Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21

Møller, Marie Sofie; Abou Hachem, Maher ; Svensson, Birte; Henriksen, Anette

Published in:
Acta Crystallographica. Section F: Structural Biology and Crystallization Communications Online

Link to article, DOI:
10.1107/S1744309112031004

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21

Marie Sofie Møller,[a,b] Maher Abou Hachem,[a] Birte Svensson[a]* and Anette Henriksen[b]*

[a]Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Søltofts Plads, Building 224, 2800 Kgs. Lyngby, Denmark, and [b]Protein Chemistry Group, Carlsberg Laboratory, Gamle Carlsberg Vej 10, 2500 Valby, Denmark

Correspondence e-mail: bis@bio.dtu.dk, athx@novonordisk.com

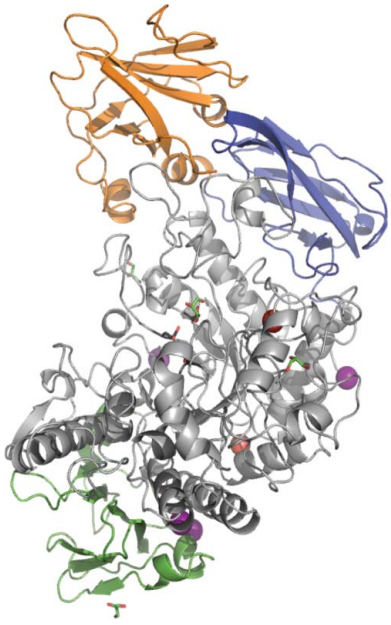
Received 22 May 2012
Accepted 7 July 2012

PDB Reference: barley limit dextrinase, 4aio.

Barley limit dextrinase (*Hv*LD) is a debranching enzyme from glycoside hydrolase family 13 subfamily 13 (GH13_13) that hydrolyses α-1,6-glucosidic linkages in limit dextrins derived from amylopectin. The structure of *Hv*LD was solved and refined to 1.9 Å resolution. The structure has a glycerol molecule in the active site and is virtually identical to the structures of *Hv*LD in complex with the competitive inhibitors α-cyclodextrin and β-cyclodextrin solved to 2.5 and 2.1 Å resolution, respectively. However, three loops in the N-terminal domain that are shown here to resemble carbohydrate-binding module family 21 were traceable and were included in the present *Hv*LD structure but were too flexible to be traced and included in the structures of the two *Hv*LD–inhibitor complexes.

1. Introduction

Barley limit dextrinase (*Hv*LD) catalyses the debranching of limit dextrins derived from amylopectin, the major constituent of barley starch. Starch amounts to 60% of the total dry weight of cereal grains and consists of an approximately 30:70 mixture of the essentially linear α-1,4-glucan amylose and the α-1,6-branched α-1,4-glucan amylopectin. Enzyme-mediated mobilization of storage starch granules in the endosperm of germinating cereal seeds involves solubilization by the concerted action of α-amylase, β-amylase, limit dextrinase (LD) and α-glucosidase, resulting in dextrins, maltooligosaccharides and glucose. Among these enzymes, only LD has the capacity to hydrolyse α-1,6-glucosidic linkages in branched α-limit and β-limit dextrins (Kristensen *et al.*, 1999). In addition, LD can hydrolyse α-1,6-glucosidic linkages in pullulan and, with low efficiency, 1,6-branch points in amylopectin (Kristensen *et al.*, 1999; Burton *et al.*, 1999). LD catalyses hydrolysis *via* the general acid/base double-displacement mechanism characteristic of glycoside hydrolase family 13 (GH13; Cantarel *et al.*, 2009; MacGregor *et al.*, 2001) by the action of a catalytic nucleophile Asp473 (numbering refers to *Hv*LD; Q9S7S8) and a catalytic acid/base proton donor Glu510, resulting in retention of the anomeric configuration of the products. Recently, expression of *Hv*LD has successfully been established in *Pichia pastoris* (Vester-Christensen, Abou Hachem, Naested *et al.*, 2010). The crystal structures of *Hv*LD in complex with the competitive inhibitors α-cyclodextrin (α-CD) and β-cyclodextrin (β-CD) have been solved and refined to 2.5 and 2.1 Å resolution, respectively (Vester-Christensen, Abou Hachem, Svensson *et al.*, 2010). *Hv*LD has four structural domains: the N-domain (residues 1–124), a carbohydrate-binding module from family 48 (CBM48; residues 125–230), the catalytic domain (residues 231–774) and the C-domain (residues 775–885). The structures of *Hv*LD–α-CD and *Hv*LD–β-CD showed overall good electron density, but the two first amino-acid residues and three loops (residues 23–27, 42–48 and 102–109) in the N-domain have low-level or no σ_A-weighted $2F_o - F_c$ electron density and were not included in the model (Vester-Christensen, Abou Hachem, Svensson *et al.*, 2010). The function of the N-domain is not clear, but it



