PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Structure of a pullulanase from *Bacillus acidopullulyticus*

Johan P. Turkenburg,[1] A. Marek Brzozowski,[1] Allan Svendsen,[2] Torben V. Borchert,[2] Gideon J. Davies,[1]* and Keith S. Wilson[1]*

[1] Department of Chemistry, Structural Biology Laboratory, The University of York, Heslington, York YO10 5YW, United Kingdom

[2] Novozymes A/S, Brudelysvej 26, 2880 Bagsvaerd, Denmark

Key words: glycoside hydrolase; X-ray structure; polysaccharide; amylase; starch; carbohydrate.

INTRODUCTION

The industrial hydrolysis of polysaccharides continues to be of considerable interest and, with the drive towards green chemical processes and renewable sources of fuel, the enzymatic hydrolysis of plant carbohydrates is only set to increase. One of the most widely used plant derived polysaccharides are the starches; predominantly α-1,4 linked D-glucopyranoside polymers with, to varying degrees, α-1,6 branches. Complete starch hydrolysis requires a consortium of enzymes including *endo*-amylases, glucoamylases and α-glucosidases as well as diverse α-1,6 cleaving enzymes including "pullulanases". Many of these α-glucan active enzymes are found in the largest of the *CA*rbohydrate-active en*ZY*mes (CAZY) sequence families, family GH13 (see www.cazy.org[1,2] recently reviewed in a historical context in Ref. 3). Considering the diversity of GH13 enzymes[4] and their burgeoning applications in environmental-friendly processes, there remains a need to study these enzymes further particularly to probe their 3-D structures and often complex modular architectures.

Here we report the 3-D structure, refined in two crystal forms at resolutions of 1.7 and 2.1 Å, respectively, of a pullulanase from the bacterium *Bacillus acidopullulyticus*. The substrate, pullulan, is a polysaccharide in which repeating maltotriosyl units (α-1,4 linked) are joined through α-1,6 links, Figure 1. Pullulanases cleave the α-1,6 links to liberate maltotriose. In industrial applications, these enzymes find use not merely for the breakdown of pullulan but also for the hydrolysis of the α-1,6 linkages in amylopectin starch. The *B. acidopullulyticus* enzyme (hereafter *Ba*Pul13A) thus finds commercial application in the starch industry where its "debranching" activity is utilized in the production of high fructose corn syrup[5] and in the production of high maltose content syrups but also in the brewing industry, especially in the production of low calorie and "light" beers where it allows more complete attenuation (i.e., more complete fermentation of the mash and less residual sugar). The 3-D structure of *Ba*Pul13A was solved by single isomorphous replacement using an "in-house" uranyl derivative and subsequently refined in two different crystal forms. Pullulanases are frequently characterized by complex multi-domain architectures in which the catalytic module is appended to several carbohydrate-binding domains (CBMs) as well as many domains of unknown function (termed "X" modules). The structure of *Ba*Pul13a indeed forms an unusual domain organization, in which the *N*-terminal CBM41 domain is disordered/partially absent in-crystal but in which the X45a-X25-X45b-CBM48-GH13 multi-domain architecture is both clear and not, to our knowledge, previously observed. Comparison with the known *Klebsiella pneumoniae* pullulanase[6] and the pullulanase/debranching enzyme from *Bacillus subtilis* reveals a conserved active centre and similar polysaccharide binding surfaces consistent with α-1,6 hydrolase activity within the α-amylase fold.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Gideon J. Davies or Keith S. Wilson, Department of Chemistry, Structural Biology Laboratory, The University of York, Heslington, York, YO10 5YW, U.K. E-mail: keith@ysbl.york.ac.uk or davies@ysbl.york.ac.uk

Received 26 January 2009; Revised 12 February 2009; Accepted 17 February 2009
Published online 9 March 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22416

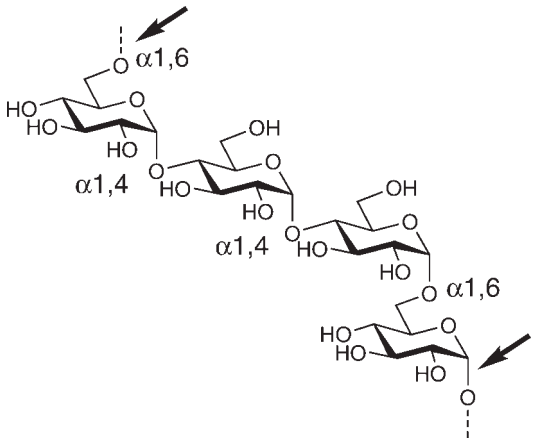

Figure 1

The structure of pullulan; a polysaccharide in which repeating (α-1,4 linked) maltotriosyl units are joined by α-1,6 linkages. Pullulanases cut the α-1,6 links indicated with bold arrows as well as finding applications for the hydrolysis of α-1,6 linkages in amylopectin.

