Structure, function and protein engineering in starch debranching enzyme systems. Barley limit dextrinase and its endogenous inhibitor

Møller, Marie Sofie; Henriksen, Anette

Publication date:
2013

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

Link back to DTU Orbit

Citation (APA):
Structure, function and protein engineering in starch debranching enzyme systems. Barley limit dextrinase and its endogenous inhibitor

Marie Sofie Møller
PhD Thesis
December 2012
Preface

This PhD thesis presents the results of my PhD project carried out at Enzyme and Protein Chemistry (EPC), Department of Systems Biology, Technical University of Denmark (DTU) in the period from December 2009 to December 2012 under supervision of Prof. Birte Svensson (EPC), Associate Prof. Maher Abou Hachem (EPC), and Senior Research Scientist Anette Henriksen (formerly Carlsberg Laboratory, now Novo Nordisk A/S). The parts of the PhD study involving crystallography took place at Carlsberg Laboratory. The PhD project was funded by a DTU PhD stipend.

The work has resulted in the following publications, which are included as chapters in the present thesis:


Furthermore, I have contributed to the following publication related to the topic of the PhD study:


Finally, work on an α-1,6-acting enzyme related to the topic of my PhD project has been finalised and published:


During my PhD I have given a number of oral presentations as well as poster contributions in various settings. They are summarised in Appendix VI.

Marie Sofie Møller, Kgs. Lyngby, December 2012
Acknowledgements

During the three years as a PhD student I have received help and advice from a number of people, whom I would like to thank. First of all my three supervisors: I am grateful to Professor Birte Svensson, EPC, for giving me the opportunity to carry out my PhD study at EPC, furthermore I am grateful for her support and expert advice throughout the three years. Associate Professor Maher Abou Hachem (EPC), who always has time to discuss scientific issues as well as issues related to life as a PhD student, is greatly acknowledged. With his support I have been growing as a researcher and as a person. My third supervisor Senior Research Scientist Anette Henriksen at Novo Nordisk (formerly Professor at Carlsberg Laboratory) is thanked for revealing some of the secrets of crystallography, and for guidance through the steps from crystal to structure. She showed me that there is a world outside DTU, and her always positive attitude and encouragements have been invaluable, when persistence was the last option.

All former and present members of EPC are thank for their help and discussions, especially Johanne Morch Jensen and Malene Bech Vester-Christensen, who established recombinant production and purification of limit dextrinase as well as limit dextrinase inhibitor. I would specifically like to thank Casper Wilkens, Alexander Holm Viborg, and Susan Andersen for fruitful discussions about scientific issues and about life as such. They have made even hard times funny and bearable. The EPC technicians Karina Rasmussen and Mette Fries are thanked for their technical assistance in lab, and Anne Blicher is acknowledged for performing amino acid analysis.

I would like to thank collaborators at the Carlsberg Laboratory: Professor Monica Palcic and present and former members of her group. Especially Lyann Sim and Michael Skovbo Windahl are thanked for technical help and support with crystallisation, and for production of protein for crystallisation. Jose A. Cuesta Seijo, is acknowledged for the collection of the final LD:LDI diffraction dataset.

"The Danish Centre for the use of Synchrotron X-ray and Neutron facilities" (DANSCATT) is acknowledged for access to synchrotron facilities at MaxLab, Sweden and ESRF, France.

Finally I would like to thank friends and family: especially my brother Niels for helping me with the layout of the present thesis and for being more nerdy than me, and my family for always believing in me, even when I had a hard time believing in myself.
Abstract

Starch is the most abundant storage carbohydrate in cereal grains. It is composed primarily of amylopectin, a polymer of glucose in which α-1,4-linked glucan chains are branched with α-1,6-bonds. Enzymatic degradation of starch in germinating barley seeds involves an initial solubilisation, mainly by α-amylase, followed by hydrolysis of the resulting dextrins to oligosaccharides and glucose by the concerted action of α- and β-amylase, limit dextrinase (LD), and α-glucosidase. Only LD is able to hydrolyse α-1,6-linkages in limit dextrans. Since LD is the sole debranching enzyme in the germinating seed, it has a key role in malting and brewing. But an endogenous inhibitor, limit dextrinase inhibitor (LDI), is present in barley seeds. It specifically inhibits LD, and thereby suppresses the degradation of branched limit dextrans to fermentable sugars. The knowledge about this enzyme:inhibitor pair and inhibitors of debranching enzymes is sparse.

During the last decades knowledge about LD has improved, recently with the crystal structures of LD in complex with the competitive inhibitors α- or β-cyclodextrin. But deeper insight into the substrate specificity determinants at molecular level is still sparse. Here I present crystal structures of LD, and LD in complex with 6\(^2\)-α-maltotriosyl-maltotriose, i.e. a limit dextrin, or two maltotriose molecules. The branched ligand is in contact with LD via interactions between all six glucose units and amino acid residues of LD. The active site cleft of LD can accommodate longer side chains, while the active site topology at both + and – subsites could cause steric hindrances. Met440 was proposed to be such a hindrance near subsite –3 based on structure comparisons between one of the first published structures of LD and a related enzyme from Klebsiella pneumoniae. LD-related enzymes with higher activity towards amylopectin as compared with LD in general have a cysteine or a glycine, at the equivalent position, thus an LD-M440G variant was produced. The kinetic constants of wild type LD on amylopectin were determined to be \(K_m=6.9\pm1.0\ \text{mg/ml}\) and \(k_{cat}=15.6\pm1.2\ \text{s}^{-1}\), resulting in a catalytic efficiency \((k_{cat}/K_m)\) of 2.3 ml(mg s)\(^{-1}\). This is more than 200-fold lower than the catalytic efficiency on pullulan. The LD-M440G variant had an even 2-fold lower \(k_{cat}/K_m\) (0.9 ml(mg s)\(^{-1}\)) on amylopectin. Based on this it is concluded that Met440 does not account for the low activity of LD on polymeric substrates. The structure of LD in complex with a branched substrate provides new possibilities for structural comparisons with other debranching enzymes. Active site topology elements of LD, like Phe553, were identified as possible substrate specificity determinants based on these comparisons.

In the second part of the thesis I present the crystal structure of the complex between LD and LDI determined to 2.7 Å. The structure revealed a novel mode of inhibition distinctly different from that of the interaction between LDI-like inhibitors and α-1,4-acting enzymes (α-amylases) from the same glycoside hydrolase family as LD (GH13), which involves the opposite face of LDI. Taking advantage of this LD:LDI structure a number of LD and LDI variants were constructed to elucidate hot spots of the interaction as well as to engineer LD to be less sensitive towards LDI. Mutations of residues in a hydrophobic cluster of LDI caused the most dramatic decreases in binding affinity. The LDI-L41G-V42D variant had close to 5×10\(^5\)-fold reduced affinity. The complex formation was independent of ionic strength, which was confirmed by the minor importance of two LDI arginines, Arg34 and Arg38, for the binding affinity. Arg38 thus interacted with two of the three catalytic site residues of LD. The LD:LDI structure also guided a successful LD mutation, D730R, reducing the sensitivity towards LDI by 180-fold without changing the LD kinetics on pullulan.
Resumé


De seneste årtier er viden om LD øget, senest med krystalstrukturen af LD i kompleks med de kompetitive inhiborer α- eller β-cyclodextrin. Men en dybere indsigd i substratspecificitetsdeterminanter på molekyleret niveau er begrænset. Her præsenterer jeg krystalstrukturerne af frit LD, og LD i kompleks med 62-α-maltotriosyl-maltotriose, dvs. en limit-dextrin, eller to maltotriosemolekyler. Den forgrenede ligand er i kontakt med LD via interaktioner imellem alle seks glykoseenheder og aminosyrer fra LD. Længere sidekæder er mulige, dog er det muligt at topologien i det aktivested, både ved + og – subsites, vil være skyld i stersiske hinderinger. Det var foreslået at Met440 kunne være sådan en hindring nær subsite –3 baseret på struktursammenligninger imellem en af de første publicerede LD-strukturer og et relateret enzym fra Klebsiella pneumoniae. LD-relaterede enzymer med højere aktivitet overfor amylopektin har i forhold til LD en cystein eller en glycine på den tilsvarende position, derfor blev en LD-M440G-variant produceret. De kinetiske konstanter for vildtype LD på amylopektin blev bestemt til at være $K_m = 6.9 \pm 1.0 \text{ mg/ml}$ og $k_{cat} = 15.6 \pm 1.2 \text{ s}^{-1}$, hvilket resulterer i en katalytiskeffektivitet $(k_{cat}/K_m)$ på ca. 2,3 ml/(mg s)$^{-1}$. Det er mere end 200 gange lavere end katalytiskeffektiviteten på pullulan. LD-M440G-varianten havde 2 gange lavere $(k_{cat}/K_m)$ på amylopektin. Baseret på det konkluderes det at Met440 ikke er skyld i LDs lave aktivitet på polymeriske substrater. Strukturen af LD i kompleks med et forgrenet substrat giver nye muligheder for struktural sammenligning med andre afgreningsenzymy. Topologi elementer, såsom Phe553, var identificeret som mulige substratspecificitetsdeterminanter baseret på sådanne sammenligninger.

I anden del af afhandlingen præsenterer jeg krystalstrukturen af komplekset imellem LD og LDI, som er bestemt til 2.7 Å. Strukturen afslørede en ny form for inhibering, som er markant forskellig fra interaktionen imellem LDI-lignende inhiborer og α-1,4-hydrolyserende enzymer (α-amylaser) fra den samme glykosidhydrolasefamilie som LD (GH13), som involverer den modsatte flade af LDI. LD:LDI strukturen blev udnyttet til at konstruere et antal LD- og LDI-varianter for at belyse vigtige aminosyrer for interaktionen og for at modificere LD til at være mindre sensitiv overfor LDI. Mutationer af rester i en hydrofobisk klynge i LDI resulterede i det mest dramatiske fald i bindingsaffinitet. LD-LDI strukturen blev udnyttet til at konstruere et antal LD- og LDI-varianter for at belyse vigtige aminosyrer for interaktionen og for at modificere LD til at være mindre sensitiv overfor LDI. Mutationer af rester i en hydrofobisk klynge i LDI resulterede i det mest dramatiske fald i bindingsaffinitet. LD-LDI strukturen blev udnyttet til at konstruere et antal LD- og LDI-varianter for at belyse vigtige aminosyrer for interaktionen og for at modificere LD til at være mindre sensitiv overfor LDI. Mutationer af rester i en hydrofobisk klynge i LDI resulterede i det mest dramatiske fald i bindingsaffinitet. LD-LDI strukturen blev udnyttet til at konstruere et antal LD- og LDI-varianter for at belyse vigtige aminosyrer for interaktionen og for at modificere LD til at være mindre sensitiv overfor LDI. Mutationer af rester i en hydrofobisk klynge i LDI resulterede i det mest dramatiske fald i bindingsaffinitet. LD-LDI strukturen blev udnyttet til at konstruere et antal LD- og LDI-varianter for at belyse vigtige aminosyrer for interaktionen og for at modificere LD til at være mindre sensitiv overfor LDI. Mutationer af rester i en hydrofobisk klynge i LDI resulterede i det mest dramatiske fald i bindingsaffinitet.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19 AI</td>
<td>0.19 α-amylase inhibitor from wheat</td>
</tr>
<tr>
<td>α-CD</td>
<td>α-cyclodextrin</td>
</tr>
<tr>
<td>BaPUL</td>
<td><em>Bacillus acidopullulyticus</em> pullulanase</td>
</tr>
<tr>
<td>β-CD</td>
<td>β-cyclodextrin</td>
</tr>
<tr>
<td>BsPUL</td>
<td><em>Bacillus subtilis</em> subsp. <em>subtilis</em> str. 168 pullulanase</td>
</tr>
<tr>
<td>CAZy</td>
<td>The Carbohydrate-Active enZYmes Database</td>
</tr>
<tr>
<td>CBM</td>
<td>carbohydrate binding module</td>
</tr>
<tr>
<td>CHFI</td>
<td>corn Hagemann factor inhibitor</td>
</tr>
<tr>
<td>CM</td>
<td>chloroform-methanol</td>
</tr>
<tr>
<td>DBE</td>
<td>debranching enzymes</td>
</tr>
<tr>
<td>dpa</td>
<td>days post anthesis</td>
</tr>
<tr>
<td>DSC</td>
<td>differential scanning calorimetry</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>G₂</td>
<td>maltosyl</td>
</tr>
<tr>
<td>G2F</td>
<td>α-maltosyl fluoride</td>
</tr>
<tr>
<td>G3-APTS</td>
<td>maltotriose-8-amino-1,3,6-pyrene trisulphonic acid</td>
</tr>
<tr>
<td>GA</td>
<td>gibberellic acid</td>
</tr>
<tr>
<td>GH13</td>
<td>glycoside hydrolase family 13</td>
</tr>
<tr>
<td>GH13_X</td>
<td>glycoside hydrolase family 13 subfamily X</td>
</tr>
<tr>
<td>ISA</td>
<td>isoamylase</td>
</tr>
<tr>
<td>KpPUL</td>
<td><em>Klebsiella pneumoniae</em> pullulanase</td>
</tr>
<tr>
<td>LD</td>
<td>limit dextrinase</td>
</tr>
<tr>
<td>M₃-M₃</td>
<td>6²-α-maltotriosyl-maltotriose</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>PUL</td>
<td>pullulanase</td>
</tr>
<tr>
<td>RBI</td>
<td>bifunctional α-amylase/trypsin inhibitor from ragi</td>
</tr>
<tr>
<td>SaPUL</td>
<td><em>Streptococcus agalactiae</em> pullulanase</td>
</tr>
<tr>
<td>SpPUL</td>
<td><em>Streptococcus pneumoniae</em> pullulanase</td>
</tr>
<tr>
<td>SPR</td>
<td>surface plasmon resonance</td>
</tr>
<tr>
<td>Trxh</td>
<td>thioredoxin h</td>
</tr>
</tbody>
</table>
Contents