is typical of GH13 enzymes that cleave or form endo-α-1,6-linkages (Jespersen *et al.*, 1991) and hence is presumed to have a functional linkage to this specificity. In this paper, we report the 1.9 Å resolution X-ray crystallographic structure of *Hv*LD, including a fully traced backbone of the N-domain which, in spite of a low sequence identity of 6%, can also be seen to possess structural similarity to the N-terminal CBM21 domain of glucoamylase from *Rhizopus oryzae* (Liu *et al.*, 2007).

2. Materials and methods

2.1. Crystallization, data collection and processing, structure determination and refinement

Recombinant *Hv*LD was prepared using *P. pastoris* as a host and was purified as described previously (Vester-Christensen, Abou Hachem, Naested *et al.*, 2010). The protein was concentrated to 10 mg ml^{-1} in 50 m*M* MES buffer pH 6.6, 250 m*M* NaCl, 0.5 m*M*

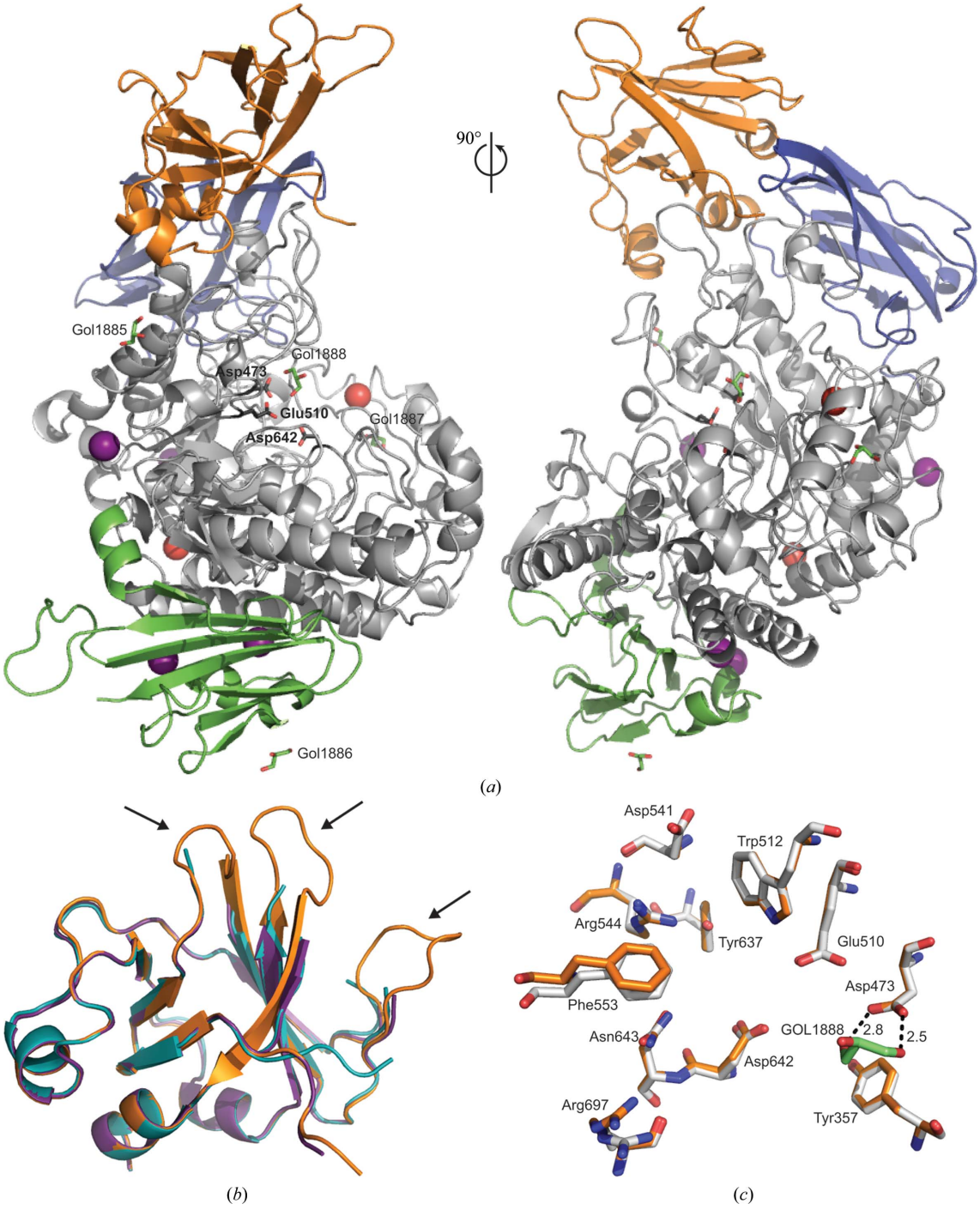


Figure 1
(*a*) Overall structure of *Hv*LD in two orientations. N-domain, orange; CBM48, blue; catalytic domain, grey; C-domain, green; Ca^{2+}, red; I^{-}, purple; GOL, green sticks. The catalytic residues (Asp473, Glu510 and Asp642) are shown as black sticks. (*b*) Comparison of the N-terminal domain of *Hv*LD (orange) with the same domain of *Hv*LD in complex with α-CD (cyan) or β-CD (purple). The missing loops are indicated by arrows. (*c*) Superposition of the amino-acid residues of the active sites of *Hv*LD (orange) and the *Hv*LD–β-CD structure (grey). GOL1888 is shown in green.