

ARTICLE

<https://doi.org/10.1038/s42003-019-0744-4>

OPEN

Structures of β -glycosidase LXYL-P1-2 reveals the product binding state of GH3 family and a specific pocket for Taxol recognition

Lin Yang^{1,3}, Tian-Jiao Chen^{2,3}, Fen Wang², Li Li², Wen-Bo Yu², Yi-Kang Si², Jing-Jing Chen², Wan-Cang Liu², Ping Zhu^{2*} & Weimin Gong^{1*}

LXYL-P1-2 is one of the few xylosidases that efficiently catalyze the reaction from 7- β -xylosyl-10-deacetyltaxol (XDT) to 10-deacetyltaxol (DT), and is a potential enzyme used in Taxol industrial production. Here we report the crystal structure of LXYL-P1-2 and its XDT binding complex. These structures reveal an enzyme/product complex with the sugar conformation different from the enzyme/substrate complex reported previously in GH3 enzymes, even in the whole glycohydrolases family. In addition, the DT binding pocket is identified as the structural basis for the substrate specificity. Further structure analysis reveals common features in LXYL-P1-2 and Taxol binding protein tubulin, which might provide useful information for designing new Taxol carrier proteins for drug delivery.

¹National Research Laboratory for Physical Sciences in Microscales, University of Science and Technology of China, 230026 Hefei, Anhui, China.

²State Key Laboratory of Bioactive Substance and Function of Natural Medicines and Key Laboratory of Biosynthesis of Natural Products of the State Health Commission, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, 1 Xian Nong Tan Street, 100050 Beijing, China. ³These authors contributed equally: Lin Yang, Tian-Jiao Chen. *email: zhuping@imm.ac.cn; wgong@ustc.edu.cn

Taxol (generic name: paclitaxel), a rare natural product mainly generated by yew bark, is the well-known blockbuster anticancer drug. It promotes tubulin assembly into microtubules and prevents their disassembly¹. However, the natural level of Taxol is extremely low², while the content of 7- β -xylosyl-10-deacetyltaxol (XDT) can be up to 25 times of Taxol³. XDT is often regarded as the waste during Taxol extraction process, causing both resource loss and potential environmental pollution. Compared with Taxol, XDT lacks the C10 hydroxyl group but harbors an additional β -xylosyl group at the C7 position. If the xylosyl group is removed, the resultant 10-deacetyltaxol (DT) can be used as a precursor for Taxol preparation. Therefore, the β -xylosidase catalyzing the removal of xylosyl from XDT (Fig. 1a) would play a prominent role in reducing both of the resource waste and pollution to environment. However, a lot of commercially available β -xylosidases have been demonstrated to possess no activity on releasing the xylosyl residue from XDT⁴. Recently, we identified two bifunctional β -xylosidase/glucosidase (named as LXYL-P1-1 and LXYL-P1-2)

from *Lentinula edodes*, which belong to GH3 family and can efficiently convert XDT into DT⁵. By combining LXYL-P1-2 and an engineered acetyltransferase, we have constructed an in vitro one-pot reaction system for converting XDT into Taxol⁶.

The catalytic specificity and higher efficiency of LXYL-P1-2 prompted further investigation of its structure-function relationship. Here we present the crystal structures of LXYL-P1-2 and its complex with XDT. The binding mode of xylose group sheds lights on the catalytic mechanism for GH3 enzymes. DT binding pocket elucidates the structural basis of substrate specificity. Structural comparison of LXYL-P1-2 and tubulin suggests a possible common feature for designing Taxol binding protein.

Results

Structures of LXYL-P1-2 in substrate free form. Highly glycosylated proteins, such as LXYL-P1-2, are greatly difficult to be crystallized. To improve formation of good quality crystals, LXYL-P1-2 was endoglycosidase-treated before crystallization⁵. The