Chapter 1: Introduction ................................. 1
  1.1 Limit dextrinase ........................................ 2
    1.1.1 Catalytic mechanism of the GH13 family .............. 4
    1.1.2 Substrate specificity of limit dextrinases/pullulanases 5
    1.1.3 Structures of pullulanases ................................ 11
    1.1.4 Transglycosylation by LD ................................ 14
  1.2 Limit dextrinase inhibitor ............................ 15
    1.2.1 CM proteins and cereal-type inhibitors .................. 16
  1.3 LD and LDI in barley .................................... 18
    1.3.1 Synthesis and localisation of LD and LDI in barley seeds 18
    1.3.2 Biological roles of LD and LDI .......................... 19
    1.3.3 In vivo regulation of LD and LDI ......................... 21
  1.4 The LD:LDI system in malting and brewing .............. 22
  1.5 Aims of the present study ............................. 24

Chapter 2: Paper: Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21 25

Chapter 3: Substrate specificity determinants of limit dextrinase 37
  3.1 Materials and methods ................................... 38
    3.1.1 Bioinformatic analysis .................................. 38
    3.1.2 Mutagenesis - LD-M440G variant ......................... 38
    3.1.3 Production and purification of LD, wild type and variants 38
    3.1.4 Assays .................................................. 39
    3.1.5 Structure determination and refinement .................. 40
  3.2 Results .................................................. 41
    3.2.1 Bioinformatic analysis .................................. 41
3.2.2 Substrate specificity analysis ............................................. 41
3.2.3 Crystal structures of LD complexed with ligands .................... 43
3.2.4 LD contra K. pneumoniae pullulanase ................................. 47
3.2.5 Active site architecture: LD contra GH13_12+14 pullulanases and de-branching enzymes from GH13_11 ........................................ 48
3.2.6 Structure based comparison of GH13 α-1,6-acting enzymes ........ 52

3.3 Discussion ............................................................................. 55
3.3.1 Structures of barley LD with and without ligand ..................... 55
3.3.2 Substrate specificity determinants of LD in relation to other α-1,6-acting enzymes ................................................... 56

Chapter 4: Manuscript: Structural rational for regulation of barley limit dextrinase activity by the endogenous proteinaceous limit dextrinase inhibitor 59

Chapter 5: Concluding remarks and perspectives 107

References 109

Appendix I: Specific activity 120

Appendix II: CM-proteins included in multiple alignment 124

Appendix III: Multiple sequence alignment - catalytic domains of GH13_8–9 and GH13_11–14 125

Appendix IV: Structure based multiple sequence alignment including the structures from GH13_8–9 and GH13_11–14 129

Appendix V: List of publications 132

Appendix VI: Presentations 133

Appendix VII: Paper: Enzymology and structure of the GH13_31 glucan 1,6-α-glucosidase that confers isomaltooligosaccharide utilization in the probiotic Lactobacillus acidophilus NCFM 135

Appendix VIII: Paper: Efficient secretory expression of functional barley limit dextrinase inhibitor by high cell-density fermentation of Pichia pastoris 147
Chapter 1

Introduction

Starch represents the major plant storage carbohydrate providing energy during the times of heterotrophic growth. Besides being an important storage carbohydrate for plants, starch represents a cornerstone for human and animal nutrition, and in addition starch is a feedstock for many industrial applications. The major source of starch for nutrition is cereal grains like: corn, rice, wheat, barley, and sorghum. In cereal grains starch constitutes up to 60% of the total dry weight. Starch is composed of glucose polymers (the α-glucans amylpectin and amylose) and occurs as semi-crystalline granules. Amylose is an essentially linear 1,4-α-glucan, while amylpectin, which typically accounts for 70% or more of the starch granule mass, is constituted by glucosyl residues linked by α-1,4-bonds to form chains that are branched via α-1,6-bonds (Figure 1.1)(Buléon et al., 1998). In barley, starch granules are stored in the seed endosperm. During germination the stored starch is mobilised by a concerted action of different amylolytic enzymes. The mobilisation involves an initial solubilisation, mainly by α-amylase, which is followed by hydrolysis of the resulting dextrans to oligosaccharides and glucose by the concerted action of α-amylase, β-amylase, limit dextrinase (LD), and α-glucosidase. Of these enzymes only LD has the capacity to hydrolyse α-1,6-glucosidic linkages in branched maltooligosaccharides (Burton et al., 1999; Kristensen et al., 1999). Since LD is the sole debranching enzyme in the germinating seed, it gets a key role in malting and brewing. But in barley seeds an endogenous inhibitor, limit dextrinase inhibitor (LDI) specifically inhibits LD and thereby suppresses the degradation of branched α-limit dextrans to fermentable sugars. The knowledge about the interaction between LD and LDI is sparse, even though their interaction has the above mentioned implications.
1.1 Limit dextrinase

Barley LD belongs to the large glycoside hydrolase family 13 (GH13) according to the Carbohydrate Active Enzymes (CAZy) database classification (http://www.cazy.org/). CAZy assigns carbohydrate active enzymes into sequence-based glycoside hydrolase (GH) families sharing structural fold and stereochemical mechanism (Cantarel et al., 2009). GH13, also known as the α-amylase family, is the largest GH-family in CAZy with more than 12,500 entries (i.e. sequences) assigned with 22 different enzyme substrate specificities towards α-glucosidic linkages, some even have both hydrolase and transglycosidase activities. GH13 groups together enzymes sharing sometimes only very limited sequence similarity, in fact only 4 short sequences are conserved between all the members (Table 1.1; MacGregor et al., 2001). The activities of the GH13 members are very diverse, but they are all retaining enzymes sharing the same catalytic residues Asp (nucleophile/base), Glu (proton donor), and a second conserved Asp that plays a role in the distortion of the substrate. Furthermore, they all share the same fold of the catalytic domain a (β/α)_{8}-barrel (Kuriki and Imanaka, 1999). GH13 has been divided into subfamilies, which are formed based on results of clustering, similarity searches and phylogenetic methods leading to subfamilies of enzymes where the correlation between sequence and enzymatic speci-
Table 1.1: Short conserved amino acid sequence motifs of the GH13 family. X, usually a hydrophobic residue; B, usually a hydrophilic residue; Z, a residue important for specificity (MacGregor et al., 2001)

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>XDXXXNH</td>
<td>GXRDXXXZZZ</td>
<td>XXXZZZZZZX</td>
<td>XXBBHD</td>
</tr>
<tr>
<td>Barley LD</td>
<td>398 MDVVYNH 404</td>
<td>469 GFRFDLMGH 477</td>
<td>506 YLYGEGGD 513</td>
<td>637 YVSAHD 642</td>
</tr>
</tbody>
</table>

Specificity has been taken into account (Stam et al., 2006). Limit dextrinases belongs to subfamily 13 (GH13_13) that contains LD-like debranching enzymes, known as pullulanases (PUL), from bacteria and eukaryotes. Enzymes with pullulanase activity from bacteria moreover are found in subfamilies 12 and 14 (Stam et al., 2006). The GH13 subfamilies 12–14 constitute the so-called GH13 PUL subfamily together with subfamilies 8–9 (starch branching enzymes), 10 (maltooligosyl trehalohydrolases), and 11 (isoamylases and glycogen debranching enzymes) (Janecek et al., 2007). The pullulanases of GH13 subfamilies 12–14 share some features, but vary in substrate specificity. Limit dextrinase shows highest activity towards limit dextrans – hence the name – derived from the degradation of amylopectin by α- and β-amylases. LD also has high activity towards pullulan (Figure 1.1), a linear polysaccharide composed of α-1,6 linked maltotriose units and produced by the yeast-like fungus *Aureobasidium pullulans* (Singh et al., 2008), why LD-like enzymes are also referred to as plant pullulanases (Manners and Yellowlees, 1971; MacGregor, 1987), although pullulan is most probably not a natural substrate for the plant enzymes. In general, the barley enzyme is called limit dextrinase, which in the biological context is the most describing name. But most other LD-like enzymes from plants are in the literature referred to as plant pullulanases.

The enzymatic nomenclature of pullulanases divides these into two groups based on specificity:

(i) pullulanase type I (EC 3.2.1.41) that exclusively cleaves the α-1,6-glucosidic linkages in pullulan and branched oligosaccharides, forming maltotriose and linear oligosaccharides, respectively, and

(ii) pullulanase type II, or amylopullulanase that hydrolyses both α-1,6- and α-1,4-glucosidic linkages in branched and linear oligosaccharides (Bertoldo and Antranikian, 2002; Doman-Pytka and Bardowski, 2004). Barley LD belongs to the pullulanase type I group based on its specificity.

The first plant debranching enzyme, which was found to act on amylopectin and its β-limit dextrans, was isolated in 1950–1951 from potato and broad bean (Hobson et al., 1951). At that time the enzyme was called the R-enzyme. Since then numerous plant pullulanases have been characterised from various sources, both seeds and leaves: barley (Manners and Yellowlees, 1971; MacGregor et al., 1994b), maize (Beatty et al., 1999; Wu et al., 2002; Li et al., 2009a; Dinges et al., 2003), mung bean (Morinaga et al., 1997), oat (Yamada, 1981; Dunn and Manners, 1975), pea (Zhu et al., 1998), rice (Yamasaki et al., 2008; Li et al., 2009a), sorghum (Hardie et al., 1976), spinach (Renz et al., 1998; Ludwig et al., 1984; Henker et al., 1998), sugar beet (Li et al., 1992), and wheat (Repellin et al., 2008). Several studies have focused on production in *Escherichia coli* of recombinant LDs from plants: maize (Wu et al., 2002), wheat (Repellin et al., 2008), and spinach (Renz et al., 1998). But the yield and quality of these recombinant enzymes were low, due to formation of inclusion bodies. Recently Vester-Christensen et al. established recombinant production of barley LD in reproducible and good yields (16 mg/l supernatant) and of high quality in the methylotrophic yeast *Pichia pastoris* (Vester-Christensen et al., 2010a), enabling growth of crystals suitable for high-resolution structure determination (Vester-Christensen et al., 2010b; see also section 1.1.3).
1.1.1 Catalytic mechanism of the GH13 family

GH13 enzymes are thought to catalyse hydrolysis of glucosidic linkages by a double displacement mechanism that involves an aspartic acid residue at the C-terminal end of $\beta$-strand 4 acting as the catalytic nucleophile and a glutamic acid residue at the end of $\beta$-strand 5 that plays the role of general acid/base catalyst. The enzymes in GH13 appear to require a third acid residue for catalytic activity; an aspartic acid situated a few residues beyond the C-terminal end of $\beta$-strand 7 (MacGregor, 2005; MacGregor et al., 2001). The cleavage point of the substrate is between the two glucose units covering subsites $-1$ and $+1$ (Figure 1.2) adopting the nomenclature for for subsites in glucosyl hydrolases (Davies et al., 1997). As shown in Figure 1.3, step one is a nucleophilic attack from a deprotonated aspartic acid resulting in a covalent $\beta$-glucosyl:enzyme intermediate. Thereafter a protonation of the leaving aglycone by a glutamic acid (the general acid/base) happens. Second: the deprotonated general acid/base acts as a catalytic base by activating the acceptor, a water molecule or a hydroxyl group, usually at C-4 or C-6 in a
glucose residue. The activated acceptor makes a nucleophilic attack on the anomeric carbon in the intermediate reforming the α-configuration either in a new glucosidic bond or a free sugar reducing end (MacGregor et al., 2001). In the double displacement mechanism, only two of the three conserved catalytic site residues directly play a role in the chemical reactions. The third conserved residue, a second aspartate, hydrogen bonds to the OH-2 and OH-3 groups of the substrate at subsite -1, and hereby plays an important role in the distortion of the substrate (Uitdehaag et al., 1999).

1.1.2 Substrate specificity of limit dextrinases/pullulanases

The substrate specificity of pullulanases varies depending on the source. An overview of the kinetics and specific activity of pullulanases from various sources can be found in Table 1.2 and Appendix I. In general, plant pullulanases show highest activity towards small branched oligosaccharides, with at least one α-1,4-glucosidic bond on either side of the α-1,6-glucosidic bond. The optimal substrate for barley LD is 6β-α-maltotriosyl-maltotetraose. As it appears from Table 1.2 both the length of the main chain, the length of the branch, and the position of the branch on the main chain is important for the hydrolytic rate. The activity drops when the substrate becomes more complex, the lowest activity is observed on amylopectin. The only exception is pullulan, where the hydrolysis rate is only 2–3 folds lower as compared with the optimal substrate. The plant pullulanases show no or very low activity toward the highly branched substrate glycogen. Most pullulanases derived from bacteria, on the other hand, show relatively good specific activity towards glycogen as compared with the activity on pullulan (Appendix I). In general, bacterial pullulanases are able to hydrolyse branched polymers at higher rates than plant pullulanases. The differences in substrate specificities can to a certain extent be referred to differences in the protein sequences. During recent years in total six structures from the three pullulanase subfamilies have been solved, including the crystal structure of barley LD. The structures have given insight into differences in the topology of the active sites, which can explain some of the differences in substrate specificities.
Table 1.2: Overview of the kinetic constants of a number of GH13 subfamilies 12–14 including a number of enzymes with no available, but which have been classified as pullulanase type I.