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer resolution shell.

Data-collection details	
Wavelength (Å)	0.976
Resolution range (Å)	33.7–1.90 (2.00–1.90)
Unit-cell parameters (Å, °)	a = 176.1, b = 82.1, c = 59.4, β = 96.2
Space group	$C2$
No. of observed reflections	156615 (23012)
No. of unique reflections	60352 (9080)
Wilson B factor (Å^2)	17.3
Completeness (%)	91.3 (94.5)
$\langle I/\sigma(I)\rangle$	7.8 (2.1)
Multiplicity	2.6 (2.5)
$R_{\rm merge}$†	0.113 (0.616)
$R_{\rm p.i.m.}$‡	0.065 (0.349)
Refinement	
Reflections used	57222
$R_{\rm cryst}/R_{\rm free}$ (%)	18.9/22.5
No. of protein atoms	7042
No. of calcium ions	2
No. of iodide ions	4
No. of glycerol molecules	4
No. of water molecules	297
Mean B factor (Å^2)	
All atoms	21.1
Protein atoms	20.7
Cruickshank's DPI for coordinate error (Å)§	0.2
R.m.s.d. values from ideal	
Bond lengths (Å)	0.008
Bond angles (°)	1.165
Ramachandran plot	
Allowed regions (%)	99.66
Disallowed regions (%)	0.34
MolProbity score	1.26

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average over all observations of reflection hkl. ‡ $R_{\rm p.i.m.}$ is the multiplicity-weighted $R_{\rm merge}$ (Weiss, 2001). § Cruickshank's diffraction-component precision index (DPI) (Cruickshank, 1999).