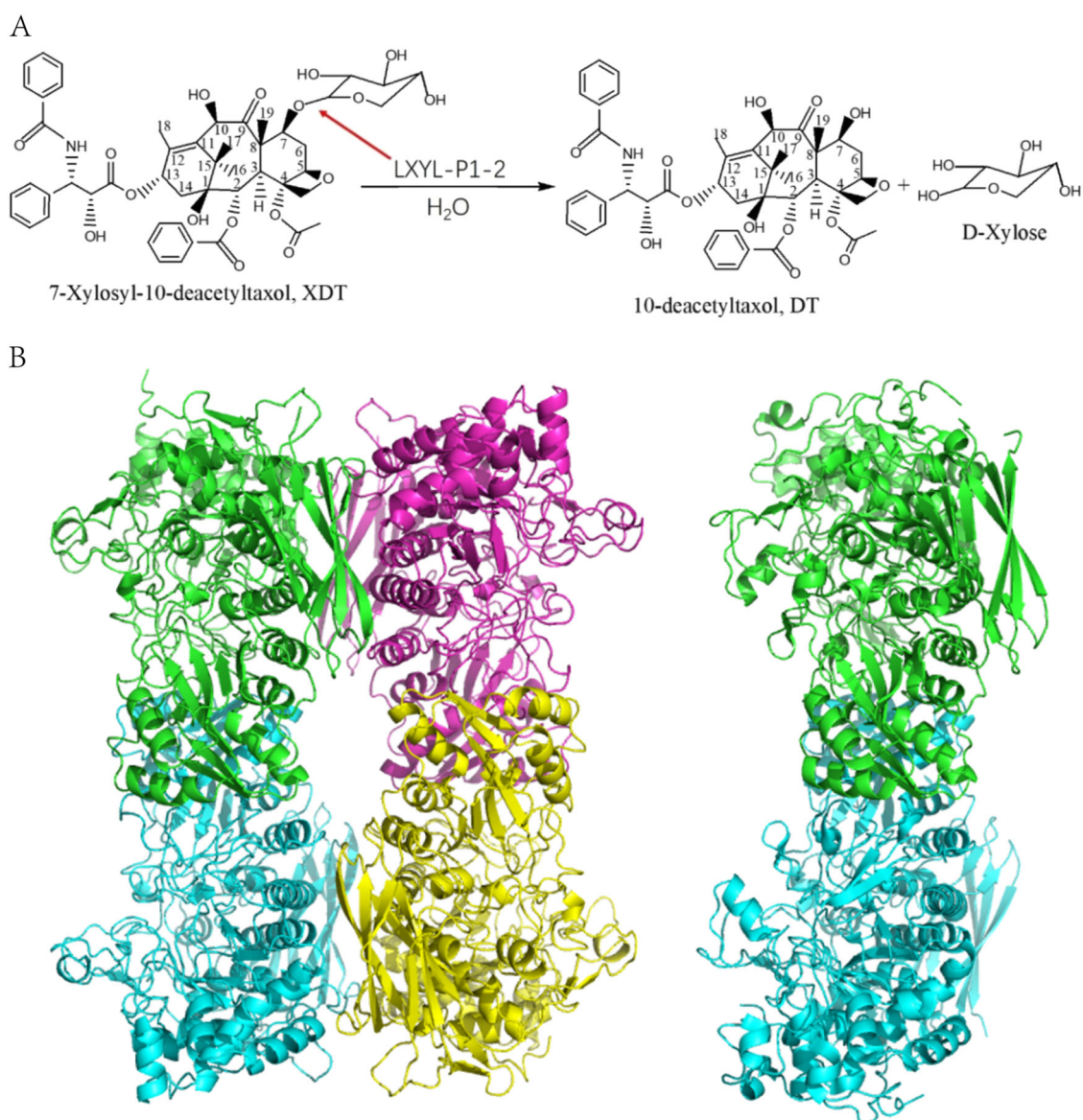


Fig. 1 a Chemical transformation of XDT to DT with LXYL-P1-2 catalyzed. **b** Tetramer structure of LXYL-P1-2. The four monomers are colored in green, cyan, magenta, and yellow, respectively. The dimer structure of AaBGL1 (PDB code 4IIB). The two monomers are colored in green and cyan and made the same orientation as the dimer in LXYL-P1-2.

crystallographic statistics for data collection and structure refinement are summarized in Table 1. The first 43 residues were missing in the electron density map, partially supporting the existence of the KEX2⁷ cleavage site (Ile³²Phe³³Arg³⁴Arg³⁵). The first amino acid of the mature protein was then verified to be Asp³⁶ by N-terminal sequencing.

LXYL-P1-2 exists as a 222 symmetric tetramer in the asymmetric unit (Fig. 1b) (PDB code 6JBS), consistent with the molecular weight measured by Gel filtration experiment. The structure of the four monomers are essentially same without any remarkable difference. Each monomer comprises three domains as some GH3 members. Domain 1 folds into a TIM barrel-like structure and contains the residues from Asp³⁶ to Gly³⁶⁵. Domain 2 (residue 398–600) is an α/β sandwich, in which five parallel β -strands and one antiparallel β -strand are sandwiched by five α -helices. Domain 1 and domain 2 is connected by linker 1 (residue 366–397). Domain 3 (residue 664–803) is connected to domain 2 by linker 2 (residue 601–663) and has a fibronectin type III (FnIII) fold. Seven Asn residues (81, 272, 342, 385, 457, 576, and 635) are found to link with different types of oligosaccharide even after treated by endoglycosidase H (Table 2).

Structure of E529Q mutant co-crystallized with XDT. To obtain the substrate binding structure, mutant E529Q was co-crystallized with XDT (PDB code 6KJ0). There are two monomers in the asymmetric unit. The electron density clearly shows the existence of xylose and DT (Fig. 2a). The xylose adopts a pyranose configuration. It was surprised to find that the glycosidic bond between xylose and DT was broken in both of the monomers. Although the E529Q mutant did not show catalytic activity in the standard enzymatic assay, it may still have weak activity to digest the glycosidic bond of XDT during the crystallization time of one week. Thus, this structure should be regarded as the enzyme/product complex.

Sequence alignment of LXYL-P1-2 indicates that Asp³⁰⁰ and Glu⁵²⁹ might act as the catalytic nucleophile and the acid/base residues, respectively. In LXYL-P1-2 structure, the distance of these two residues is 5.5 Å, consistent with the proposed retaining catalytic mechanism in GH3 family. In the substrate free enzyme, Glu⁵²⁹ forms hydrogen bond to the side-chain of Arg²¹⁸ and the main-chain NH of Ser⁴⁶⁶. In the complex, the amide group of Gln⁵²⁹ rotates about 40 degree to form hydrogen bonds to the side-chain OH of Ser⁴⁶⁶ and the O7 atom of DT. This confirms that Glu⁵²⁹ plays as acid/base to attack the glycosidic bond. Other residues remain the same conformations as the free state. The side-chains of Asp¹⁰⁹, Arg¹⁷⁴, Lys²⁰⁷, His²⁰⁸, and Asp³⁰⁰ stabilize the xylose by forming hydrogen bonds (Fig. 2d).