<table>
<thead>
<tr>
<th>GH13, 12</th>
<th>Source, enzyme type, accession no.</th>
<th>Substrate and assay conditions</th>
<th>$K_m$</th>
<th>$k_{cat}$</th>
<th>$V_{max}$</th>
<th>$k_{cat}/K_m$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus sp. KSM-1378, Amylopullulanase, AAS36537.1</td>
<td>pH 9.0, 50°C</td>
<td>Pullulan</td>
<td>0.72</td>
<td>169</td>
<td></td>
<td></td>
<td>Ara et al., 1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amylose</td>
<td>0.27</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus pneumoniae, glycogen-degrading virulence factor, SpuA, AAK74446.1</td>
<td>25°C</td>
<td>pNP-$\alpha$-maltopentaoside</td>
<td>0.36±0.06</td>
<td>4.5±0.2</td>
<td></td>
<td></td>
<td>van Bueren et al., 2011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GH13, 13</th>
<th>Source, enzyme type, accession no.</th>
<th>Substrate and assay conditions</th>
<th>$K_m$</th>
<th>$k_{cat}$</th>
<th>$V_{max}$</th>
<th>$k_{cat}/K_m$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barley (Hordeum vulgare), limit dextrinase (recombinant), AAD04189</td>
<td>pH 5.5, 37°C</td>
<td>Pullulan</td>
<td>0.08±0.003</td>
<td>61±13</td>
<td>753</td>
<td>(16±0.02)</td>
<td>Vester-Christensen et al., 2010a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.78 mM)</td>
<td>215</td>
<td>273</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barley (Hordeum vulgare), limit dextrinase, AAD04189</td>
<td>pH 5.0, 40°C</td>
<td>Pullulan</td>
<td>0.44</td>
<td>60.0</td>
<td>140</td>
<td></td>
<td>Greffe et al., 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6^3$-$\alpha$-maltotriosyl-maltotetraose</td>
<td>0.81</td>
<td>215</td>
<td>280</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6^4$-$\alpha$-maltosyl-maltotetraose</td>
<td>0.49</td>
<td>164</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6^4$-$\alpha$-maltosyl-maltotriose</td>
<td>0.47</td>
<td>85</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6^3$-$\alpha$-maltotriosyl-maltotetraose</td>
<td>0.78</td>
<td>218</td>
<td>279</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6^3$-$\alpha$-maltotriosyl-maltotriose</td>
<td>3.3</td>
<td>157</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\alpha$-cyclodextrin</td>
<td>0.087</td>
<td>10</td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\beta$-cyclodextrin</td>
<td>0.47</td>
<td>5</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\gamma$-cyclodextrin</td>
<td>0.19</td>
<td>39</td>
<td>205</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\alpha$-cyclodextrin</td>
<td>0.49</td>
<td>164</td>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\beta$-cyclodextrin</td>
<td>0.47</td>
<td>85</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\gamma$-cyclodextrin</td>
<td>0.78</td>
<td>218</td>
<td>279</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\alpha$-cyclodextrin</td>
<td>3.3</td>
<td>157</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\beta$-cyclodextrin</td>
<td>0.087</td>
<td>10</td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\gamma$-cyclodextrin</td>
<td>0.47</td>
<td>5</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$G_2$-$\alpha$-cyclodextrin</td>
<td>0.19</td>
<td>39</td>
<td>205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klebsiella pneumoniae (Aerobacter aerogenes), pullulanase type I, AAA25124</td>
<td>pH 5.0, 40°C</td>
<td>$6^3$-$\alpha$-maltotriose</td>
<td>0.98</td>
<td>2071</td>
<td>2100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source, enzyme type, accession no.</td>
<td>Substrate and assay conditions</td>
<td>$K_m$ (mg/ml)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>$V_{max}$ (ml/(mg s))</td>
<td>$k_{cat}/K_m$</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>---------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td><strong>Klebsiella pneumoniae</strong> (Aerobacter aerogenes), pullulanase type I, AAA25124</td>
<td>pH 5.5, 40$^\circ$C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yokobayashi et al., 1973</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pullulan</td>
<td>0.017</td>
<td>103.3</td>
<td>6076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Potato amylopectin</td>
<td>10.1</td>
<td>14.1</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sweet potato amylopectin</td>
<td>8.3</td>
<td>8.8</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wheat amylopectin</td>
<td>7.6</td>
<td>10.3</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maize amylopectin</td>
<td>9.9</td>
<td>8.6</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waxy maize amylopectin</td>
<td>9.8</td>
<td>9.0</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glutinous rice amylopectin</td>
<td>9.1</td>
<td>9.4</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waxy maize amylopectin, acid treated</td>
<td>8.9</td>
<td>19.9</td>
<td>2.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glutinous rice amylopectin, acid treated</td>
<td>7.7</td>
<td>13.7</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oyster glycogen</td>
<td>33</td>
<td>4.5</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rabbit liver glycogen</td>
<td>20</td>
<td>1.8</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sweet corn phytoglycogen</td>
<td>50</td>
<td>8.2</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Klebsiella pneumoniae</strong> (Aerobacter aerogenes), commercial pullulanase</td>
<td>pH 5.4, 37$^\circ$C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Malle et al., 2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pullulan</td>
<td>0.617</td>
<td>116</td>
<td>188</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maize (Zea maize) pullulanase, AAD11590</td>
<td>pH 6.0, 37$^\circ$C</td>
<td>mM</td>
<td>µmol/min</td>
<td></td>
<td></td>
<td>Wu et al., 2002</td>
</tr>
<tr>
<td></td>
<td>Pullulan</td>
<td>0.37±0.05</td>
<td>95.0±4.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-limit dextrin of amylopectin</td>
<td>4.10±0.95</td>
<td>344.4±55.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rice (Oryza sativa L. japonica), pullulanase, BAA28632</strong></td>
<td>pH 6.0, 37$^\circ$C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yamasaki et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pullulan</td>
<td>0.625</td>
<td>23.1</td>
<td>37.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amylopectin</td>
<td>1.538</td>
<td>2.18</td>
<td>1.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Soluble starch</td>
<td>1.538</td>
<td>2.89</td>
<td>1.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glycogen</td>
<td>0.909</td>
<td>3.31</td>
<td>3.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rice (Oryza sativa, L), limit dextrinase</strong></td>
<td>pH 5.0, 30$^\circ$C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dunn and Manners, 1975</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pullulan</td>
<td>-</td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amylopectin</td>
<td>6</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amylopectin β-limit dextrin</td>
<td>4</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rabbit-liver glycogen β-limit dextrin</td>
<td>7</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phytohexylase β-limit dextrin</td>
<td>10</td>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$6^\alpha$-α-maltotriosylmaltotetraose</td>
<td>0.67 (0.6 mM)</td>
<td></td>
<td></td>
<td>260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source, enzyme type, accession no.</td>
<td>Substrate and assay conditions</td>
<td>Km</td>
<td>$k_{cat}$</td>
<td>$V_{max}$</td>
<td>$k_{cat}/K_m$</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------------------</td>
<td>----</td>
<td>----------</td>
<td>----------</td>
<td>---------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Sorghum bicolor, limit dextrinase, ABK63595 (fragment)</td>
<td>pH 5.0, 30°C</td>
<td>mg/ml (mM)</td>
<td>relative to pullulan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.2 (0.4)</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td>Hardie et al., 1976</td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin β-limit dextrin</td>
<td>2.5 (1.5)</td>
<td>2.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycogen β-limit dextrin</td>
<td>30 (26)</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6³-α-maltotriosyl-maltotetraose</td>
<td>1.7 (1.5)</td>
<td>5.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinach (Spinacia oleracea), pullulanase, AAN26412</td>
<td>pH 6.0, 37°C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.78</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td>Ludwig et al., 1984</td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>7</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spinach (Spinacia oleracea), pullulanase (recombinant)</td>
<td>pH 6.0, 37°C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.70</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td>Renz et al., 1998</td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>7.2</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaerobranca gottschalkii, pullulanase type I (recombinant), AAS47565</td>
<td>pH 8, 60°C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td>Bertoldo et al., 2004</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.75</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus acidopullulyticus, pullulanase, CAC00156</td>
<td>pH 5.0, 70°C</td>
<td>mg/ml</td>
<td>U/min</td>
<td>$V_{max}/K_m$</td>
<td></td>
<td>Singh et al., 2010</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>4.0</td>
<td>0.23</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soluble starch</td>
<td>11.1</td>
<td>0.20</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dextran</td>
<td>40.0</td>
<td>1.25</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus acidopullulyticus, pullulanase, CAC00156</td>
<td>pH 5.0, 40°C</td>
<td>mM</td>
<td>s⁻¹</td>
<td>$1/(mM s)$</td>
<td></td>
<td>Jensen, 2004</td>
<td></td>
</tr>
<tr>
<td>6³-α-maltosyl-maltotetraose</td>
<td>0.3</td>
<td>164</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6³-α-maltosyl-maltopentaose</td>
<td>0.47</td>
<td>61</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6³-α-maltotriosyl-maltotetraose</td>
<td>0.23</td>
<td>87</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6³-α-maltotriosyl-maltopentaose</td>
<td>1.5</td>
<td>65</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G₂-α-cyclodextrin</td>
<td>0.21</td>
<td>8.0</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G₂-β-cyclodextrin</td>
<td>1.4</td>
<td>50</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G₂-γ-cyclodextrin</td>
<td>0.08</td>
<td>31</td>
<td>390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacillus subtilis strain 168, pullulanase type I/glycogen debranching enzyme, AAC00283</td>
<td>pH 5.4, 37°C</td>
<td>mg/ml</td>
<td>s⁻¹</td>
<td>ml/(mg s)</td>
<td></td>
<td>Malle et al., 2006</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>1.284</td>
<td>97</td>
<td>75.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fervidobacterium pennavorans Ven5, pullulanase type I, AAD30387</td>
<td>pH 6.0, 80°C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td>Bertoldo et al., 1999</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.4</td>
<td>1.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source, enzyme type, accession no.</td>
<td>Substrate and assay conditions</td>
<td>$K_m$</td>
<td>$k_{cat}$</td>
<td>$V_{max}$</td>
<td>$k_{cat}/K_m$</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Bacillus sp. AN-7, pullulanase type I</strong></td>
<td>pH 6, 80$^\circ$C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td>Kunamneni and Singh, 2006</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>1.3</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bacillus sp. S-1, alkaline pullulanase type I</strong></td>
<td>pH 9.0, 50$^\circ$C</td>
<td>mg/ml</td>
<td>mg/s/ml</td>
<td></td>
<td></td>
<td>Kim et al., 1993b</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>7.92</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>1.63</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha,\beta$-limited dextran</td>
<td>3.1</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bacillus flavocaldarius KP1228, thermostable pullulanase type I, BAB18516 (classified into GH13, but with EC no. 3.2.1.41)</strong></td>
<td>pH 6.8, 75$^\circ$C</td>
<td>(%)</td>
<td>1/(%)</td>
<td></td>
<td></td>
<td>Suzuki et al., 1991</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>1.7</td>
<td>48.5</td>
<td></td>
<td>28.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>1.8</td>
<td>12.6</td>
<td></td>
<td>7.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-limit dextrin</td>
<td>1.1</td>
<td>80.0</td>
<td></td>
<td>72.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soluble starch</td>
<td>13.0</td>
<td>51.2</td>
<td></td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Broad bean, Vicia faba L., limit dextrinase</strong></td>
<td>30$^\circ$C</td>
<td>mg/ml</td>
<td>Relative units</td>
<td></td>
<td></td>
<td>Gordon et al., 1975</td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>1.2</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin, $\beta$-limit dextrin</td>
<td>1</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit-liver glycogen $\beta$-limit dextrin</td>
<td>17</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6$^\alpha$-maltotriosyl-maltotetraose</td>
<td>0.18</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Clostridium thermohydro-sulfuricum (strain ATCC33223), pullulanase</strong></td>
<td>pH 6.0, 60$^\circ$C</td>
<td>mg/ml s$^{-1}$</td>
<td>ml/(mg s)</td>
<td></td>
<td></td>
<td>Saha et al., 1988</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.675</td>
<td>271</td>
<td></td>
<td>410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lactococcus lactis IBB 500, pullulanase</strong></td>
<td>pH 4.5, 45$^\circ$C</td>
<td>mg/ml</td>
<td>$\mu$M/(ml mg)</td>
<td></td>
<td></td>
<td>Wasko et al., 2011</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.34±0.02</td>
<td>2.04±0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oat (Avena sativa), limit dextrinase</strong></td>
<td>pH 5.0, 30$^\circ$C</td>
<td>mg/ml</td>
<td>relative to pullulan (%)</td>
<td></td>
<td></td>
<td>Dunn and Manners, 1975</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.17</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>1.4</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin $\beta$-limit dextrin</td>
<td>1.5</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit-liver glycogen $\beta$-limit dextrin</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytoglycogen $\beta$-limit dextrin</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6$^\alpha$-maltotriosyl-maltotetraose</td>
<td>0.23</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oat (Avena sativa), debranching enzyme</strong></td>
<td>pH 5.6, 30$^\circ$C</td>
<td>mg/ml</td>
<td>Relative (%)</td>
<td></td>
<td></td>
<td>Yamada, 1981</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.23</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutinous rice starch $\beta$-limit dextrin</td>
<td>3.00</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutinous rice starch</td>
<td>1.00</td>
<td>6.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source, enzyme type, accession no.</td>
<td>Substrate and assay conditions</td>
<td>$K_m$</td>
<td>$k_{cat}$</td>
<td>$V_{max}$</td>
<td>$k_{cat}/K_m$</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>------</td>
<td>---------</td>
<td>----------</td>
<td>-------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td><strong>Sugar beet, <em>Beta vulgaris</em> var. <em>altissima</em>, pullulanase from mature roots</strong></td>
<td>pH 5.6, 37°C</td>
<td>mg/ml</td>
<td></td>
<td></td>
<td></td>
<td>Masuda et al., 1987</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>4.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH 7.0, 55°C</td>
<td>% (w/v)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Odibo and Obi, 1988</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermoactinomyces thalophilus No. 15, pullulanase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-limit dextrin from amylopectin</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$-limit dextrin from glycogen</td>
<td>1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH 5.6, 70°C</td>
<td>mg/ml</td>
<td>$s^{-1}$</td>
<td>ml/(mg s)</td>
<td></td>
<td></td>
<td>Plant et al., 1987</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.07</td>
<td>$6.7 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-1,6-maltotriosyl-maltotriose</td>
<td>0.24</td>
<td>$1.8 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn amylopectin</td>
<td>0.03</td>
<td>$2.3 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potato amylopectin</td>
<td>0.01</td>
<td>$8.3 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit liver glycogen</td>
<td>0.04</td>
<td>$3.7 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oyster glycogen</td>
<td>0.01</td>
<td>$1.0 \times 10^5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lintner’s starch</td>
<td>0.05</td>
<td>$2.2 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-chain amylose</td>
<td>0.02</td>
<td>$1.5 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potato amylose</td>
<td>0.02</td>
<td>$2.6 \times 10^4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermus aquaticus YT-1 (strain ATCC25104)</td>
<td>pH 5.9, 70°C</td>
<td>K_m (%)</td>
<td>nmol/min/ml</td>
<td></td>
<td></td>
<td>Plant et al., 1986</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.0063</td>
<td></td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylopectin</td>
<td>0.0025</td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lintner’s starch</td>
<td>0.0034</td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermus caldophilus GK-24, pullulanase type I</td>
<td>pH 7.0, 73°C</td>
<td>mg/ml</td>
<td>U/mg</td>
<td></td>
<td></td>
<td>Kim et al., 1996</td>
<td></td>
</tr>
<tr>
<td>Pullulan</td>
<td>0.42</td>
<td></td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Results in brackets are based on the Michaelis-Menten equation for uncompetitive substrate inhibition.