$CaCl_2$, 0.67 m*M* maltotriose, resulting in a sixfold molar excess of maltotriose, and crystals of *Hv*LD were obtained by hanging-drop vapour diffusion at 293 K. Optimized crystals were obtained by streak-seeding using a reservoir solution consisting of 30%(*w*/*v*) polyethylene glycol (PEG) 3350, 5% glycerol, 0.3 *M* NaI. Cysteine was added to the crystallization drops to a final concentration of 5–7 m*M*. Crystals appeared within one week. The *Hv*LD crystals were cryoprotected by changing the PEG 3350 concentration of the drops to approximately 35% by stepwise addition of 35%(*w*/*v*) PEG 3350, 5% glycerol, 0.3 *M* NaI to the drop until cryoprotection was achieved. The crystals were mounted on Mesh LithoLoops (0.2 mm loop size and 40 µm mesh size; Molecular Dimensions, Newmarket, England) and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected on beamline ID23-1 at the European Synchroton Radiation Facility (ESRF; Grenoble, France) with λ = 0.976 Å. The data were integrated using *MOSFLM* (Leslie, 1992) and scaled with *SCALA* from the *CCP*4 program suite (Winn *et al.*, 2011). The resulting structure factors were used for molecular replacement (MR) using *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP*4 suite and the *Hv*LD–*β*-CD model (PDB entry 2y4s; Vester-Christensen, Abou Hachem, Svensson *et al.*, 2010) including only the protein moiety. The model was refined using *REFMAC*5 (Murshudov *et al.*, 2011). Manual inspection, rebuilding and addition of water molecules and ions were performed with *Coot* (Emsley *et al.*, 2010). In addition to the *Coot* validation functions, final analysis of model geometry optimization was performed using the output from *PROCHECK* and *MolProbity* (Laskowski *et al.*, 1993; Chen *et al.*, 2010).

Two structure-based alignment tools were used in order to advance insight into the possible role of the *Hv*LD N-domain (residues 2–124): a *DALI* search (Holm & Rosenström, 2010) against all PDB entries and *FATCAT* structural alignment (Ye & Godzik, 2003). In addition, a search using *PDBeFold* (Krissinel & Henrick, 2004) was performed, but no additional information was gained. The structure-based searches were also performed using the N-domain from the *Hv*LD–*β*-CD structure (PDB entry 2y4s), but the number of significant hits was low compared with the searches with the N-domain from native *Hv*LD owing to the missing loop regions and did not include the CMB21 domain.

3. Results and discussion

3.1. Structure determination and model quality

Two calcium ions, four iodide ions, four glycerol molecules and 294 water molecules were modelled in *Hv*LD. Refinement statistics are listed in Table 1. The geometry of the models is good, with 99.7% of the residues in the allowed regions of the Ramachandran plot and three residues (Lys107, Leu116 and Ala439) in the disallowed region. Ala439 is found in a similar position and intramolecular arrangement as in the *Hv*LD–*β*-CD structure (PDB entry 2y4s) used for molecular replacement. Lys107 resides in a flexible loop and Leu116 resides in the third α-helix of the N-terminal domain.

3.2. Overall structure

The *Hv*LD structure (Fig. 1*a*) and the protein moiety of *Hv*LD–*β*-CD are virtually identical, with an r.m.s.d. of 0.2 Å for all C^α atoms. The major difference between the structures is that the three short loops (residues 23–27, 42–48 and 102–109) in the N-domain are included in the N-domain of the *Hv*LD structure (Fig. 1*b*), which consists of seven *β*-strands arranged in an antiparallel fashion and three α-helices.

Four glycerol molecules (Gol) from the crystallization buffer and the cryoprotectant were found in *Hv*LD (Fig. 1*a*). Gol1885 is located at the interface between CBM48 and the catalytic domain and Gol1886 is located on the exposed surface of the C-domain. Gol1887 is buried in part of loop 2, similar to Gol306 in *Hv*LD–*β*-CD. Gol1888 is found in the active site, interacting with the catalytic nucleophile Asp473, and shows the same interaction pattern as a glycerol molecule in *Hv*LD–*β*-CD (Fig. 1*c*).

3.3. Active site

The amino-acid residues in the active site of *Hv*LD are found in a similar arrangement and adopt the same rotamers as the amino-acid residues in the *Hv*LD structures with α-CD and *β*-CD bound in the active site (Fig. 1*c*).