METHODS

*Ba*Pul13A was provided to the University of York as a kind gift from Novozymes A/S Denmark where it had been produced as described previously (Ref. 7 and references cited therein). The protein was crystallized in two forms "I" and "II". Form I crystals were grown from 20 to 35% ammonium sulfate, 0.1 M acetate buffer pH 4.5, and 5–10 m*M* of one of several additives including NDSB 195 (or 256), or detergent SB12. Form I crystals are in space group C2 with unit cell dimensions a = 206.8 Å, b = 55.6 Å, c = 112.0 Å, β = 116°. Seeding of these crystals into the same crystallization conditions gave rise to a crystal form, termed form II which has space group C2, with unit cell dimensions a = 162.7 Å, b = 61.6 Å, c = 109.2 Å, β = 109.1°. Native data were collected on form I using "in house" Cu Kα radiation to 2.1 Å resolution (not shown) and on form II to approximately 1.7 Å resolution at beamline ID14-4 of the European Synchrotron Radiation Facility (ESRF) at a wavelength of 0.95 Å. Data were processed and reduced using DENZO/SCALEPACK,[8] Table I. The structure was solved using an "in house" uranyl acetate heavy atom derivative collected to 2.2 Å resolution (100% complete, <I>/<σI> of 5.7 in the outer bin). Three heavy atom sites were found by inspection of the difference Patterson(s), the data were merged with the native data (mean fractional isomorphous difference 0.31 unweighted, 0.23 weighted) and single isomorphous replacement (with anomalous scattering) phases calculated using the CCP4 suite[9] implementation of MLPHARE. Initial phasing yielded an overall R_{cullis} of 0.76, phasing power (centric/acentric) 1.2/1.5, figure of merit (centric/acentric) of 0.37/0.54). Phases were improved with DM and used in conjunction with the 1.7 Å native data and the structure built semiautomatically using ARPwARP[10] with QUANTA (Accelrys, San Diego) and more latterly COOT[11] used for model building with maximum likelihood refinement using REFMAC.[12] Final X-ray data and structure refinement statistics are given in Table I.

RESULTS AND DISCUSSION

*Ba*Pul13A is a mature protein of 921 amino acids whose catalytic domain belongs to family GH13 of the CAZY classification. Family GH13 is one of the largest of the glycosidase families with over 5000 members (Oct. 2008), 115 of which have a 3-D structure determined (corresponding to 273 PDB entries). This massive family has recently been subdivided with 35 subfamilies currently disclosed.[4] The sequence-similarity of *Ba*Pul13A suggests a complex modular architecture, Figure 2A, consisting of CBM41-X45a-X25-X45b-CBM48-GH13_14 where CBM is a carbohydrate-binding domain,[14] "X" a module of unknown function and GH13_14 indicates a CAZY GH13 domain (itself consisting of the three classical A,B and C domains of generic α-amylases) recently subclassified as a subfamily 14 domain.[4] BaPul13A is therefore a seven domain protein with the three-domain catalytic module linked to four N-terminal, auxiliary domains.

The 3-D structure (Fig. 2B), at 1.7 Å resolution, is entirely consistent with this CAZYModo[1] derived modular assignment although the CBM41 *N* terminal domain

Table I

X-ray Data and Structure Refinement Statistics

Space group	C2
Cell dimensions, a b, c (Å)	162.8, 61.6, 109.2
Cell dimensions, α,β,γ (°)	90, 109.1, 90
Data processing	
Resolution range[a]	20–1.65 (1.68–1.65)
R_{merge}	0.072 (0.69)
<I>/<σI>[a]	14.5 (1.5)
Completeness (%)[a]	96 (90)
Redundancy[a]	3.1 (2.7)
Model refinement	
No. reflections used	118628
R_{cryst}	0.17
R_{free}	0.21
No. protein atoms	7311
Ligands	4 Sulfate, 8 glycerol, 4 acetate
r.m.s.d. bonds (Å)	0.010
r.m.s.d angles (°)	1.24
Mean *B* value protein atoms (Å)[b]	Main 18, side 20
ESU[b] (Å)	0.09
Ramachandran plot[c] (%)	
Favored	95.2
Allowed	4.2
Outliers	0.6
PDB Code	2wan

[a]Numbers in parenthesis correspond to the high resolution outer shell.
[b]Estimated standard uncertainty, based upon R_{free}, calculated using REFMAC.
[c]Calculated using validation options in COOT.