$^b$ 0.077 µg pullulanase with pullulan and 31 µg with other substrates

$^c$ Molarity expressed with respect to $\alpha$-1,6-glycosidic linkages using chain length of 10 for amylopectin $\beta$-limit dextrin and 6 for glycogen $\beta$-limit dextrin.

$^d$ Initial velocity at 40 mg/ml amylopectin – the conc. needed to saturate the enzyme.
1.1.3 Structures of pullulanases

In 2010 two crystal structures of barley LD were published in complex with its competitive inhibitors; α- and β-cyclodextrin (α- and β-CD) (Figure 1.4; Vester-Christensen et al., 2010b) as a result of the successful production of recombinant (Vester-Christensen et al., 2010a). These structures provided the first insight into the active site topology of a plant pullulanase and they gave rise also to further questions regarding substrate binding. Recently, the structure of free, uncomplexed LD was published (Møller et al., 2012a). In 2006 the first GH13_13 pullulanase was structure determined, i.e. the pullulanase type I from Klebsiella pneumoniae (KpPUL) (Mikami et al., 2006). In total six different crystal structures of KpPUL were published: The enzyme without ligand and the enzyme in complex with, glucose, maltose, isomaltose, maltotriose, or maltotetraose (see Table 1.3 for PDB entries). These structures, especially the complex with maltotetraose, illustrated details of ligand binding in debranching enzymes, as two maltotetraose molecules were accommodated at the active site in a manner assumed to mimic a branched substrate, i.e. a substrate consisting of a maltotetraose main chain with a maltotetraose as branch. In addition to the crystal structures of the two enzymes from GH13_13, two pullulanases from GH13_12 and two pullulanases from GH13_14 have been structure determined (see Table 1.3 for overview of the published pullulanase structures). The two GH13_12 pullulanases are both from the Streptococcus genus; S. agalactiae and S. pneumoniae. The two enzymes from streptococci function as virulence factors by binding α-glucans and degrading glycogen. They are cell wall-attached multi-modular proteins (Table 1.4) (Gourlay et al., 2009; van Bueren et al., 2011).

Table 1.3: Overview of the published crystal structures of pullulanases from the three GH13 subfamilies 12, 13, and 14.

<table>
<thead>
<tr>
<th>GH subfamily</th>
<th>Organism</th>
<th>Enzyme</th>
<th>Reference</th>
<th>PDB entries</th>
<th>Ligand(s) in active site</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>13_12</td>
<td>Streptococcus agalactiae</td>
<td>PUL type I (SaPUL)</td>
<td>Gourlay et al., 2009</td>
<td>3FAW 3FAX</td>
<td>None  β-CD</td>
<td>Maltotetraose Structures lack the N-terminal domain containing a CBM41</td>
</tr>
<tr>
<td></td>
<td>Streptococcus pneumoniae</td>
<td>Glycogen-degrading virulence factor (SpPUL) (PUL type I)</td>
<td>van Bueren et al., 2011</td>
<td>2YA0 2YA1 2YA2</td>
<td>None  2×maltotetraose</td>
<td>G-moranoline&lt;sup&gt;a&lt;/sup&gt; 2YA0+2YA2: lack CBM41-1+2 and N-domain 2YA1: lacks N-domain</td>
</tr>
<tr>
<td></td>
<td>Hordeum vulgare, barley</td>
<td>Limit dextrinase (plant PUL type I (LD))</td>
<td>Møller et. al., 2012; Vester-Christensen et al., 2010b</td>
<td>2FGZ 2FH6 2FH8 2FHB 2FHC 2FHF</td>
<td>None  β-CD  α-CD</td>
<td>Glucose Isomaltose 2×maltose 2×maltotriose 2×maltotetraose</td>
</tr>
<tr>
<td>13_13</td>
<td>Klebsiella pneumoniae</td>
<td>PUL type I (KpPUL)</td>
<td>Mikami et al., 2006</td>
<td>2E8Y 2E8Z 2E9B</td>
<td>None  α-CD</td>
<td>Only preliminary X-ray analysis is described</td>
</tr>
<tr>
<td></td>
<td>Bacillus acidopullulolyticus</td>
<td>PUL type I (BaPUL)</td>
<td>Turkenburg et al., 2009</td>
<td>2WAN</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bacillus subtilis subsp. subtilis str. 168</td>
<td>PUL type I (Glycogen debranching enzyme, BsPUL)</td>
<td>Malle et al., 2006</td>
<td>2E8Y 2E8Z 2E9B</td>
<td>None  α-CD  Maltotriose</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>4-O-α-D-glucopyranosylmoranoline
Figure 1.4: Top: Domain organisation of the six structure-determined members of GH13_12, GH13_13, and GH13_14. The colours refer to: N-domains with unknown function (orange), CBMs (green), catalytic-domain (blue) and C-domain (grey). The amino acid numbering and the annotation of domains are based on information from literature references associated with each structure (see Table 1.3 for references). As there is no publication for BsPUL, the annotation is based on information from www.pdb.org and structural alignments to the five other pullulanase structures.

Bottom: Structures of the six structure-determined pullulanases. The domains are coloured according to the color code defined for the domain organisation diagram. The structures of BaPUL and SaPUL lack one and two CBM41, respectively.
The two pullulanases from GH13 with known structure are both from the Bacillus genus; *B. acidopullulyticus* (Turkenburg et al., 2009) and *B. subtilis* (Malle et al., 2006). The pullulanases from bacilli are industrially important enzymes. These pullulanases find use for the hydrolysis of α-1,6 linkages in amylpectin and limit dextrins. *B. acidopullulyticus* pullulanase (*BaPUL*) is thus used commercially in the starch industry in the production of high fructose corn syrup (Norman, 1982) and in the production of high maltose content syrups, as well as in the brewing industry, especially in the production of low calorie and "light" beers, where it allows more complete fermentation of the mash resulting in less residual sugar (Turkenburg et al., 2009).

The pullulanases from the three subfamilies are multi-modular proteins (Figure 1.4). Besides the catalytic domain with the common fold shared by all GH13 members, *i.e.* a (β/α)₈-barrel (Kuriki and Imanaka, 1999), the pullulanases have one or more N-terminal domains and a C-terminal domain. The function of some of the N-terminal domains are unknown (Møller et al., 2012a), but they share fold with characterised carbohydrate binding modules (CBMs). The N-domain of LD has homology to CBM21 (Møller et al., 2012a). A CBM is defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discreet fold having carbohydrate-binding activity (www.cazy.org/Carbohydrate-Binding-Modules). The recognised functions of these auxiliary domains are to bind polysaccharides, bringing the catalytic domain into close and prolonged vicinity with its substrate, and thus advancing carbohydrate hydrolysis. CBMs are grouped into 64 families (according to CAZy, Cantarel et al., 2009), which show different polysaccharide binding specificities. Such recognition of polysaccharides by CBMs plays a role in relation to metabolism, pathogen defence, polysaccharide biosynthesis, virulence, plant development, etc. (Guillen et al., 2010; Boraston et al., 2004). LD has a CBM48 (Figure 1.4), a CBM occurring with the GH13 enzymes: pullulanase, isoamylase, maltooligosyl trehalohydrolase, as well as glycogen branching enzyme and starch branching enzyme (Machovic and Janecek, 2008).

Despite the overall structural similarity of the structures of the six pullulanases, the sequence identities and similarities are low, even for the catalytic domains. At the level of the catalytic domain LD and the other GH13 pullulanase, *KpPUL*, have a sequence identity and similarity of 40% and 59%, respectively, while the identities and similarities between LD and the four other structure-determined pullulanases from subfamilies GH13_12 and GH13_14 are lower (identities; 25–31%, and similarities; 37–45%).

Besides the structures from the three GH13 subfamilies containing pullulanases/limit dextrinases several structures are solved from other GH13 subfamilies (Stam et al., 2006), which contain enzymes hydrolysing α-1,6-glucosidic linkages, *i.e.* isoamylases and glycogen debranching enzymes (GH13_11), and oligo- and glucan-1,6-α-glucosidases (GH13_31). Structures moreover are available from subfamilies 8 and 9 with branching enzymes, *i.e.* enzymes which hydrolyse α-1,4-glucosidic linkages and form α-1,6-linkages. All these structures of enzymes with somehow comparable substrates, but with very different specificity, can add some important information to the understanding of the elements in the active site of pullulanases and limit dextrinases, which determine the specificity of these enzymes. The differences and similarities between structures of the GH13_12–14 enzymes and of those from the subfamilies mentioned above will be further discussed in section 3.3.2.
1.1.4 Transglycosylation by LD

Among the GH13 enzymes predominantly functioning as hydrolases, it is not uncommon to find some degree of transglycosylation activity, i.e., they transfer glucose or small oligosaccharides from one glucoside to another resulting in the formation of an α-glucosidic linkage. McDougall et al. have shown that barley LD has endotransglycosylase activity. Linear maltodextrins in the range from maltose to maltoheptaose have been shown to activate LD until a certain concentration, 5–12.5 mM depending on length of the maltodextrin, above that LD activity is inhibited, when activity is measured by the Limit-Dextrizyme assay (azarine-crosslinked-pullulan). This activation also happened when using Red Pullulan (Procion Red-crosslinked-pullulan) as substrate and maltotriose as activator (McDougall et al., 2004). This result confirmed previous finding by MacGregor et al., that barley LD was activated by maltodextrins derived from malt extract, when activity was assayed using dyed pullulan substrates. The activation, however, was not observed when β-limit dextrin, a natural substrate, was used for assaying activity (MacGregor et al., 2002). These results led to the hypothesis that LD was able to transglycosylate. MacDougall et al. used an assay, with pullulan as substrate and a fluorescent maltotriose derivative (maltotriose-8-amino-1,3,6-pyrene trisulphonic acid; G3-APTS) as activator, resulting in LD catalysed transfer of G3-APTS to form products of higher molecular weight (Figure 1.5). These products were degraded by LD into fluorescent products.

Figure 1.5: Acceptors, donors, and products from transglycosylation reactions. A) The acceptor (G3-ATPS; ATPS is symbolised by a red P) and donor (pullulan) from the transglycosylation by LD (McDougall et al., 2004). B) The structure of aesculin used as acceptor in the transglycosylation with pullulan as donor by the type I PUL from Thermotoga neapolitana. This reaction resulted in maltotriosyl-aesculin as determined by NMR (C) (Kang et al., 2011). D) 6-O-α-maltosyl-β-CD.
LD from rice is activated by linear maltodextrins and also catalysed transglycosylation (McDougall et al., 2004), and transglycosylation products are even formed from a reaction mixture with only pullulan (Yamasaki et al., 2008). The transglycosylation capacity of barley LD has been further investigated using α-maltosyl fluoride (G2F) and α-maltotriosyl fluoride (G3F) as donors and linear maltoligosaccharides (maltotriose through maltohexaose) or α- or β-CD as acceptors (Vester-Christensen, 2009). The formation of transglycosylation products was followed by thin layer chromatography (TLC) and selected products were identified by mass spectrometry. The reactions with the linear maltoligosaccharides as acceptors resulted in small branched products, some of which were even double substituted. α- and β-CD were very good acceptors for the transglycosylation reaction, explained by the high affinity of these CDs to LD (Kd of α- and β-CD is 27.2 and 0.7 µM, respectively; Vester-Christensen et al., 2010a), and the positions of CDs as seen in the structures leave space for simultaneously binding of G2F in the active site. The products from the reaction between G2F and the β-CD were mono- through tetra-substituted β-CD (Figure 1.5D), while the reaction between α-CD and G2F resulted in accumulation of G2-α-1,6-α-CD produced in a reaction with low transfer rate (Vester-Christensen, 2009).