Mikami *et al.* (2006) observed a substrate-induced conformational change of the active-site residues connecting the acid/base catalytic residue (Glu706) and the C2 binding site (Trp708) in the case of the GH13 pullulanase from *Klebsiella pneumoniae*, which belongs to the same subfamily as *Hv*LD according to CAZy (Cantarel *et al.*, 2009). They observed two different main-chain conformations of the loop (residues 706–710; EGWDS) depending on whether or not a ligand (in this case glucose, isomaltose, maltose, maltotriose or maltotetraose) was bound. In addition, the side chain of Trp708 made about a 90° rotation to enable a stacking interaction at the active-site +2 subsite. In the native pullulanase structure (PDB entry 2fgz) and in the structures with bound glucose (PDB entry 2fh6) or isomaltose (PDB entry 2fh8) the loop was in the 'inactive' free conformation,

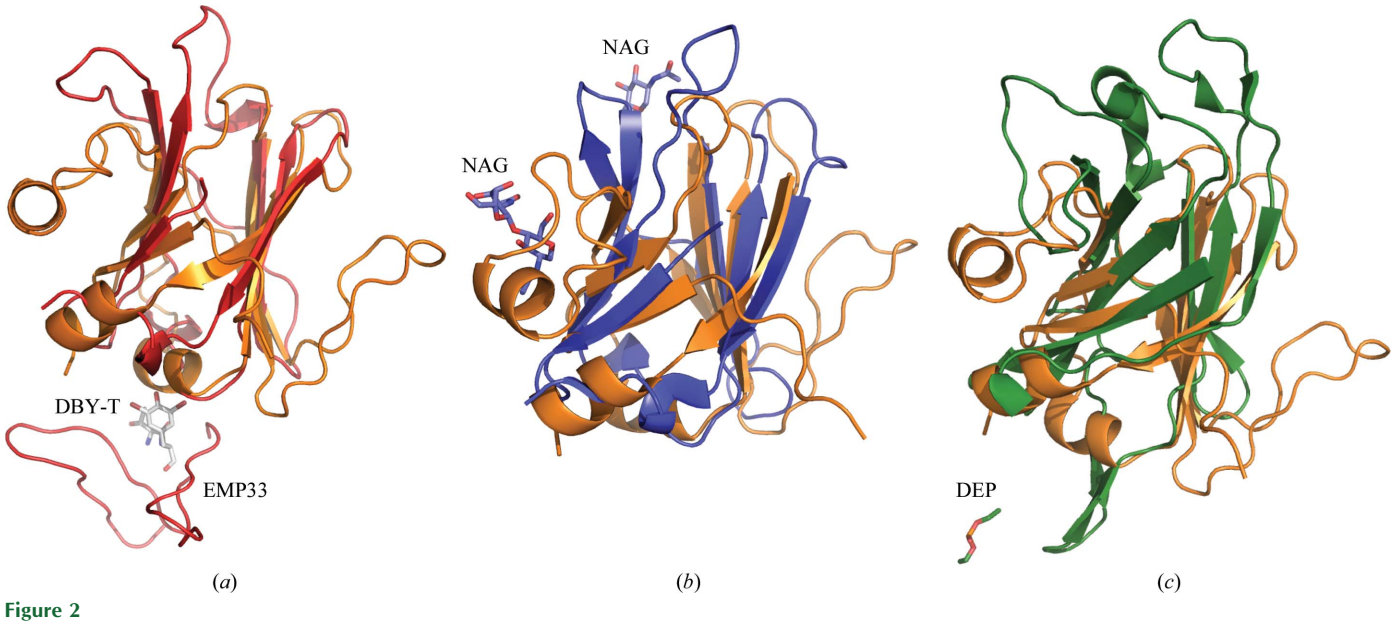


Figure 2
Superposition of the N-domain from *Hv*LD (orange) and structurally similar domains with documented functions identified by the *DALI* search. (*a*) Erythropoietin receptor (PDB entry 1eba; Livnah *et al.*, 1998; red) and the ligands from the structure: EMP33, an erythropoietin-mimic peptide, and DBY-T, 3,5-dibromotyrosine. (*b*) Cytokine receptor γ-chain (PDB entry 2b5i; Wang *et al.*, 2005; blue) and *N*-acetyl-D-glucosamine (NAG). (*c*) Esterase (PDB entry 3doi; Levisson *et al.*, 2009; green) and diethyl phosphonate (DEP).