Compared with the free enzyme, the only prominent conformational changes in the overall structure upon XDT binding is the movement of loop Ile²²²-Gln²²⁹ with the longest distance of 2.5 Å (Fig. 2b, c). This movement suggests that LXYL-P1-2 possesses an open conformation in free enzyme and a closed conformation in substrate binding state. DT molecule binds at a pocket formed by several loops: residues 220–232 and 324–328 from domain 1, residues 379–383 which links domain 1 and 2, and residues 446–450 and 529–530 from domain 2 (Fig. 3). Therefore, domain 1 and domain 2 together make the DT binding site. Besides the hydrogen bond between O7 of DT and Glu(Gln)⁵²⁹, no other hydrogen bonds or static electric interactions were observed between LXYL-P1-2 and DT. The hydrophobic environment, contributed by Leu²²⁰, Ile²²², Ile²²⁴, Val²²⁷, Ile²³², Phe³²⁴, Leu³²⁵, Ala³²⁷, Leu³²⁸, Ile³⁷⁹, Thr³⁸³, and Ala⁴⁴⁷ form the hydrophobic wall of DT binding pocket. The modified oligosaccharide chains seem not contact with the substrate, which

Table 1 Statistics of data-collection and refinement.

Data sets	Native(6JBS)	E529Q-XDT(6KJ0)
Diffraction data		
Space group	<i>P</i> _{4₃} <i>2</i> ₁ <i>2</i>	<i>C</i> 222 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	131.9, 131.9, 385.9	79.9, 182.2, 241.5
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution range (Å)	50–2.4 (2.49–2.40)	50–2.27 (2.30–2.27)
Number of unique reflections	133487	76271
Data completeness (%)	99.9 (99.1)	93.5 (94.1)
Redundancy	12.8 (7.1)	8.6 (8.4)
$\langle I \rangle / \langle \sigma(I) \rangle$	23.6 (2.9)	27.1 (4.0)
<i>R</i> _{merge} ^a	0.155 (0.667)	0.122 (0.548)
Refinement		
<i>R</i> -factor/ <i>R</i> _{free} ^b	0.184/0.251	0.192/0.251
Number of reflections used	126287	72397
r.m.s.d. bond length (Å)	0.009	0.017
r.m.s.d. bond angles (°)	1.527	2.040
Mean <i>B</i> factor (Å ²)		
Protein main-chain atoms	33.9	33.8
Protein side-chain atoms	35.4	34.5
Water molecules	32.9	33.8
Glycan	51.8	55.6
Tris	42.1	–
DT	–	55.1
Xylose	–	34.8
No. of atoms		
Protein	22984	11433
Water molecules	1544	683
Glycan	952	493
Tris	32	–
DT	–	118
Xylose	–	20

The free *R* factor was calculated using 5% of reflections omitted from the refinement
^a*R*_{merge} = $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the main value of $I(hkl)$
^b*R*-factor = $\sum |F_{obs}| - |F_{calc}| / \sum |F_{obs}|$, where *F*_{obs} and *F*_{calc} are observed and calculated structure factors

Table 2 Glycan modification of LXYL-P1-2.

Residue	Glycan structure
Asn81	GlcNAc- β 1-4-GlcNAc- α 1-4-Man
Asn272	GlcNAc- β 1-4-GlcNAc- β 1-4-Man- α 1-4-Man
Asn342	GlcNAc
Asn385	GlcNAc
Asn457	GlcNAc- β 1-4-GlcNAc- α 1-4-Man- α 1-3-Man(- β 1-6-Man)- α 1-2-Man
Asn576	GlcNAc
Asn635	GlcNAc

indicates that the glycosylation does not contribute to the enzymatic activity.

Site-directed mutations of the XDT binding region. According to the enzyme-substrate complex structure, a number of residues at xylose and DT binding site were subjected to site-directed mutations to inspect their influence to the enzyme activities. Ala-scanning mutations were carried out and the activities of the mutants are summarized in Fig. 2e. Unsurprisingly, the mutations of conserved catalytic residues all showed an inactive or very low active. Deceased activities were observed on F324A, L328A, and L325A mutants, which locate at the recognition area for benzoate group of DT. Mutation S228A destroyed the potential hydrogen bond between OG-Ser²²⁸ and N-Tyr²²¹ (3.0 Å in substrate free structure, 4.5 Å in E529Q), making it easier for loop Ile²²²-Gln²²⁹ to move. Mutation S449A might provide more hydrophobicity in DT binding pocket. It is surprised that mutation S91A, which hydrogen bonds to Asp¹⁰⁹, increased the enzyme activity, too.