The production of α-1,6-substituted CDs by transglycosylation catalysed by pullulanase has also been reported for *B. acidopullulyticus* (Kitahata et al., 1987) and *K. pneumoniae* (Yoshimura et al., 1988). The bacterial pullulanases from e.g. *B. acidopullulyticus* and *K. planticola*, on the other hand, did not catalyse transglycosylation of G3-APTS to pullulan (McDougall et al., 2004), emphasising that the transglycosylation activity is dependent on donor and acceptor. Several bacterial pullulanases have been used to synthesise compounds with α-1,6-linkages at very high concentrations of reactants, e.g. pullulanase from *Aerobacter aerogenes* (*Klebsiella pneumoniae*), which synthesised branched β-CD from 4β-O-β-D-galactosyl-maltose (1.6 M) and β-CD (0.16 M) (Kitahata et al., 2000). Furthermore, a PUL type I from the hyperthermophilic bacterium *Thermotoga neapolitana* has been shown to have α-1,6-transfer activity (Kang et al., 2011). The product after reaction of this PUL with pullulan and aesculin (Figure 1.5B) was determined to be α-maltotriosyl-(1,6)-aesculin (Figure 1.5C) by using NMR and isoamylase treatment.

### 1.2 Limit dextrinase inhibitor

More than two decades ago, when the thorough research began on barley LD, it was found that in germinating seeds LD occurs in a free, active form and as a bound, inactive form. The nature of this inactive form was not known, but it was reported that long extraction time and reducing agents, like cysteine or dithiothreitol (DTT), or proteases were needed to extract active barley LD (McCleary, 1992). The different observations led to the hypothesis that inactive LD was bound to a protein. In 1993 Macri et al. showed the presence of a proteinaceous inhibitor named limit dextrinase inhibitor (LDI). Two forms of LDI were found, which differed in isoelectric point (pI) (MacGregor et al., 1993). Partially purified extracts of barley or malt were later shown to contain 3–4 LD inhibitory protein species. But only the two LDI forms present in highest amount were isolated: a low pI form (pI=6.7) and a high pI form (pI=7.2). Both forms were equally effective at inhibiting LD (MacGregor et al., 1994b). Amino acid sequencing showed both forms to have the same sequence, but to differ in modification at a cysteine thiol group (MacGregor et al., 2000).
LDI from barley is the only characteised proteinaceous inhibitor of a debranching enzyme even though numerous pullulanase-like enzymes have been characteised from both seeds and leaves from various plants as described in section 1.1. An early study on limit dextrinase/pullulanase from mature cereal grains; oat, wheat, barley, and rye indicated that the mature grains could contain a proteinaceous inhibitor against LD, since the pullulanase activity in flour samples increased after incubation with sodium dithionite or papain (Yamada, 1981). In the same study oat pullulanase was purified to homogeneity from flour in buffer with sodium dithionite, this resulted in much higher specific activity as compared with oat pullulanase purified earlier without any reducing agent present (Dunn and Manners, 1975).

Until the establishment of production of recombinant LD (Vester-Christensen et al., 2010a) and LDI (Jensen et al., 2011), no details were known of their interaction except for the 1:1 stoichiometry of the complex determined by electrospray time-of-flight mass spectrometry (MacGregor et al., 2003). In addition, the effectiveness of the inhibition was shown to vary with pH, being strongest at pH 5.5-6.5, which is the pH optimum range of the LD activity. The inhibitor activity was weak at pH 4.5 and 8 (MacGregor et al., 1994b). Furthermore, LDI was shown to be heat stable, but susceptible to protease digestion (Macri et al., 1993). No quantitative determination of the strength of the interaction or type of inhibition had been carried out (MacGregor, 2004). The effect of LDI on various debranching enzymes was tested, since related enzyme inhibitors are thought to play a defensive role against extra-cellular enzymes from parasitic insects and fungal pathogens. But barley LDI was reported to be relatively ineffective in inhibiting pullulanase and isoamylase from Aerobacter aerogenes, isoamylase from Pseudomonas amyloferonnaosa (MacGregor et al., 1994b), and pullulanase from B. acidopullulyticus (Bryce et al., 1995). Although LDI is a member of the CM-protein family (see Section 1.2.1 below), which contains known inhibitors of α-amylases and/or trypsin, LDI showed no inhibitory activity against trypsin and barley, insect or mammalian α-amylases (MacGregor et al., 2000).

1.2.1 CM proteins and cereal-type inhibitors

Inhibitors of GH13 enzymes, mainly α-amylases, are widespread in plants, where they act as a defence system against pests. These α-amylase inhibitors can be classified according to their tertiary structure in six different types, namely: lectin-like, knottin-like, cereal-type, Kunitz-like, γ-purothionin-like and thaumatin-like inhibitors (Franco et al., 2002). α-amylase inhibitors from the group of cereal-type inhibitors are known to inhibit enzymes from GH13, notably α-amylases from mammals and insects. LDI is the only studied member of this family that inhibits a debranching enzyme. The cereal-type inhibitors are members of a larger family of proteins, which can be extracted from flour by chloroform-methanol (CM), why this family of proteins is called CM-proteins (Svensson et al., 2004). Included in this family are lipid transfer proteins, hybrid proline-rich proteins, the seed storage proteins like the glutamine-rich proteins 2S-albumins, and the cereal-type α-amylase/trypsin inhibitors (José-Estanyol et al., 2004). In addition, different CM-proteins may contain one or two more cysteines, which are conserved only within the individual CM-protein group. This is the case for the cereal-type inhibitors, which makes it possible to distinguish these from the other CM-proteins (Figure 1.6).

Three-dimensional structures have been determined for three members of the cereal-type in-
Figure 1.6: Multiple sequence alignment of a number of CM-proteins generated using PROMALS3D (Pei et al., 2008b) and visualised using ESPript (Gouet et al., 1999). The cereal type inhibitors are marked with asterisks (LDI, Q2V8X0; RBI, P01088; AI 0.19, BAA20139; CHFI, P01088). See Appendix II.1 for the list of all proteins included. Cysteines are marked by green boxes, except for the completely conserved cysteines, which have red background. The alignment is modified from (José-Estanyol et al., 2004).
1.3 LD and LDI in barley

1.3.1 Synthesis and localisation of LD and LDI in barley seeds

During the last two decades insight has been gained on the synthesis and localisation of LD and LDI, but detailed knowledge is still limited. The synthesis and localisation of LD in barley have been studied utilising different systems like, e.g. developing barley kernels (Sissons et al., 1993), de-embryonated half-kernels (Schroeder and MacGregor, 1998), and germinated barley seeds (Schroeder and MacGregor, 1998). The determination of total LD levels in mature barley and germinating barley is challenging as the determination is hampered by the presence of LDIs (Schroeder and MacGregor, 1998), necessitating long extractions to release enzyme from inhibitor (MacGregor et al., 1994a). The extent of the problem depends on the experimental setup, but during extraction LD and LDI can interact, like it happens during mashing (Schroeder and MacGregor, 1998). Besides issues with the extraction, as barley germinates the level of inhibitor diminishes, adding a further complication to the determination of the total LD content and activity in the period of germination (Schroeder and MacGregor, 1998).

It has been shown that a single gene encoding LD is expressed in the aleurone layer (1.7) during germination and to a lower extent in the immature kernels as shown by northern-blot analysis of extracted mRNA from developing and mature barley kernels (Burton et al., 1999; Kristensen et al., 1999). The transcription of the LD gene seems to reach a maximum after five days of germination, and thus the peak of translation is also somewhat later in germination as compared with α-amylase (Kristensen et al., 1999). Furthermore, response elements for the plant hormone gibberellic acid (GA), which stimulates the cells of germinating seeds to produce mRNA molecules that code for hydrolytic enzymes, were found in the promoter region of the LD gene (Burton et al., 1999). A study has, however, shown that LD development was possible without the addition of GA, and that the maximum level of LD activity was similar whether or not GA was present, but the maximum level of activity was achieved sooner when GA was present (Schroeder and MacGregor, 1998). LD was present before treatment with GA in the study of de-embryonated half-seeds at a level up to 15% of the maximum achieved during incubation of the half-seeds with GA (Schroeder and MacGregor, 1998). It was suggested that it reflects LD being present in ungerminated barley (as shown by Manners and Yellowlees, 1973; Sissons et al., 1993). Early studies using an LD-immunoassay showed that LD is present in embryo and endosperm of maturing barley kernels (Sissons et al., 1993). During kernel development the amount of immunoreactive free LD increased six days post anthesis (dpa), reached a maximum at 21 dpa, and then decreased rapidly, resulting in an activity level of free LD in mature kernels (42 dpa), which is slightly above the level 6–8 dpa. The total amount of LD remained at the

![Figure 1.7: Overview of barley seed compartments.](image-url)
maximum level (Sissons et al., 1993), and LD was therefore suggested to be converted to a bound form, which most likely is the complex with LDI. Like LD, two of the three other enzymes involved in the complete degradation of starch during germination, namely α-amylase and α-glucosidase, are produced in the aleurone layer (and scutellum) in response to GA, and subsequently released into the starchy endosperm (Finnie et al., 2011). β-amylase, however, is synthesised and stored in the starchy endosperm during grain filling (Hara-Nishimura et al., 1986). Unlike α-amylase and α-glucosidase, LD is only slowly released from the aleurone layer, which may indicate that secretion of LD is controlled by a different mechanism than the one controlling secretion of the other hydrolytic enzymes (Schroeder and MacGregor, 1998). This observation fits well with the fact that LD is first really needed when most of the starch reserves have been mobilised (Schroeder and MacGregor, 1998).

1.3.2 Biological roles of LD and LDI

Today it is widely accepted that LD, in particular barley LD, contributes to the degradation of starch during germination, while the function of the enzyme during biosynthesis is subject to more debate. Starch research traditionally recognises two sorts of starch: transitory starch and storage starch. Transitory starch accumulates in leaves during the day and is degraded at night, while storage starch accumulates over longer periods, often in specialised storage organs like potato tubers, cereal grains or pea seeds. The stored starch is degraded to provide carbon and energy for re-growth or germination after periods of dormancy (Smith, 2012). Arabidopsis has been used as a model plant for the research in the degradation of starch in leaves, while cereal grains have been the model for storage starch degradation in seeds. In both systems LD/PUL mutants have been constructed.

Barley LD has been shown to be expressed during seed development (Sissons et al., 1993; Burton et al., 1999) and its debranching activity has been postulated to play a role in the control of amylopectin biosynthesis in monocots, e.g. maize and rice (Pan and Nelson, 1984; Nakamura et al., 1997). Stahl et al. showed by antisense down-regulation of barley LDI shown, that the presence of LDI has an impact on the starch granule size distribution, starch composition and amylopectin structure. The amylopectin chain length distribution was changed towards less long chains (>25 units) and enhanced number of medium long chains (10–15 units), when LDI was down-regulated. Furthermore, the ratio between A- and B-type starch granules was changed from being 1:20 (A:B) in wild type to 1:3 in the plants with reduced LDI levels (Stahl et al., 2004). The A-type granules are large (15–32×6–10 µm) and disk-shaped and appear four dpa and continue to increase in size throughout the grain filling period, while B-type granules are small (2–3 µm) and spherical and are initiated 12–14 dpa and remain considerably smaller (Jane et al., 1994). Gene expression and activity for both LD (Burton et al., 1999) and LDI occur in the endosperm 2–4 weeks post anthesis (Stahl et al., 2007), and thus correlates with the second wave of nucleation of B-type granules (Stahl et al., 2004). These results support the hypothesis that LD plays a role during biosynthesis of starch. Biosynthesis, however, is much more complex as compared with starch degradation, since several types of debranching enzymes (see below) are involved.

In total four classes of enzymes: ADP-glucose pyrophosphorylase (AGPase), soluble starch synthase (SS), starch branching enzyme (BE), and starch debranching enzyme (DBE, α-1,6-glucan hydrolase), are involved in amylopectin synthesis in higher plants (Table 1.4). The forth group, DBE, has been shown to play a role in starch biosynthesis of crystalline amylopectin. All plants characterised to date contain two conserved types of starch debranching enzymes: pullulanase-type (LD) and isoamylase-type (ISA, EC 3.2.1.68) of which at least three genes are
Table 1.4: Overview of the enzymes involved in biosynthesis of branched α-glucans in different organisms (adapted from Nakamura, 2002)

<table>
<thead>
<tr>
<th></th>
<th>Plants</th>
<th>Bacteria</th>
<th>Animals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Polyglucan</strong></td>
<td>Amylopectin</td>
<td>Glycogen</td>
<td>Glycogen</td>
</tr>
<tr>
<td></td>
<td>(4–5.5% branches,</td>
<td>(8–10% branches,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>chain length 18–25</td>
<td>chain length 12–15</td>
<td></td>
</tr>
<tr>
<td>residues&lt;sup&gt;a&lt;/sup&gt;)</td>
<td></td>
<td>residues)</td>
<td></td>
</tr>
<tr>
<td><strong>Donor formation</strong></td>
<td>AGPase</td>
<td>AGPase</td>
<td>UGPase</td>
</tr>
<tr>
<td><strong>Chain Elongation</strong></td>
<td>SSI</td>
<td>Glycogen synthase</td>
<td>Glycogen synthase</td>
</tr>
<tr>
<td></td>
<td>SSII</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSIII</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Branching</strong></td>
<td>BEI</td>
<td>Glycogen branching enzyme</td>
<td>Glycogen branching enzyme</td>
</tr>
<tr>
<td><strong>Debranching</strong></td>
<td>ISA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PUL(?)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>From (Nakamura, 1996)

The physiological function of ISA in biosynthesis is well established based on mutants deficient in one or more of the ISA isotypes of barley (Burton et al., 2002), Arabidopsis (Wattebled et al., 2005, 2008; Streb et al., 2008), maize (James et al., 1995), and rice (Nakamura et al., 1989; Fujita et al., 2009; Li et al., 2009a; Fujita et al., 2003). Results from these studies show that a lack of ISA will result in the formation of the disordered water-soluble polysaccharide, phytoglycogen. Phytoglycogen resembles glycogen in structure, but it lacks the structural organization that allows crystallisation into insoluble granules. Generally seen, ISA is assumed to function in the editing of excessively branched chains or in removing improper branches of amylopectin formed by branching enzymes in order to maintain the cluster structure of amylopectin during starch biosynthesis (Buléon et al., 1998; Jeon et al., 2010; Nakamura, 2002). Compared with ISA, the physiological function of the pullulanase-type DBE in starch biosynthesis is less well established. Besides during germination, substantial pullulanase activity can be detected in the developing in rice and maize endosperm (Beatty et al., 1999; Nakamura et al., 1996), and the presence of mRNA for barley LD has been detected as well (Burton et al., 1999).