while the loop was in the 'active' conformation in the complexes with maltose, maltotriose or maltotetraose (PDB entries 2fhb, 2fhc or 2fhf, respectively; Mikami *et al.*, 2006). The loop is one of the conserved regions of GH13 (MacGregor *et al.*, 2001) and is also conserved in *Hv*LD (residues 510–514; EGWDF). In *Hv*LD the loop is found in the 'active' form both in the case of the native structure presented here and in the *Hv*LD–CD complexes, in which the loop and Trp512 in particular participate in binding. Noticeably, Trp512 of native *Hv*LD is also in the 'active' rotamer position. It may be argued that the HvLD structure is not in its native state and that the glycerol molecule (Gol1888; Fig. 1*c*) in the active site could induce the change to the 'active' form. However, this does not seem to be a valid explanation since the glycerol molecule is interacting with the catalytic nucleophile Asp473, which is not part of the abovementioned conserved loop that changes conformation and makes no interactions with it. A conformational change upon substrate binding has been observed for several GH13-like enzymes (Barends *et al.*, 2007; Przylas *et al.*, 2000; Hondoh *et al.*, 2003; Woo *et al.*, 2008), among which is a GH13 glycogen-debranching enzyme from *Sulfolobus solfataricus* (Woo *et al.*, 2008), in which the Trp adopts the same rotamer and is in the same position as *Hv*LD when substrate is bound.

These findings suggest that *Hv*LD activity is not dependent on conformational changes of active-site amino-acid residues, unlike the pullulanase discussed above. This may indicate that the active site of *Hv*LD is less flexible, possibly explaining the lower hydrolytic activity of *Hv*LD towards large substrates such as amylopectin and the high activity towards the oligosaccharide limit dextrins compared with bacterial pullulanases.

3.4. The N-terminal domain

Superposition of the N-terminal domain of *Hv*LD with those of the deposited *Hv*LD–α-CD and *Hv*LD–β-CD complex structures (Fig. 1*b*) shows no significant variability in the conformation except for a different tucking in of the N-terminal amino-acid residues 2–5 to the rest of the molecule in the *Hv*LD–α-CD structure (PDB entry 2y5e; Vester-Christensen, Abou Hachem, Svensson *et al.*, 2010) and the previously mentioned well defined loop density of the three flexible loops in the native *Hv*LD structure.

Several alignment methods were explored to advance insight into the possible role of the *Hv*LD N-domain. A *DALI* search (Holm & Rosenström, 2010) with this domain against the entire PDB archive identified nine unique structures with *DALI* Z-scores of above 5 (Supplementary Table 1[1]). Only five of these proteins are α-1,6-acting pullulanases belonging to GH13_13 and GH13_14 and the sequence identity to *Hv*LD is in general low (see Supplementary Fig. 1[1]). Common to the hits is that they, like the *Hv*LD N-domain, do not harbour the active-site residues. Three of the identified domains have documented, albeit diverse, functions. These include binding of a peptide ligand, domain multimerization and *N*-acetyl-D-glucosamine (NAG) binding (Supplementary Table 1[1]). Noticeably, the parts of the domains involved in these interactions are not structurally similar (Fig. 2). The discrepancy between the amino-acid residues involved in intermolecular interactions and the lack of structural conservation of the same residues indicate that the various functionalities have evolved independently, suggesting that the N-terminal domain is a stable generic scaffold for mediating intermolecular interactions. *FATCAT* structural alignment (Ye & Godzik, 2003) with the complete N-terminal domain as present in *Hv*LD identified only pullulanase N-terminal domains with a *FATCAT P*-value of below 1.0×10^{-3}. Noticeably, the N-terminal starch-binding domain of the CBM21 glucoamylase from *R. oryzae* (PDB entry 2djm; Liu *et al.*, 2007) and the N-domain of *Hv*LD align with a *P*-value of 1.44×10^{-3} despite having a sequence identity of only 6% (Supplementary Fig. 2[1]). Ser76, Tyr78, Ser86 and Lys94 of *Hv*LD are the only surface-exposed residues among the identical residues from the structure-based sequence alignment, and although they are clustered from a steric point of view they are located in a part of the domain which is not structurally conserved (Supplementary Fig. 2[1]). The starch-

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: HV5219).

binding residues identified in *R. oryzae* CBM21 (Tung *et al.*, 2008) are not conserved or are replaced by residues with similar biophysical properties in *Hv*LD (Supplementary Fig. 2). It therefore seems unlikely that these residues play similar roles in the two molecules unless major structural changes occur in *Hv*LD in the presence of starch. In conclusion, the N-terminal domain of *Hv*LD may participate in intermolecular interactions that are important for the *in vivo* functionality of *Hv*LD, but there are no indications of whether the interactions involve multimerization, interactions with other proteins or interactions with substrate.