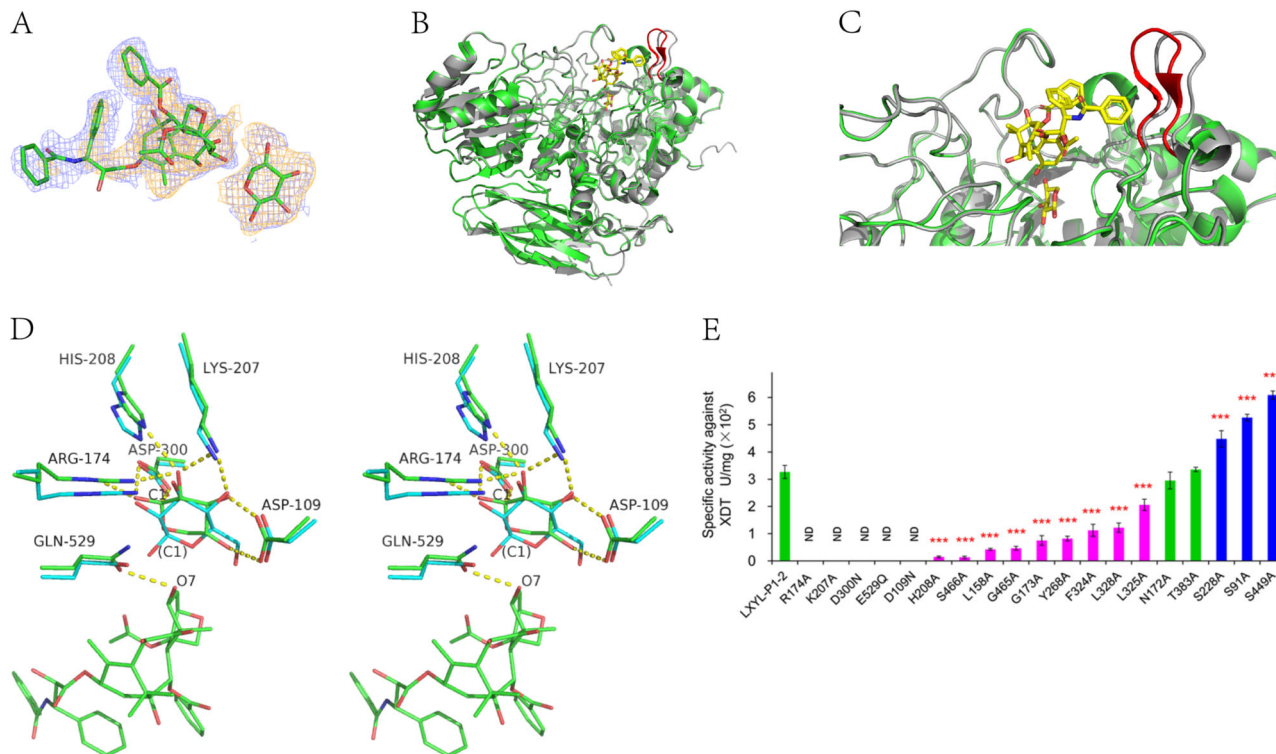


Fig. 2 **a** The 2Fo-Fc electron density map (blue, 1.0 σ) and Fo-Fc omit map (orange, 3.0 σ) for xylose and DT. **b** Comparison of the overall structure of LXYL-P1-2 in substrate free form (gray) and XDT binding form (green). The moving loop Ile²²²-Gln²²⁹ is shown in red. DT and xylose are shown in a stick model. **c** A big view of comparison of the detailed structure of LXYL-P1-2 in substrate free form (gray) and XDT binding form (green). The moving loop Ile²²²-Gln²²⁹ is shown in red. DT and xylose are shown in a stick model. **d** Stereo view of the active site comparison of LXYL-P1-2 (green) and *HjCel3A* (PDB code 3ZYZ, cyan). The residues from LXYL-P1-2, atom C1 of xylose and atom O7 of DT are labeled. The C1 atom of glucose bound in *HjCel3A* is also labeled in brackets. **e** The specific activities of LXYL-P1-2 and its site-directed mutants on XDT. The data represent the means \pm s.d., $n = 3$. * $P < 0.05$ versus LXYL-P1-2 (Control), ** $P < 0.01$ versus LXYL-P1-2 (Control), *** $P < 0.001$ versus LXYL-P1-2 (Control), One-way analysis of variance was used for data analysis using SPSS 17.0.

This might indicate that Asp¹⁰⁹ that interacts with xylose needs more flexibility during catalysis.

Discussion

Comparison with other GH3 enzymes shows that the crystal structures of LXYL-P1-2 provide the detailed information of xylosidase with the activity to hydrolyze 7- β -xylosyl-10-deacetylaxol for the first time to our knowledge. Other structures similar to LXYL-P1-2 were found by using Dali server⁸. As expected, the overall fold of LXYL-P1-2 is resembled to some three-domain GH3 members. The RMSD of main-chain atoms is 1.3 Å when superimposing LXYL-P1-2 to the most similar structure *TnBgl3B*⁹. In the structure of LXYL-P1-2, eight cis-peptide bonds are found in Asp¹⁰⁹-Gly¹¹⁰, Ala¹⁵⁹-Pro¹⁶⁰, Gly¹⁶⁴-Pro¹⁶⁵, Lys²⁰⁷-His²⁰⁸, Trp²⁰⁹-Ile²¹⁰, Met³¹⁹-Pro³²⁰, Ala³⁸¹-Pro³⁸², and Leu⁴¹⁹-Pro⁴²⁰. Six of these cis-peptide bonds are conserved in many GH3 members, except for Ala³⁸¹-Pro³⁸² that makes a sharp turn involved in DT binding pocket. In addition, two intra-molecular disulfide bonds are formed by Cys²⁶⁶-Cys²⁷⁷ and Cys⁴⁴⁴-Cys⁴⁵¹. These disulfide bonds are also conserved in most of GH3 members. Structure comparison also shows that the conformations of the residues around the sugar binding site, including the catalytic nucleophile Asp³⁰⁰ and the acid/base Glu⁵²⁹ are strictly conserved (Fig. 2d). This suggests LXYL-P1-2 share the common hydrolytic mechanism in GH3 family.