In rice, the *Pul* gene encoding a pullulanase is highly expressed during the entire period of seed development, with peaks in the middle and late stages (Li et al., 2009a; Fujita et al., 2009), suggesting that pullulanase could be involved in starch biosynthesis. The amylose content, seed morphology, and morphology as well as crystallinity of starch granules of rice *Pul* mutant lines were almost the same as those of the wild type. But slight changes in the chain-length distribution of the mutant lines were detected. Rice *Pul*-deficient mutant lines show increased short-chains of DP ≤ 12 to a lesser extent than in the *Isa* mutant line (Fujita et al., 2009).

A complete loss of pullulanase-type DBE activity in maize did not result in any obvious morphological differences in the leaf or the kernel, while a reduced rate of endosperm starch mobilization during seedling establishment was observed. This emphasises that the pullulanase-type DBE is required for normal starch degradation (Dinges et al., 2003). In maize mutant lacking pullulanase activity accumulation of branched maltoligosaccharides was seen in the developing
endosperm, this was not seen in wild-type. Furthermore, in a background deficient in ISA the lack of pullulanase activity resulted in significant accumulation of phytoglycogen in the kernel, and worse than when only Isa is defect. This may indicate that pullulanase in maize also partly compensates for the defect in Isa and functions during starch biosynthesis as well as degradation (Dinges et al., 2003).

Analysis of Pul single-mutant lines of Arabidopsis did not lead to a distinct phenotype. However, Isa2-Pul-defective double mutant lines display a 92% decrease in starch content. This suggests that the function of PUL in Arabidopsis partly overlaps with that of Isa1, although its implication remains negligible when ISA1 is present within the leaf cell (Wattebled et al., 2005). Further mutational studies with Isa3 and Pul-defective mutant lines and the double mutant showed that the Isa3/Pul double mutant has a more severe starch-excess phenotype and a slower rate of starch breakdown than Isa3 single mutants. The double mutant accumulates branched oligosaccharides (limit dextrins) that are undetectable in the wild-type and the single mutants (Delatte et al., 2006). Based on this it was concluded that ISA3 and PUL have redundant function for degradation of transitory starch in Arabidopsis leaves (Delatte et al., 2006).

In addition, it was suggested that PUL is involved in starch degradation based on an observed higher rate of starch degradation at night in Isa3 mutant lines as compared to the Isa3-Pul double mutant (Wattebled et al., 2008).

The different mutational studies of the DBEs in different higher plants display distinctive expressivities of mutant phenotypes leading to contradictory interpretations. One explanation for these small, but significant differences can be found if different DBEs display some level of functional overlap. In that case if the balance between the DBEs or their degree of functional redundancy varies from one species to another these small differences might be explained and the contradictions may be resolved (Wattebled et al., 2008).

1.3.3 **In vivo regulation of LD and LDI**

The presence of the endogenous LD inhibitor (LDI) is likely to be partly responsible for the low levels of LD activity in germinating seeds and malt (MacGregor, 2004), but is probably also important for controlling LD activity in starch synthesis during grain filling (Stahl et al., 2004), as down-regulation of LDI result in changed starch granule size distribution, starch composition, and amylopectin structure (Stahl et al., 2007, 2004), as mentioned in the previous section. The in vivo regulation of LD is connected directly to LDI. Overexpression of wheat thioredoxin h (Trxh) in the endosperm of transgenic barley grain resulted, however, in up to 4-fold increase in LD activity (Cho et al., 1999). The increased LD activity was significantly higher as compared with the control even after 3 days of germination, where LDI was shown to be inactive. This led to the hypothesis that the increase in LD activity when Trxh is overexpressed is not caused by enhanced inactivation of LDI, but could be due to increased de novo synthesis of LD or decreased binding of the mature enzyme to the starch endosperm (Cho et al., 1999). A similar experiment has been done in wheat, where Trxh was under- and overexpressed, respectively. Increased Trxh level resulted in increased activity of LD throughout the germination period, while LD activity was marginally suppressed for 2 days after seed imbibition when Trxh was underexpressed, but then rose to a level roughly equal to that of the control (Li et al., 2009b). LD activity was furthermore lowered by 30% during seed development, when Trxh was underexpressed. It persisted from 30 to 70 dpa (Li et al., 2009b). Even though increased activity of barley LD during germination cannot solely be ascribed to inactivation of LDI when wheat Trxh is over-expressed there is in vitro evidence, that barley Trxh can inactivate LDI by reducing the structure stabilising disulphide bonds following the reaction schematised in Figure 1.8 (Jensen
et al., 2012). Maize Trxh in combination with DTT has been shown to increase the activity of recombinant maize LD by 2-fold, while the two reducing agents DTT and Trxh alone did not affect the activity significantly (Wu et al., 2002).

A recent study of redox sensitive starch metabolising enzymes in Arabidopsis thaliana showed that LD was dependent on DTT to show full activity, while using CuCl₂ as oxidising agent caused a decrease in activity to less than 10% of the maximum activity. The loss of activity could be completely recovered by subsequent treatment with DTT, thus the oxidation-reduction reaction was shown to be reversible (Glaring et al., 2012). DTT treatment of recombinant barley LD, on the other hand, resulted in only slightly reduced activity (~90% residual activity) (Jensen et al., 2012). Spinach and wheat LD have also been identified as redox sensitive in vitro (Schindler et al., 2001; Repellin et al., 2008). Activation of spinach LD by reduction resulted in a decrease in the number of isomeric forms. But unlike barley LD, spinach leaf LD could not be activated by thioredoxin, but it was activated by reduced glutathione (GSH). Based on this observation it was suggested that since the spinach leaf LD is functioning in the chloroplastic starch metabolism, reduction e.g. by GSH will result in maintenance of a specific set of isoforms of LD, and at the same time extend the pH range of activity beyond pH 7, allowing enzyme activity during the dark period (pH 7), but not during daytime (pH 8). The oxidized LD did not show activity in the range beyond pH 7 (Schindler et al., 2001). Three cysteines of spinach LD were proposed to be involved in the redox sensitivity and additionally in isoform microheterogeneity (Cys390, Cys452, and Cys677) (Schindler et al., 2001). These three cysteines are conserved in Arabidopsis LD. But since only one of the cysteines is conserved between both Arabidopsis, spinach, wheat LD, and barley LD, it was suggested that there could be a different mechanism of redox sensitivity in the LD enzymes (Glaring et al., 2012).

1.4 The LD:LDI system in malting and brewing

The presence of one or more starch-debranching enzymes in malt has been recognised for more than 50 years, but the importance of the debranching activity in brewing is still under debate. Research on plant debranching enzymes in 1970–1975 demonstrated that only one debranching enzyme is present in malt, namely LD. The major problem about the early studies on the importance for LD in brewing is that different extraction procedures are used for each of the malting enzymes, α- and β-amylase, LD, and α-glucosidase), as well as the assay methods. Most of the assays are based on carbohydrate-dye complexes. The substrates are specific for each individual enzyme, but may be inadequate to use in studies of the synergistic role of each individual enzyme in starch degradation. Newer studies have tried to overcome some of the problems by using different assays and modern technologies. Limit dextrinase is present in

---

**Figure 1.8:** Reduction of protein disulphide bonds by the Trx system. The system involves: Trx, the cofactor nicotinamide adenine dinucleotide phosphate (NADP⁺), and the NADPH-dependent thioredoxin reductase (NTR).
very small amounts in malt and is laborious to purify. Suitable substrates for analysis of limit dextrinase activity became commercially available only recently. For these reasons little research on malt limit dextrinase has been carried out. Furthermore, many earlier results need to be re-evaluated now that the shortcomings of the techniques used previously have been recognised (Stenholm and Home, 1999).

Commercial malts contain low levels of LD, and the LD activity is moreover low compared with that of the other starch hydrolases (MacGregor et al., 1999). Furthermore, the level of LDI in malt is sufficient to inhibit most of the LD activity in a mash (Schroeder and MacGregor, 1998). As a result of the low LD activity, limit dextrans persist in worts and beer (Bringhurst et al., 2001). Limit dextrans cannot be fermented by yeast and represent a loss of alcohol yield to the brewer or distiller (Panchal and Stewart, 1979). On the other hand, it may be desirable to retain some of the branched dextrans to contribute to the mouth feel of the beer (Schroeder and MacGregor, 1998).

The importance of LD in the production of fermentable sugars has been the topic for much discussion during the last couple of decades. The activity levels of the other enzymes, i.e. α- and β-amylase involved in mobilisation of starch in germinating barley seeds are much higher than of LD. This emphasise the critical factor of the poor thermostability of LD. Recently, Evans et al. concluded that the levels of α-amylase, β-amylase and LD activities were all important for wort fermentability and that it was the balance of these enzymes and their relative thermostabilities that was important for the final fermentability rather than the individual levels of enzyme activity (Evans et al., 2005), furthermore the efficiency of any one starch degrading enzyme in a mash is influenced by the presence of other starch degrading enzymes (MacGregor et al., 1999).

But LD is a special case, because it was later shown that the level of LD thermostability was inversely correlated with total LD activity in malt (Evans et al., 2010). A study focusing on the activity level of LD during fermentation in malt and grain distilleries concluded that barley LD can survive the conditions encountered during mashing and is not only present in the vessel but its activity can increase during fermentation. One of the major differences between beer brewing and a distillery process like the Scotch whisky production is that the fermentation substrate (mash) is not boiled, LD will be therefore available to degrade dextrans into fermentable sugars, and can potentially increase the yield of alcohol (Walker et al., 2001).

The impact of LDI in the brewing process is not well understood. But as mentioned above the level of LDI in malt is sufficient to inhibit the LD present in the mash. LDI, therefore, has important implications for the malting and brewing industries (MacGregor et al., 2000). Due to this importance of LDI and LD, the interest in understanding the interactions between LD and LDI has increased during the last two decades.
1.5 Aims of the present study

The work presented in this thesis involves the characterisation of barley LD, its endogenous inhibitor, LDI, and their interaction investigated by X-ray crystallography, structure guided mutagenesis, and surface plasmon resonance analysis. The study benefits from the recent establishment of recombinant production of LD and LDI in *P. pastoris*, which made it possible to investigate the two proteins and their interaction in detail.

The scientific aims are:

i) Investigation of substrate specificity determinants of LD by X-ray crystallography and structure guided mutagenesis.

ii) Analysis of differences and similarities between LD and related α-glucan processing enzymes from different plants and microorganisms using bioinformatics tools.

iii) Determination of the structure of the complex between LD and LDI by X-ray crystallography.

iv) Investigation of the interaction between LD and LDI by structure guided mutagenesis and surface plasmon resonance analysis.
Chapter 2

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

The present chapter presents the complete crystal structure of LD (PDB entry 4AIO) solved with only a glycerol molecule bound in the active site. The glycerol molecule is coordinated by the nucleophile, Asp473, and originates from the reservoir solution used for growing the crystal. The structure was solved to 1.9 Å, and included three short loops of the N-terminal domain, which lacked in the two previously solved structures of LD in complex with α- or β-cyclodextrin (PDB entries 2Y4S and 2Y5E). The function of this N-terminal domain, which is typical of GH13 enzymes that cleave or form endo-α-1,6-linkages, is unknown. But, as presented in the paper, it shares homology to carbohydrate-binding module family 21. The uncomplexed LD structure will be used in the following Chapter 3 for comparison with complexed LD structures and structures of related GH13 enzymes.

*Supplementary material is included at the end of the chapter.*
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, Slotløts Plads, Building 224, 2800 Kgs. Lyngby, Denmark, and bProtein 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.

1. Introduction

Barley limit dextrinase (HvLD) 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 HvLD 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 HvLD 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 HvLD structure but were too flexible to be traced and included in the structures of the two HvLD–inhibitor complexes.
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 HvLD, 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 HvLD 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 mM MES buffer pH 6.6, 250 mM NaCl, 0.5 mM...
CaCl$_2$, 0.67 mM maltotriose, resulting in a sixfold molar excess of maltotriose, and crystals of HvLD 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 mM. Crystals appeared within one week. The HvLD 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, Molecular Dimensions, Newmarket, England) and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) with $\lambda = 0.976$ Å. The data were integrated using MOSFLM (Leslie, 1992) and scaled with SCALA from the CCP4 program suite (Winn et al., 2011). The resulting structure factors were used for molecular replacement (MR) using MOLREP (Vagin & Tepleyakov, 1997) from the CCP4 suite and the HvLD–β-CD model (PDB entry 2y4s; Vester-Cristensen, Abou Hachem, Svensson et al., 2010) including only the protein moiety. The model was refined using REFMAC5 (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 produced 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 HvLD 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 PDBRefold (Krissinel & Henrick, 2004) was performed, but no additional information was gained. The structure-based searches were also performed using the N-domain from the HvLD–β-CD structure (PDB entry 2y4s), but the number of significant hits was low compared with the searches with the N-domain from native HvLD 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 HvLD. 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 HvLD–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 HvLD structure (Fig. 1a) and the protein moiety of HvLD–β-CD are virtually identical, with an r.m.s.d. of 0.2 Å for all C$\alpha$ 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 HvLD structure (Fig. 1b), 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 HvLD (Fig. 1a). 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 HvLD–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 HvLD–β-CD (Fig. 1c).