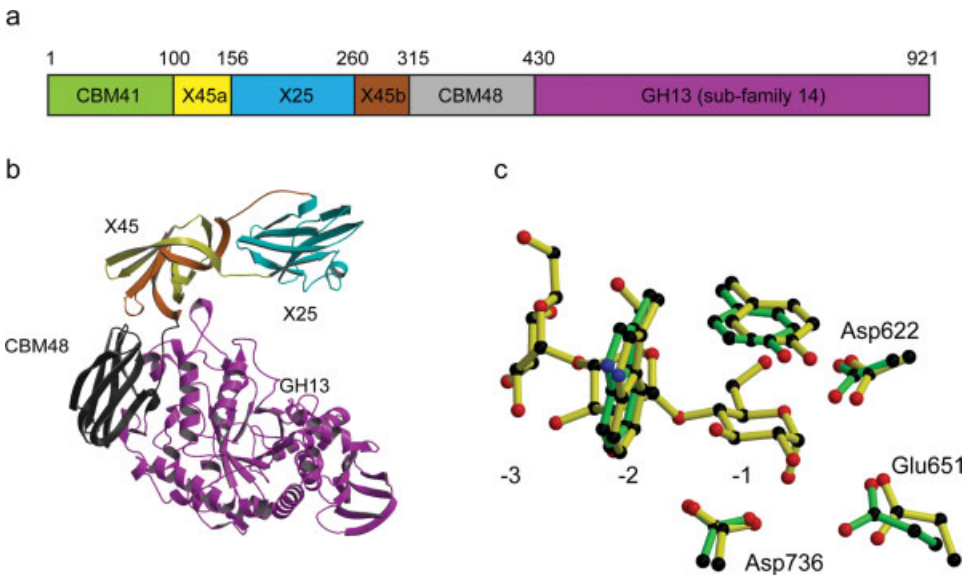

Figure 2

3-D structure of the *Bacillus acidopullulyticus* family GH13 pullulanase. (**a**) Schematic diagram of the modular architecture of *Ba*Pul13A with the approximate (CAZy-defined[1]) domain boundaries indicated. (**b**) The structure, as a protein cartoon, in which the domains are colored approximately as in (a); drawn with BOBSCRIPT.[13] The CBM41 domain is disordered/low occupancy in-crystal and cannot be modeled. (**c**) active centre overlap (BaPul13 in green) with the *B. subtilis* pullulanase type I/glycogen-debranching enzyme complex with maltotriose (unpublished, PDB code 2E9B, in yellow). The catalytic "triads" are shown along with some conserved aromatic residues involved in binding the −3 to −1 subsite sugars in both enzymes.

is disordered in crystal (notably in both crystal forms) and cannot be modeled. At very low contour levels, there is some evidence in difference density maps for the N-terminal domain, and there is sufficient space in the packing for this domain to be accommodated, but attempts to model this domain did not give any meaningful results. The final crystallographic R-values indicate that this domain contributes little to the scattering experiment. It is therefore most likely that this domain is very disordered, and possibly also partly cleaved-off in solution. CBM41 domains themselves are known to have glycogen/starch binding functions (e.g., Ref. 15) and it is likely that the CBM41 domain of this enzyme will play a similar role. The observed 3-D structure therefore consists of X45a-X25-X45b-CBM48 GH13_14. There is one available structure of a subfamily GH13_14 enzyme, the *Bacillus subtilis* pullulanase type I/glycogen-debranching enzyme (PDB codes: 2E8Y, 2E8Z, and 2E9B) but in this case the architecture is far simpler with just 104 residues of unknown function followed by a CBM48 and the catalytic domain. The CBM48 domain is, as with CBM41, a domain for which a glycogen/starch binding function is known (e.g., in the AMP-activated protein kinase[16]).

Of particular note in the structure of *Ba*Pul13A is the insertion of a new domain of unknown function "X25" in a loop present in the X45 domain. The insertion of one domain, splitting another, is a feature rarely observed in glycosidase 3-D structures but, judging by sequence alone, it is quite common for the X45-X25 pair. X25 itself is also found in tandem copies, such as in the highly modular amylopullulanase from *Geobacillus stearothermophilus TS-23* (sequence id AAG44799.1), whilst un-split X45 may exist, occasionally, without the X25 insertion, notably in pulB from *Bacillus acidopullulyticus*[17] (GenBank CAC60158) and perhaps more distantly in pullulanases from *Anaerobranca gottschalkii*.[18] Given the widespread prevalence of X25 and X45 domains in pullulanases, it is tempting to suggest that these β-sheet domains (reported here, we believe, for the first time) are carbohydrate-binding domains which target mixed α-1,6/α-1,4 linked D-glucan polysaccharides such as pullulan.