Interestingly, the xylose ring orientation rotated by about 60 degrees compared to the glucoses found in other GH3 enzyme structures as in *HjCel3A*¹⁰. As shown in Fig. 2d, the OH-1,-2,-3

groups of xylose is corresponded to the OH-2,-3,-4 of glucose. The OH-1 of glucose points outwards of the active site, where the +1 group of substrate could be linked. The glucose soaked in crystals mimics the substrate binding state. In LXYL-P1-2, the xylose is generated from the hydrolysis of XDT and represent the product binding state, in which the OH-1 group is away from DT, with the distance between C1 of xylose and O7 of DT being 4.3 Å. If the xylose is orientated as the in *HjCel3*, the distance between C1 and O7 would be only 3.0 Å, indicating a good position to form the glycosidic bond. Therefore, the glucose soaked in *HjCel3* crystals might mimic the substrate binding state. In LXYL-P1-2, the xylose is generated from the hydrolysis of XDT and might represent the product binding state. Both in the substrate and the product states, although the C1 atom of xylose in LXYL-p1-2 and the C1 atom of glucose in *HjCel3* are at different positions, both of them are close to the nucleophile residue (Asp300 in LXYL-P1-2 and Asp236 in *HjCel3*, with the C1-OD1 distances of 2.7 Å and 2.8 Å).

In the proposed retaining mechanism of glycosidase, a covalent intermediate complex is expected to be formed by the sugar group of substrate and the nucleophile residue. The structures of xylose and glucose described above might depict the two states before and after the covalent intermediate stage. This might be the first report elucidating the rotation of the sugar ring from substrate binding to product forming states.

Besides comparison above, the tetramer formation attracts our attention. The catalytic pocket of each monomer faces to the outside of tetramer. The tetramer formation seems not influence on the active pocket, but contributes to the highly thermal



Fig. 3 Sequence alignment. Sequence alignment was done on the online sever of T-coffee²³ and ESPrnt sever was used for result display²⁴. Cysteine in the protein are shown by green figure, and the DT specific residues are shown by red points. XDT binding and catalytic domain are highlight in yellow.

stability of LXYL-P1-2. Two interfaces are defined in the LXYL-P1-2 tetramer. On interface A, His¹¹⁸, Tyr³⁹⁰, Asn³⁹², Arg³⁹⁴, Arg⁴²⁶, Tyr⁴³⁶, Glu⁴⁷⁹, Gln⁴⁸², Asn⁴⁸⁹, and Glu⁴⁹¹ form a number of hydrogen bonds between the two monomers, while Y^{473-1486*}, L⁴⁷⁸.Y^{436*}, L⁴⁵⁶.V^{438*} (*indicates the residue from the neighboring monomer) form hydrophobic pairs. On interface B, Arg⁶⁷⁸.Tyr^{363*}, Asp⁶⁷⁵-Lys^{666*} form hydrogen bonds. In addition, Van der Waals interactions of Ser⁶².Trp^{679*}, Thr³⁶⁸-Phe^{710*}, and Asn³⁶⁹-Val⁷⁵⁷ also contribute to interface too. The interface A seems conserved as in AaBGL (Fig. 1b). However, the modified oligosaccharides are involved in dimer formation¹¹. In LXYL-P1-2,

there is no glycan modification close to this interface. The interface B was never reported in GH3 family. *KmBglI*¹² and *AoBglI*¹³ were reported to be a tetramer, but their dimer-dimer interfaces are different from that in LXYL-P1-2.

Compared to other GH3 enzymes, the sequence and structural variations mainly come from the loops, especially those forming the substrate binding pocket (Fig. 3). Loops of residues 220-232, 324-328, 379-383, 446-450, and 529-530 form the DT binding pocket. As shown in the sequence alignment, the first four loops are conserved in LXYL-P1-1 and LXYL-P1-2 but are quite different from other GH3 enzymes. In fact, when superimposing

other GH3 enzymes structures with LXYL-P1-2, they may have residues occupy the position of DT binding, such as Val²⁹⁸ and Phe²⁹ in *TnBgl3B*⁹, Trp³⁷, Phe²⁶⁰, and Trp⁴⁴³ in *HjCel3A*¹⁰, Trp²⁷⁴, Leu²⁹⁵, and Tyr⁵¹⁰ in *KmBglI*¹². The DT specificity results from the loop sequences, with the residues contributing hydrophobic side-chains towards the binding DT molecule. This suggests the pocket size, shape and hydrophobic environment are critical to DT recognition. As shown in Fig. 2e, the mutants with large hydrophobic side-chain removed decrease the enzymatic activity. In contrast, mutations of the surrounding residues S91A and S449A, which increase the hydrophobicity, could improve the catalytic ability.

To confirm the importance of the DT-binding loops, we purified the recombinant *TnBgl3B* that shares the same sugar binding site but different DT-binding loops, and tested its activity on PNP-Glc, PNP-Xyl, and XDT, respectively. The results showed that although *TnBgl3B* exhibited considerable activities on PNP-Glc and PNP-Xyl, the activity on XDT was undetectable (Table 3). In order to find more candidate enzymes with XDT xylosidase activity, we searched the whole genome in EXPASY blast server [<https://web.expasy.org/blast/>]. No protein except LXYL-P1-1 and LXYL-P1-2, are found with the similar DT-preferred loops. Therefore, LXYL-P1-1 and LXYL-P1-2 seem to be the only available enzymes that could be applied in DT production to date. The activity-increased mutants show the potential of successful engineering of better enzymes in the future.

Another key point of the research is the binding site of Taxol in LXYL-P1-2. Due to the poor solubility in water, Taxol is usually formulated in a mixture of cremophor EL and dehydrated ethanol, which however may have severe side effects on patients (Ma and Mumper¹⁴). Therefore, the specific carrier proteins are needed for developing Taxol delivery system. To date, Nab-paclitaxel (Abraxane[®], approved by FDA in 2005) has been the first FDA approved taxoid formulation based on albumin nano-delivery systems, and a number of novel Taxol nano-particle formulations are in clinical trials¹⁴. The structure of LXYL-P1-2 bound with XDT may provide useful information for designing new Taxol binding proteins.

