrate complex structures of a GH39 beta-xylosidase from Geobacillus stearothermophilus. J Mol Biol 353: 838-846.
- Zeqiraj E, Filippi BM, Deak M, Alessi DR, van Aalten DM (2009) Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation. Science 326: 1707-1711.
- Qiu C, Tarrant MK, Choi SH, Sathyamurthy A, Bose R, et al. (2008) Mechanism of activation and inhibition of the HER4/ErbB4 kinase. Structure 16: 460-467.
- Chan HC, Huang YT, Lyu SY, Huang CJ, Li YS, et al. (2011) Regioselective deacetylation based on teicoplanin-complexed Orf2\* crystal structures. Mol Biosyst 7: 1224-1231.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774-797.
- Yu J, Zhou Y, Tanaka I, Yao M (2010) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics 26: 46-52.
- Vuong TV, Wilson DB (2010) Glycoside hydrolases: catalytic base/nucleophile diversity. Biotechnol Bioeng 107: 195-205.