The *Ba*Pul13A active centre in which hydrolysis of α-1,6 linkages occurs with net retention of anomeric configuration, via a covalent glycosyl-enzyme intermediate, (glycosidase mechanisms are recently reviewed in Ref. 19). Classically, the active centre of α-amylases is characterized by three Asp/Glu residues: an enzymatic nucleophile, the Brønsted acid/base and a third carboxylate side-chain that interacts with the O2 and O3 sugar hydroxyls in the catalytic centre. The active centre of *Ba*Pul13A is no exception with Asp622 as nucleophile, Glu651 as acid/base and Asp736 the transition-state stabilizer that would interact with O3/O3, Fig 2C. Consistent with the CAZy subfamily 14 assignment, the active centre is extremely similar to the *Bacillus subtilis* pullula-

nase type I/glycogen-debranching enzyme with the maltotriose complex of that enzyme (unpublished, PDB codes 2E9B) likely a good model of maltotriose binding in the −1 to −3 subsites[20] of *Ba*Pul13A with all interacting residues invariant, Fig 2C.

The use of enzymatic catalysts for plant polysaccharide degradation remains at the forefront of industrial biocatalysis, the 3-D structure of the *B. acidopullulyticus* is a powerful reminder of the complex multi-modular architectures that these enzymes often display. The study, and dissection, of this modularity should enable the development of more powerful biocatalyst for the demanding chemical processes required for a greener future.

ACKNOWLEDGMENTS

GJD is a Royal Society Wolfson-Merit Award Recipient. Dr Bernard Henrissat (CNRS, Marseille) is thanked for discussions concerning the complex modular architecture of GH13 enzymes.

REFERENCES

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-active enzymes database(CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 2009;37:D233–D238.
2. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 1997;7:637–644.
3. Davies GJ, Sinnott ML. Sorting the diverse: the sequence-based classifications of carbohydrate-active enzymes. Biochem J 2008;DOI 10.1042/BJ20080382 (online only).
4. Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng Des Sel 2006;19:555–562.
5. Norman BE. A novel debranching enzyme for application in the glucose syrup industry. Starke 1982;34:340–346.
6. Mikami B, Iwamoto H, Malle D, Yoon HJ, Demirkan-Sarikaya E, Mezaki Y, Katsuya Y. Crystal structure of pullulanase: Evidence for parallel binding of oligosaccharides in the active site. J Mol Biol 2006;359:690–707.
7. Svendsen A, Andersen C, Borchert TV. Pullulanase variants and methods for preparing such variants with predetermined properties. PCT Patent application WO/2001/051620, 2001.
8. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
9. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50(Pt 5):760–763.
10. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463.
11. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(Pt 12 Pt 1):2126–2132.
12. Murshudov GN. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53(Pt 3):240–255.
13. Esnouf RM. An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. J Mol Graph Model 1997;15:132–134.
14. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine tuning polysaccharide recognition. Biochem J 2004;382:769–781.
15. van Bueren AL, Higgins M, Wang D, Burke RD, Boraston AB. Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. Nat Struct Mol Biol 2007;14:76–84.
16. Polekhina G, Gupta A, van Denderen BJ, Feil SC, Kemp BE, Stapleton D, Parker MW. Structural basis for glycogen recognition by AMP-activated protein kinase. Structure 2005;13:1453–1462.
17. Kelly AP, Diderichsen B, Jorgensen S, McConnell DJ. Molecular-genetic analysis of the pullulanase b-gene of Bacillus acidopullulyticus. FEMS Microbiol Lett 1994;115:97–105.
18. Bertoldo C, Armbrecht M, Becker F, Schafer T, Antranikian G, Liebl W. Cloning, sequencing, and characterization of thermoalkalistable type I pullulanase from *Anaerobranca gottschalkii*. Appl Environ Microbiol 2004;70:3407–3416.
19. Vocadlo D, Davies GJ. Mechanistic insights into glycosidase chemistry. Curr Opin Chem Biol 2008;12:539–555.
20. Davies GJ, Wilson KS, Henrissat B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J 1997;321:557–559.