In Protein Data Bank, Taxol is only found in tubulin/Taxol complex with the highest resolution of 3.5 Å. Our study shows the first high resolution structure of Taxol analogs bound in proteins. In order to search the possible common recognition features for DT or Taxol binding, the complexes of LXYL-P1-2/XDT and tubulin/Taxol¹⁵ are compared. Although the overall structures of LXYL-P1-2 and tubulin are totally different, the conformations of the DT or Taxol main body are superimposable (Fig. 4a, c) and two similar binding pockets could be identified, but with the binding groups are swapped.

As shown in Fig. 4b, loop of residues 220–232 in LXYL-P1-2 forms the first pocket that recognizes the benzoylamino of DT. The non-polar amino acid residues, Ile²²², Ile²²⁴, Ile²³², and

Val²²⁷, form a semi-enclosed hydrophobic pocket, in which the benzoylamino is half surrounded, while the side-chain of Gln²²⁹ and the main-chain O of Val²²⁷ form the hydrophilic area at the bottom of the pocket. In the similar way, the pocket for benzoate binding in tubulin is also a semi-enclosed hydrophobic pocket formed by Leu²¹⁷, Leu²¹⁹, Leu²³⁰, and Leu²⁷⁵, while Asp²²⁶ and His²²⁹ form the hydrophilic bottom.

Loop324–328 of LXYL-P1-2 forms the second pocket, which is bound with the benzoate of DT. This pocket is formed by the side chain benzene of Phe³²⁴, the side chain methyl group of Ala³²⁷, the side chain of Leu³²⁵, the CG2 atom of Thr³⁸³. Tubulin has a similar pocket, which is formed by Phe²⁷², Ala²³³, Val²³, and Pro³⁶⁰. Instead of binding the benzoate in LXYL-P1-2, however, this binding pocket is for the phenylpropanoyl group binding in tubulin.

Further structural analysis indicates that Ile²²², Ile²²⁴, Ile²³², Val²²⁷, and Gln²²⁹ in LXYL-P1-2 have the similar spatial distribution as Leu²¹⁷, Leu²¹⁹, Leu²³⁰, Leu²⁷⁵, and Asp²²⁶ in tubulin (Fig. 4c). It is interesting that the three benzene rings of DT and Taxol are also in the same spatial position when superposed (Fig. 4c), which partially supports the conserved spatial distribution of binding pockets in Taxol binding. Besides the structural information, the results of our enzyme catalytic experiments demonstrate that Phe³²⁴ and Leu³²⁵ in the benzoate recognition region is critical for substrate recognition (Fig. 2d). Therefore, this spatial distribution of hydrophobic residues would make an interface with two pockets specific for Taxol binding. This is the first Taxol analog binding structure except tubulin and might be the only enzyme to catalyze XDT, which may provide useful information for Taxol analogs design.

Methods

Materials and strains. The plasmid *pPIC3.5K-lxyl-p1-2* was cloned in our lab^{5, 16}. The *TnBgl3B* gene (GenBank: ABI29899.1) was synthesized by SynBio Research Platform at Tianjin University (Tianjin, China). Phusion polymerase, restriction enzymes, and T4 ligase were purchased from New England Biolabs (Ipswich, MA). *Escherichia coli* Transeetta (DE3) competent cells and plasmids were purchased from TransGen Biotech (Beijing, China). The pET-28a plasmid was purchased from Novagen (Malaysia). XDT and DT (HPLC purity >99%) were purified in our laboratory. All other chemicals were analytical grade unless otherwise indicated.

Construction of active-site mutants of LXYL-P1-2 and *TnBgl3B* recombinant strain. The plasmid *pPIC3.5K-lxyl-p1-2* was used as DNA template. The L-alanine scanning mutations and other active-site mutations were all obtained by means of the site-directed mutagenesis technique with Phusion High-Fidelity DNA Polymerase (NEB) by using whole-plasmid amplification PCR. All products were sequenced to ensure that no base change other than designed. The plasmids were extracted, linearized with *Sac* I and transformed into the *Pichia pastoris* GS115 competent cells for expression.

The full-length *TnBgl3B* gene was amplified by using forward primer (F: 5'-AA GGATCCATGGAAAAGGTTAACGATC-3') with *Bam*HI site (underlined) and reverse primer (R: 5'-TAGCGCGCGCTTAAGGCTTGAATCTTC-3') with *Not*I site (underlined). The PCR products were digested by *Bam*HI and *Not*I for directional ligation into vector pET-28a. After ligation, the construct was sequenced and transformed into *E. coli* Transeetta (DE3) competent cells for expression.

Protein expression, purification and activity assay. The heterologous expression and deglycosylation of LXYL-P1-2, as well as the mutants were same to the previous article^{5, 16}. *E. coli* cells with pET-28a-*TnBgl3B* recombinant plasmid were grown overnight at 37 °C and 200 r.p.m. in 10 ml Luria-Bertani (LB) medium containing kanamycin (50 µg/ml) in a shaking flask. The overnight culture was suspended in 100 ml fresh LB medium at a final concentration of 1% (v/v), and grown at 37 °C and 200 r.p.m. for 2–3 h until OD₆₀₀ reached 0.8. Then, isopropyl-β-D-thiogalactopyranoside (IPTG) was added at a final concentration of 1 mM and the cell culture was incubated for an additional 20 h at 24 °C and 200 r.p.m. The protein was purified by Ni Sepharose 6 Fast Flow resin and Agilent ZORBAX GF-450 gel-filtration column. The enzyme activities of LXYL-P1-2 were tested as reported previously⁵.