#### 3.3. Active site

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

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 pullulanaise from Klebsiella pneumoniae, which belongs to the same subfamily as HvLD according to CAZY (Cantarel et al., 2009). They observed two different main-chain conformations of the loop (residues 706–710; EGWD9) 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 pullulanaise structure (PDB entry 2fgz) and in the structures with bound glucose (PDB entry 2fh6) or isomaltose (PDB entry 2h1h) the loop was in the ‘inactive’ free conformation,
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 HvLD (residues 510–514; EGWDF). In HvLD the loop is found in the ‘active’ form both in the case of the native structure presented here and in the HvLD–CD complexes, in which the loop and Trp512 in particular participate in binding. Noticeably, Trp512 of native HvLD 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. 1c) 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., 2000; Hondoh et al., 2003; Woo et al., 2008), among which is a GH13 glycosyl-debranching enzyme from Sulfolobus solfataricus (Woo et al., 2008), in which the Trp adopts the same rotamer and is in the same position as HvLD when substrate is bound.

These findings suggest that HvLD 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 HvLD is less flexible, possibly explaining the lower hydrolytic activity of HvLD 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

Supposition of the N-terminal domain of HvLD with those of the deposited HvLD–α-CD and HvLD–β-CD complex structures (Fig. 1b) 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 HvLD–α-CD structure (PDB entry 2y5s; Vester-Christensen, Abou Hachem, Svensson et al., 2010) and the previously mentioned well defined loop density of the three flexible loops in the native HvLD structure.

Several alignment methods were explored to advance insight into the possible role of the HvLD 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). Only five of these proteins are α,β-acting pullulanases belonging to GH13_13 and GH13_14 and the sequence identity to HvLD is in general low (see Supplementary Fig. 1). Common to the hits is that they, like the HvLD 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-β-glucosamine (NAG) binding (Supplementary Table 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 HvLD identified only pullulanase N-terminal domains with a FATCAT P-value of below 1.0 × 10–5. Noticeably, the N-terminal starch-binding domain of the CBM21 glucoamylase from R. oryzae (PDB entry 2dm; Liu et al., 2007) and the N-domain of HvLD align with a P-value of 1.44 × 10–3 despite having a sequence identity of only 6% (Supplementary Fig. 2). Ser76, Tyr78, Ser86 and Lys94 of HvLD 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). The starch
binding residues identified in R. oryzae CBM21 (Tung et al., 2008) are not conserved or are replaced by residues with similar biophysical properties in HvLD (Supplementary Fig. 2). It therefore seems unlikely that these residues play similar roles in the two molecules unless major structural changes occur in HvLD in the presence of starch. In conclusion, the N-terminal domain of HvLD may participate in intermolecular interactions that are important for the in vivo functionality of HvLD, 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 Ma¨ller-Dieckmann (ESRF beamline ID23-1) for assistance during data collection. MSM was supported by a DTU PhD stipend.

References


SUPPLEMENTARY MATERIAL

Native structure of the starch debranching enzyme barley limit dextrinase

Marie Sofie Møller\textsuperscript{ab}, Maher Abou Hachem\textsuperscript{a}, Birte Svensson\textsuperscript{a*} and Anette Henriksen\textsuperscript{b*}

\textsuperscript{a}Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Søltofts Plads, Building 224, Kgs. Lyngby, DK-2800, Denmark, and \textsuperscript{b}Protein Chemistry Group, Carlsberg Laboratory, Gamle Carlsberg Vej 10, Valby, DK-2500, Denmark

Correspondence email: bis@bio.dtu.dk; athx@novonordisk.com
Supplementary Fig. 1 Top: Structural alignment of domains with Z-score ≥ 5 in a DALI search (Holm & Rosenstrom, 2010) with the HvLD N-terminal domain (4aioA). The figure is generated using ALINE (Bond & Schuetzkoepf, 2009) based on output from DALI. Conserved residues are highlighted in blue colours. The upper panel in the alignment show the secondary structure of the PDB entries in the alignment. Bottom: Pairwise sequence identity scores from a structural alignment between the protein domains identified in the DALI search.
Supplementary Fig. 2  Top: FATCAT alignment of the LD N-terminal domain and the N-terminal starch binding domain (CBM21) of glucoamylase from *Rhizopus oryzae* (2djm). The residues in red are identical residues between LD and the CBM21. The residues in bold are residues involved in ligand binding of the CMB21 (Tung *et al.*, 2008). Bottom: A superposition of the *HvLD* N-terminal domain (orange) and the N-terminal starch binding domain (grey) with conserved residues from the alignment shown as sticks.
**Supplementary Table 1** Unique domains structurally similar to the LD N-terminal domain. The domains are identified by a DALI structural search (http://ekhidna.biocenter.helsinki.fi/dali_server/)

<table>
<thead>
<tr>
<th>PDB entry</th>
<th>Organism</th>
<th>Protein activity</th>
<th>DALI Z-score</th>
<th>Rmsd</th>
<th>Alignment length</th>
<th>Sequence identity (%)</th>
<th>Reference</th>
<th>Possible function of aligned sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Y5E</td>
<td><em>Hordeum vulgare</em></td>
<td>Limit dextrinase</td>
<td>19.9</td>
<td>1.0</td>
<td>104</td>
<td>100</td>
<td>(Vester-Christensen et al., 2010)</td>
<td>Whole domain - unknown function</td>
</tr>
<tr>
<td>2FH6</td>
<td><em>Klebsiella aerogenes</em></td>
<td>Pullulanase</td>
<td>16.7</td>
<td>1.8</td>
<td>112</td>
<td>31</td>
<td>(Mikami et al., 2006)</td>
<td>Domain core - unknown function</td>
</tr>
<tr>
<td>2E8Y</td>
<td><em>Bacillus subtilis</em> str. 168</td>
<td>Pullulanase</td>
<td>7.3</td>
<td>2.8</td>
<td>82</td>
<td>13</td>
<td>Malle et al., unpublished</td>
<td>Domain core - unknown function</td>
</tr>
<tr>
<td>2WAN</td>
<td><em>Bacillus acidopullulyticus</em></td>
<td>Pullulanase</td>
<td>6.6</td>
<td>2.4</td>
<td>78</td>
<td>14</td>
<td>(Turkenburg et al., 2009)</td>
<td>Domain core - unknown function</td>
</tr>
<tr>
<td>3DOI</td>
<td><em>Thermotoga maritima</em></td>
<td>Esterase with preference for (C2–C10) esters</td>
<td>5.1</td>
<td>3.2</td>
<td>82</td>
<td>9</td>
<td>(Levisson et al., 2009)</td>
<td>Domain core – Multimerisation of hexamer</td>
</tr>
<tr>
<td>1N10</td>
<td><em>Phleum pratense</em></td>
<td>Unknown, Grass pollen allergen Phi P1</td>
<td>5.1</td>
<td>2.6</td>
<td>73</td>
<td>10</td>
<td>Fedorov et al., unpublished</td>
<td>Domain core - unknown function</td>
</tr>
<tr>
<td>1EBA</td>
<td><em>Homo sapiens</em></td>
<td>Peptide-EPO receptor</td>
<td>5.1</td>
<td>3.3</td>
<td>77</td>
<td>13</td>
<td>(Livnah et al., 1998)</td>
<td>Peptide ligand binding</td>
</tr>
<tr>
<td>3FAW</td>
<td><em>Group B Streptococcus agalactiae</em></td>
<td>Pullulanase</td>
<td>5.0</td>
<td>2.7</td>
<td>72</td>
<td>7</td>
<td>(Gourlay et al., 2009)</td>
<td>Domain core - unknown function</td>
</tr>
<tr>
<td>2BSI</td>
<td><em>Homo sapiens</em></td>
<td>Cytokine receptor gamma chain</td>
<td>5.0</td>
<td>3.0</td>
<td>76</td>
<td>7</td>
<td>(Wang et al., 2005)</td>
<td>Domain core – NAG binding</td>
</tr>
</tbody>
</table>

Entries with more than 90% identical amino acid sequence are only included with one representative entry.

Only entries with a Dali Z-score ≥ 5.0 are included.

**References**


Chapter 3

Substrate specificity determinants of limit dextrinase

The structures of six different pullulanases from GH13, 12–14, including barley LD, have been solved to date, as described in Section 1.1.3. These structures are solved in complex with different ligands, among these are the hydrolysis products, i.e., maltooligosaccharides. But none of the pullulanases have been structure determined in complex with a natural substrate, i.e., a limit dextrin (branched maltooligosaccharide). In the present work catalytic LD variants has been constructed to obtain a crystal structure of LD in complex with a natural substrate. This was done to gain new insight into the substrate binding mechanism of LD, and to identify substrate specificity determinants, as kinetic analysis has shown that in order for LD to be highly active on a limit dextrin, there has to be a glucose-unit at both sides of the glucose-unit, to which the branch is connected (Jensen, 2004), i.e., the subsite 0′ (see Section 1.1.1 has to be covered. The resulting crystal structures and those of other pullulanases are compared in order to get insight into structural determinants that could explain differences in substrate preference among pullulanases (Section 1.1.2). In addition to the crystal structure determinations a possible specificity determinant of barley LD, Met440, was investigated by mutagenesis and enzymatic assays.
3.1 Materials and methods

3.1.1 Bioinformatic analysis

The catalytic domains of protein sequences used for multiple alignment were extracted using the NCBI Batch Web CD-search tool (Marchler-Bauer and Bryant, 2004) in combination with a script (unpublished work by Alexander Holm Viborg). Multiple sequence alignments were done using MUSCLE from the MEGA version 5 (Tamura et al., 2011), except for the multiple sequence alignment based on protein structures from GH13 subfamilies 8–9 and 11–14, which was generated with the PROMALS3D (PROfile Multiple Alignment with predicted Local Structures and 3D constraints) server using default settings (Pei et al., 2008b,a). The multiple sequence alignments generated using MEGA version 5 were visualised using ESPript (Gouet et al., 1999).

3.1.2 Mutagenesis - LD-M440G variant

The LD variants (M440G, D473A, and D473A-E510S) were obtained using the pPIC9K/LD construct as template (Vester-Christensen et al., 2010a) and the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) with primers LD-M440G-Fw+Rv, LD-D473A-Fw+Rv, and LD-D473A-E510S-Fw+Rv (Table 3.1). Transformation into \textit{P. pastoris} strain GS115, and selection for expression and secretion of LD was done as previously described (Vester-Christensen et al., 2010a).

Table 3.1: Mutational primers for introduction of mutations in LD. Fw, forward primer; Rv, reverse primer.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD-M440G-Fw</td>
<td>GCCAGAGTTGAGAACAGTGCAGCTGGTAACAGCAAGTGAGCA</td>
</tr>
<tr>
<td>LD-M440G-Rv</td>
<td>GCTCACTTGCTGTATTGTACCAGCTGCACGGTCTTCATCTGGCC</td>
</tr>
<tr>
<td>LD-D473A-Fw</td>
<td>TTGACGGGTTCAGATTTGCTCTTATGGGCCATATCATGAAACG</td>
</tr>
<tr>
<td>LD-D473A-Rv</td>
<td>CGTTTCATGATATGGGCCAAATCTGTAACCCGTCAACG</td>
</tr>
<tr>
<td>LD-D473A-E510S-Fw</td>
<td>ATATACGTGTATGTTAGGATGGACTTCGTCGAAAGTTGACGC</td>
</tr>
<tr>
<td>LD-D473A-E510S-Rv</td>
<td>GCGTGCAACCTCAGCGAAAGCTCCATCTGAAACCTCATAGTATAT</td>
</tr>
</tbody>
</table>

3.1.3 Production and purification of LD, wild type and variants

Recombinant LD, wild type and variants, were produced using \textit{P. pastoris} as a host and purified as described (Vester-Christensen et al., 2010a). In short, the recombinant proteins were obtained by secretory expression during high cell-density fermentation in 5 L scale, and the proteins were purified by a two-step procedure involving affinity chromatography using \(\beta\)-cyclodextrin (\(\beta\)-CD) conjugated to Sepharose (GE Healthcare, Sweden) followed by gel filtration (Hiload Superdex 200 26/60; GE Healthcare). Protein purity was assessed by SDS-PAGE using NuPAGE Novex Bis-tris 4–12% gels (Invitrogen, Carlsbad, CA). Protein concentrations were determined spectrophotometrically at 280 nm using a molar extinction coefficient of \(1.52 \times 10^5 \, \text{M}^{-1} \, \text{cm}^{-1}\) determined by aid of amino acid analysis (Barkholt and Jensen, 1989). The variant LD-D473A-E510S, which was used for crystallography, had an additional purification step to remove residual \(\beta\)-CD. Thus pooled and concentrated protein from the gel filtration step was mixed with maltotetraose to a final concentration of 250 mM maltotetraose. The mix
was then dialysed (Spectra/Por dialysis membrane cut-off 25 kDa; Spectrum Laboratories, Rancho Dominguez, CA) 6 × 24 h at 4°C against 6 × 1.5 l gel filtration buffer (50 mM MES/NaOH pH 6.6, 250 mM NaCl, 0.5 mM CaCl₂). After the dialysis the protein was run over a Hiload Superdex 200 16/60 column (GE Healthcare) equilibrated in the gel filtration buffer.