Access to synchrotron beam time was made possible by support from DANSCATT. We would like to acknowledge beamline scientist Christoph Mueller-Dieckmann (ESRF beamline ID23-1) for assistance during data collection. MSM was supported by a DTU PhD stipend.

References

Barends, T. R., Bultema, J. B., Kaper, T., van der Maarel, M. J., Dijkhuizen, L. & Dijkstra, B. W. (2007). *J. Biol. Chem.* **282**, 17242–17249.

Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* D**65**, 510–512.

Burton, R. A., Zhang, X. Q., Hrmova, M. & Fincher, G. B. (1999). *Plant Physiol.* **119**, 859–871.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.

Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Gourlay, L. J., Santi, I., Pezzicoli, A., Grandi, G., Soriani, M. & Bolognesi, M. (2009). *J. Bacteriol.* **191**, 3544–3552.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Hondoh, H., Kuriki, T. & Matsuura, Y. (2003). *J. Mol. Biol.* **326**, 177–188.

Jespersen, H. M., MacGregor, E. A., Sierks, M. R. & Svensson, B. (1991). *Biochem. J.* **280**, 51–55.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Kristensen, M., Lok, F., Planchot, V., Svendsen, I., Leah, R. & Svensson, B. (1999). *Biochim. Biophys. Acta*, **1431**, 538–546.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Levisson, M., Sun, L., Hendriks, S., Swinkels, P., Akveld, T., Bultema, J. B., Barendregt, A., van den Heuvel, R. H., Dijkstra, B. W., van der Oost, J. & Kengen, S. W. (2009). *J. Mol. Biol.* **385**, 949–962.

Liu, Y.-N., Lai, Y.-T., Chou, W.-I., Chang, M. D.-T. & Lyu, P.-C. (2007). *Biochem. J.* **403**, 21–30.

Livnah, O., Johnson, D. L., Stura, E. A., Farrell, F. X., Barbone, F. P., You, Y., Liu, K. D., Goldsmith, M. A., He, W., Krause, C. D., Pestka, S., Jolliffe, L. K. & Wilson, I. A. (1998). *Nature Struct. Biol.* **5**, 993–1004.

MacGregor, E. A., Janecek, S. & Svensson, B. (2001). *Biochim. Biophys. Acta*, **1546**, 1–20.

Mikami, B., Iwamoto, H., Malle, D., Yoon, H.-J., Demirkan-Sarikaya, E., Mezaki, Y. & Katsuya, Y. (2006). *J. Mol. Biol.* **359**, 690–707.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Padavattan, S., Flicker, S., Schirmer, T., Madritsch, C., Randow, S., Reese, G., Vieths, S., Lupinek, C., Ebner, C., Valenta, R. & Markovic-Housley, Z. (2009). *J. Immunol.* **182**, 2141–2151.

Przylas, I., Terada, Y., Fujii, K., Takaha, T., Saenger, W. & Sträter, N. (2000). *Eur. J. Biochem.* **267**, 6903–6913.

Tung, J.-Y., Chang, M. D.-T., Chou, W.-I., Liu, Y.-Y., Yeh, Y.-H., Chang, F.-Y., Lin, S.-C., Qiu, Z.-L. & Sun, Y.-J. (2008). *Biochem. J.* **416**, 27–36.

Turkenburg, J. P., Brzozowski, A. M., Svendsen, A., Borchert, T. V., Davies, G. J. & Wilson, K. S. (2009). *Proteins*, **76**, 516–519.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Vester-Christensen, M. B., Abou Hachem, M., Naested, H. & Svensson, B. (2010). *Protein Expr. Purif.* **69**, 112–119.

Vester-Christensen, M. B., Abou Hachem, M., Svensson, B. & Henriksen, A. (2010). *J. Mol. Biol.* **403**, 739–750.

Wang, X., Rickert, M. & Garcia, K. C. (2005). *Science*, **310**, 1159–1163.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Woo, E. J., Lee, S., Cha, H., Park, J. T., Yoon, S. M., Song, H. N. & Park, K. H. (2008). *J. Biol. Chem.* **283**, 28641–28648.

Ye, Y. & Godzik, A. (2003). *Bioinformatics*, **19**, ii246–ii255.