Crystallization, data collection, and structure determination. LXYL-P1-2 protein was concentrated to 10 mg/ml for crystallization. Both native crystal and complex crystal were grown at 16 °C with the hanging drop vapor diffusion

Table 3 Specific activity of *TnBgl3B* and LXYL-P1-2 against PNP-Xyl, PNP-Glc and XDT.

	PNP-Xyl (U/mg × 10 ⁴)	PNP-Glc (U/mg × 10 ⁴)	XDT (U/mg × 10 ²)
LXYL-P1-2	7.19 ± 0.87	15.59 ± 2.09	3.27 ± 0.24
<i>TnBgl3B</i>	1.60 ± 0.03***	11.86 ± 0.70***	ND

Values represent mean ± s.d. of triplicates from a representative experiment (n = 3 experiment replicates)

ND Not detected

*P < 0.05 vs LXYL-P1-2

**P < 0.01 vs LXYL-P1-2

***P < 0.001 versus LXYL-P1-2 (Control). One-way analysis of variance was used for data analysis using SPSS 17.0

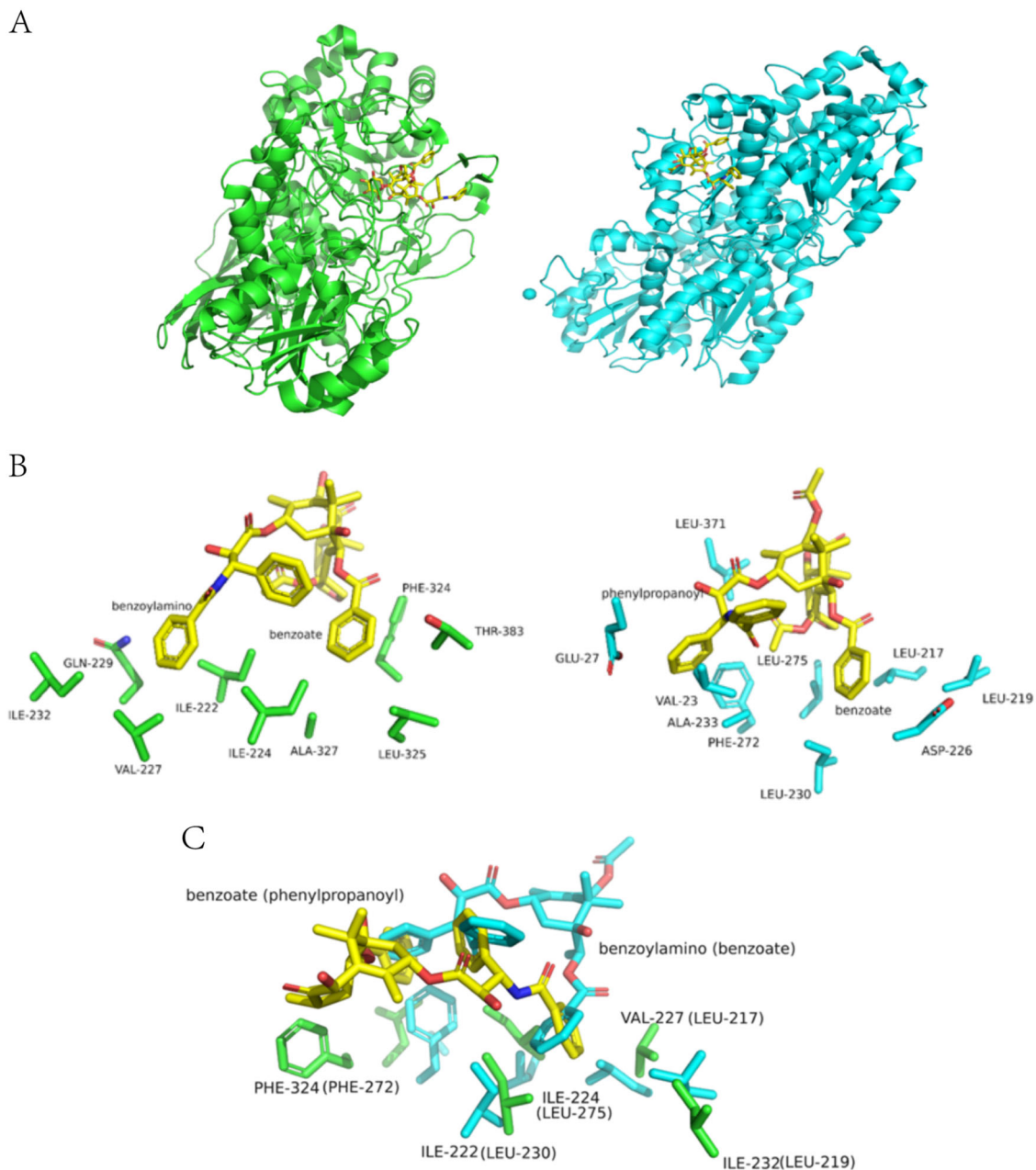


Fig. 4 a Structure comparison of the overall structure of LXYL-P1-2 (PDB code 6KJ0, green) and tubulin (PDB code 1JFF, cyan). DT and Taxol are shown as a yellow stick model. **b** Structure comparison of DT/Taxol two benzene rings recognition pocket of LXYL-P1-2 (PDB code 6KJ0, green) and tubulin (PDB code 1JFF, cyan). DT and Taxol are shown as a yellow stick model. The residues involved in DT/Taxol binding are shown as stick. **c** Structure superposition of DT/Taxol binding pocket of LXYL-P1-2 (green) and tubulin (cyan). DT is shown as a yellow stick model. Taxol is shown as model color stick. The residues involved in DT/Taxol binding are shown as stick. Label of tubulin are shown in parentheses.

method. For complex crystallization, E529Q protein was mixed with XDT before setting up. Native crystal and complex crystal grew in the solution contain 13% PEG3350, 0.1 M Tris-HCl pH8.5, 0.2% MgCl₂ and 15% PEG3350, 0.1 M Hepes pH 7.5, 0.2% MgCl₂, respectively. Diffraction data were collected in BL17U of the Shanghai Synchrotron Radiation Facility (SSRF)¹⁷. Data were processed using HKL2000¹⁸. The native crystal diffracted to 2.4 Å (PDB code 6JBS), while the E529Q-XDT complex crystal diffracted to 2.27 Å (PDB code 6KJ0).

The native structure was solved by Phaser¹⁹ in CCP4 suit²⁰ with *TnBgl3B* structure (PDB code 2 × 40) as the searching model. The complex structure was solved using native structure as the searching model. Refmac5 was used for structure refinement²¹. Coot was used for model building²². All the structure figures were prepared using PyMol (<https://pymol.org>).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The coordinates of the crystal structures are deposited in Protein Data bank with the entrance codes of 6JBS and 6KJ0.

Received: 23 September 2019; Accepted: 16 December 2019;
Published online: 10 January 2020

References

- Schiff, P. B., Fant, J. & Horwitz, S. B. Promotion of microtubule assembly in vitro by taxol. *Nature* **277**, 665 (1979).
- Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93**, 2325–2327 (1971).

- Chattopadhyay, S. K., Kumar, S., Madhusudanan, K. P., Sharma, R. P. A process for the production of taxol. Vol. EP patent 0905130 B1 (2002).
- Hanson, R. L. et al. Enzymic hydrolysis of 7-xylosyltaxanes by xylosidase from *Moraxella* sp. *Biotechnol. Appl. Biochem.* **26**, 153–158 (1997).
- Cheng, H.-L. et al. Cloning and characterization of the glycoside hydrolases that remove xylosyl groups from 7- β -xylosyl-10-deacetyltaxol and its analogues. *Mol. Cell. Proteom.* **12**, 2236–2248 (2013).
- Li, B. J. et al. Improving 10-deacetylbaocatin III-10-beta-O-acetyltransferase catalytic fitness for Taxol production. *Nat. Commun.* **8**, 15544 (2017).
- Rozañ, L., Krysan, D. J., Rockwell, N. C. & Fuller, R. S. Plasticity of extended subsites facilitates divergent substrate recognition by Kex2 and furin. *J. Biol. Chem.* **279**, 35656–35663 (2004).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Pozzo, T., Pasten, J. L., Karlsson, E. N. & Logan, D. T. Structural and functional analyses of beta-glucosidase 3B from *Thermotoga neopolitana*: a thermostable three-domain representative of glycoside hydrolase 3. *J. Mol. Biol.* **397**, 724–739 (2010).
- Karkehabadi, S. et al. Biochemical characterization and crystal structures of a fungal family 3 beta-glucosidase, Cel3A from *Hypocrea jecorina*. *J. Biol. Chem.* **289**, 31624–31637 (2014).
- Suzuki, K. et al. Crystal structures of glycoside hydrolase family 3 beta-glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* **452**, 211–221 (2013).
- Yoshida, E. et al. Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 beta-glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39–49 (2010).
- Agirre, J. et al. Three-dimensional structures of two heavily N-glycosylated *Aspergillus* sp. family GH3 beta-D-glucosidases. *Acta Crystallogr. Sect. D* **72**, 254–265 (2016).
- Ma, P. & Mumper, R. J. Paclitaxel nano-delivery systems: a comprehensive review. *J. Nanomed. Nanotechnol.* **4**, 1000164 (2013).
- Löwe, J., Li, H., Downing, K. & Nogales, E. Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).
- Zhu, P. et al (2015) Glycosyl hydrolase with beta-xylosidase and beta-glucosidase activities and uses thereof. Google Patents.
- Wang, Q.-S. et al. Upgrade of macromolecular crystallography beamline BL17U1 at SSRF. *Nucl. Sci. Tech.* **29**, 68 (2018).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Cryst. D.* **67**, 235–242 (2011).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Cryst. D.* **60**, 2184–2195 (2004).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **66**, 486–501 (2010).
- Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

Acknowledgements

This work was supported by Ministry of Science and Technology of China (Grant No 2017YFA0504900), the National Natural Science Foundation of China (Grant Nos. 31270796 and 81573325), CAMS Initiative for Innovative Medicine (CAMS-2017-I2M-2-004) and the National Mega-project for Innovative Drugs (Grant No. 2018ZX09711001-006). We also thank the staff of Shanghai Synchrotron Radiation Facility, instrument BL17U for the support in diffraction data collection.

Author contributions

L.Y. conducted the X-ray experiments and analyzed the experimental data. T.C., F.W., L.L., W.Y., Y.S., J.C. and W.L. prepared the protein samples and finished the activity assay. W.G. P.Z., L.Y. and T. C. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-019-0744-4>.

Correspondence and requests for materials should be addressed to P.Z. or W.G.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020