3.1.4 Assays

Enzyme kinetics - wild type LD and LD-M440G

The kinetic constants of wt LD and LD-M440G on different substrates were determined from initial velocities at 37°C using a modified reducing sugar assay (Vester-Christensen et al., 2010a). The starting volume of the assay was 1.1 ml containing substrate (0.02–1 mg/ml pullulan or 0.5–10 mg/ml amylopectin) and wild type LD (3.6 nM with pullulan and 25.7 nM with amylopectin) or LD-M440G (3.6 nM with pullulan and 102.8 nM with amylopectin) in assay buffer; 20 mM sodium acetate pH 5.5, 5 mM CaCl₂, 0.005% TritonX-100. Aliquots (100 or 200 µl dependent on substrate) were removed at 5 time points (3, 6, 9, 12, and 15 min) and added to 500 µl of stop solution (0.4 M sodium carbonate pH 10.7, 2.5 mM CuSO₄, 2.5 mM 4,4'-dicarboxy-1,2'-biquinoline, 6 mM L-serine) and Milli-Q water to a final volume of 1 ml. The absorbance was measured after 30 min at 80°C at $A_{540}$. The release of reducing sugar was quantified using a maltose standard curve (0–55.5 µM). The kinetic constants, the Michaelis constant $K_m$ and the catalytic constant/turnover number $k_{cat}$, were determined by fitting either the Michaelis-Menten equation (3.1) or the equation for uncompetitive substrate inhibition (3.2) to the initial velocities, where $K_{i_s}$ is the dissociation constant for the inhibitory [substrate-enzyme]-substrate ternary complex. The fitting and plotting were done using the Enzyme Kinetics Module 1.0 of the program Sigmaplot 9.01 (Systat Software, Chicago, IL).

$$V = \frac{V_{max}}{1 + \frac{K_m}{[S]}}$$  \hspace{1cm} (3.1)

$$V_{i,sub} = \frac{V_{max}}{1 + \frac{K_m}{[S]} + \frac{[S]}{K_{i_s}}}$$  \hspace{1cm} (3.2)

Specific activity of LD-D473A, and LD-D473A-E510S

The specific activity of the LD-D473A and LD-D473A-E510S variants were determined using the reducing sugar assay as described above with pullulan (0.3 mg/ml) as substrate and 100 µl samples were removed at 5 time points during 250 min. The protein concentrations of LD-D473A and LD-D473A-E510S were 250 nM and 500 nM, respectively. Wild type LD was included in the experiment with an assay concentration of 4.2 nM and 100 µl aliquots were removed at 5 time points (3, 6, 9, 12, and 15 min). In addition to the standard reducing sugar assay with 20 mM sodium acetate pH 5.5 as assay buffer, a parallel assay was run with 20 mM MES pH 5.5 as the assay buffer.
3.1.5 Structure determination and refinement

Crystallisation

The final concentration of the LD-D473A and LD-D473A-E510A variants for crystallisation was 10 mg/ml in 50 mM MES/NaOH pH 6.6, 250 mM NaCl, 0.5 mM CaCl₂, 10 mM 6²-α-maltotriosyl-maltotriose (gift from the late Bent S. Enevoldsen). Crystals were obtained by hanging-drop vapour diffusion at 293 K. Optimised crystals were obtained by streak-seeding using a reservoir solution consisting of 20% (w/v) polyethylene glycol (PEG) 3350, 0.3 M NaI, 5% glycerol. Crystals appeared within 5 days. The crystals were cryo-protected by changing the PEG 3350 concentration of the drops to approximately 35% by stepwise addition of 35% (w/v) PEG 3350, 0.3 M NaI, 5% glycerol, 150 mM 6²-α-maltotriosyl-maltotriose, and the crystals were flash-cooled in liquid nitrogen.

The LD-E510A variant used for crystallisation with a branched substrate was kindly produced and purified at the Carlsberg Research Center. The residual activity of the variant on 0.3 mg/ml pullulan was 0.0004% as compared with wild type LD. The working concentration of the protein for crystallisation was 10 mg/ml in 20 mM MES buffer pH 6.5, 150 mM NaCl, 10 mM DTT. Optimised crystals were obtained by streak-seeding using a reservoir solution consisting of 19–21% (w/v) PEG 3350, 0.3 M NaI. Crystals appeared within 5 days. The LD-E510A crystals were soaked with 6²-α-maltotriosyl-maltotriose by addition of 1 µl of a 100 mM substrate stock. Paratone-N (Hampton Research, CA, USA) was used as cryoprotection and crystals were flash-cooled in liquid nitrogen.

Data collection, processing, and refinement

X-ray diffraction data for all three LD variants were collected at beamline I911-2 at MAX-lab, Lund, Sweden with λ = 1.041 Å. The data were integrated using MOSFLM (Leslie, 1992) and scaled with SCALA from the CCP4 program suite (Winn et al., 2011). The resulting structure factors were used for molecular replacement (MR) using MOLREP (Vagin and Teplyakov, 1997) from the CCP4 suite and the LD-β-CD model (PDB entry 2Y4S; (Vester-Christensen et al., 2010b)) including only the protein moiety. The model was refined using REFMAC5 (Murshudov et al., 2011). Manual inspection, rebuilding, and addition of water molecules, ligand, and ions were performed with Coot (Emsley et al., 2010). The ligand was built and restraints were generated using a combination of the programs PRODRG (GlycoBioChem, UK) and Sketcher from the CCP4 program suite (Winn et al., 2011). In addition to the Coot validation functions, final analysis of model geometry optimisation was performed using the output from PROCHECK and MolProbity (Laskowski et al., 1993; Chen et al., 2010). All visualisation of the structures was done using the PyMOL Molecular Graphics System version 1.3 (Schrödinger, LLC).
3.2 Results

3.2.1 Bioinformatic analysis

Based on comparisons of the LD structure in complex with β-CD and the structure of K. pneumoniae pullulanase in complex with two maltotetraose molecules, it was suggested that Met440 of LD causes steric hindrance in the active site and thereby reduces the activity towards larger branched substrates like amylopectin (Vester-Christensen et al., 2010b). A multiple alignment (Figure 3.9) between the catalytic domain of all protein sequences of GH13 subfamilies 12, 13 and 14, i.e. the pullulanase subfamilies, shows that all the pullulanases from GH13_12 has a glycine at the position equal to Met440 of LD, furthermore a comparison of the LD structure (PDB entry 4AIO) and the two GH13_12 structures from S. agalactiae (PDB entry 3FAW) and S. pneumoniae (PDB entry 2YA0) confirms that the three-dimensional position of the glycines equals the position of Met440, even though the loop where they are situated has another course as compared with LD. The members of GH13_13, i.e. pullulanases/limit dextrinases from both bacteria and eukaryotes, have mainly a cysteine at the position of Met440 of LD. All the pullulanases from bacteria had a cysteine, while the only exceptions are observed among the eukaryotes (in total 12 proteins). The plants have either a methionine as LD or a valine, while the algae have a cysteine like the bacterial GH13_13 pullulanases. Finally, the GH13_14 pullulanases have either a valine or a cysteine at the position corresponding to that of Met440 of LD.

3.2.2 Substrate specificity analysis

A LD-M440G variant was constructed to determine if this methionine could be a specificity determinant. The substitution of methionine with glycine was chosen as all the GH13_12 pullulanases, which have higher activity towards larger branched substrates as compared with LD, had glycine at the equivalent position. The kinetics of LD-M440G on pullulan were essentially unchanged as compared with wild type (Table 3.2). The activity on amylopectin differed on the other hand. In general the catalytic efficiency of wild type LD on amylopectin is very low as compared with pullulan. Depending on which of the kinetic models that were fitted to the data, the catalytic efficiency on amylopectin was reduced 210- or 270-fold as compared with pullulan. As it can be recognised from the standard deviations for the kinetic values, the uncompetitive substrate inhibition model might not give a good description of the data. The plots in Figure 3.1 show the uncompetitive substrate inhibition model to fit nicely to the data with the best statistics ($R^2$=0.999), but the lack of data from higher substrate concentrations cause uncertainty of the kinetic values. Looking only at the results from the classical Michaels-Menten equation, it is seen that the $K_m$ of the LD-M440G is reduced approximately 1.5-fold as compared with wild type. In addition, the turnover number of LD-M440G is reduced 4-fold. As a result of these changes the catalytic efficiency is reduced 2.6-fold compared to that of wild-type.
Table 3.2: Kinetic constants of wild type LD and LD-M440G on pullulan and amylopectin.

<table>
<thead>
<tr>
<th></th>
<th>Wild type LD</th>
<th></th>
<th></th>
<th></th>
<th>LD_M440G</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_m$ (mg/ml)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>$k_{cat}/K_m$ (ml/(mg s))</td>
<td>$K_i$ (mg/ml)</td>
<td>$K_m$ (mg/ml)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>$k_{cat}/K_m$ (ml/(mg s))</td>
</tr>
<tr>
<td>Pullulan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncompetitive</td>
<td>0.16±0.02 78±10</td>
<td>488±16</td>
<td>1.5±0.4</td>
<td></td>
<td>0.15±0.06 72±15</td>
<td>480±216</td>
<td>1.7±1.1</td>
</tr>
<tr>
<td>substrate inhibition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potato amylopectin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classical Michaelis-</td>
<td>6.9±1.0 15.6±1.2</td>
<td>2.3±0.4</td>
<td></td>
<td>4.4±0.3</td>
<td>3.9±0.1</td>
<td>0.9±0.1</td>
<td></td>
</tr>
<tr>
<td>Menten</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncompetitive</td>
<td>25.6±33.2 46.8±54.8</td>
<td>1.8±3.1</td>
<td>6.0±10.4</td>
<td>9.9±2.4</td>
<td>7.2±1.4</td>
<td>0.7±0.2</td>
<td>13.6±5.7</td>
</tr>
<tr>
<td>substrate inhibition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.1: A) Michaelis-Menten plot of the kinetics of amylopectin hydrolysis by wild type (wt) LD (black dots) and LD-M440G (white dots). The solid line is the Michaelis-Menten fit to the initial rate data. B) The same data as shown in plot A. But the fit is the uncompetitive substrate inhibition model.
3.2.3 Crystal structures of LD complexed with ligands

In order to gain insight into the substrate binding of LD, three catalytic residue variants were produced and purified: a nucleophile mutant, LD-D473A, a general acid/base mutant, LD-E510A, and a double catalytic residues mutant, LD-D473A-E510S. The nucleophile mutant had 0.23% residual activity, while the double mutant had 0.01% residual activity. The residual activity of LD-E510A was 0.0004% as compared with wild type LD. The residual activity was measured both in a 20 mM sodium acetate buffer, which is the standard assay buffer, and in a 20 mM MES buffer, to elucidate if the acetate could play a role causing nucleophile rescue (Ly and Withers, 1999). But no difference in the level of residual activity was observed. Crystallisation was set up with all three variants. LD-D473A and LD-D473A-E510S were co-crystallised with Α-α-maltotriosyl-maltotriose (M3-M3) and additional ligand was soaked in after crystal formation. In case of LD-E510A crystals the ligand was soaked in afterwards. Datasets were collected for all three LD variants.

Figure 3.2: A) Overall structure of LD in complex with Α-α-maltotriosyl-maltotriose (LD:M3-M3). N-domain, orange; CBM48, green; catalytic domain, grey; C-domain, blue; Ca$^{2+}$, red; I$^-$, yellow. The ligand is shown as purple sticks. The unsolved loops of the N-domain are indicated by circles. B) Superimposition of the active site residues of LD:M3-M3 (green sticks) and the corresponding residues of uncomplexed LD (grey; PDB entry 4AIO), the general acid/base of LD (Glu510) is mutated to an alanine in the LD:M3-M3 structure. The M3-M3 is shown as purple sticks, while the glycerol from uncomplexed LD (GOL1888) is shown as grey sticks. The water molecules are shown as small spheres (LD:M3-M3, red; 4AIO, grey). The major difference between the active site residues of the two LD structures is indicated by a punctured circle. The nomenclature for the subsites follows that of Davies et al., 1997. C) Superimposition of the three catalytic residues of LD:M3-M3 and uncomplexed LD, colour code as in B.
Overall structure of LD-E510A in complex with a branched substrate

The structure of LD-E510A in complex with \(6^2\)-\(\alpha\)-maltotriosyl-maltotriose (M3-M3) comprises four structural domains (Figure 3.2A): the N-domain (residues 3–124); a CBM48 (residues 125–230), the catalytic domain (residues 231–774); and the C-domain (residues 775–885). In addition, M3-M3, two calcium ions, four iodide ions, and 378 water molecules were modelled. Refinement statistics are listed in Table 3.3. The structure of LD-E510A:M3-M3 and the uncomplexed LD structure (PDB entry 4AIO) are virtually identical, with an r.m.s.d. of 0.2 for all matched atoms (4910 atoms). The major difference between the two structures is the missing structure of three loops (residues 23–27, 40–48, and 103–108) in the N-terminal domain of LD-E510A:M3-M3. These loops were also lacking in the two first structures of LD; LD in complex with \(\alpha\)-CD (PDB entry 2Y5E) or \(\beta\)-CD (PDB entry 2Y4S) (Vester-Christensen et al., 2010b). The amino acid residues of LD-E510A:M3-M3 involved in the substrate binding are found in a similar arrangement and adopt the same rotamers as in uncomplexed LD (Figure 3.2B). There are two exceptions: the third catalytic site residue Asp642 changes rotamer (Figure 3.2C), and Phe553 at subsite +2 changes orientation. At the positions where the oxygens of Glu510 would be expected to be, two water molecules were observed (Figure 3.2C).

Overall structure of LD-D473A and LD-D473A-E510S

The overall structure of LD-D473A and LD-D473A-E510S resemble the structure of LD-E510A. The refinement statistics are listed in Table 3.3. Both structures lack three loops of the N-terminal domain like in the LD-E510A structure. The structures of LD-D473A and LD-D473A-E510S had products, i.e. linear maltooligosaccharides, bound in the active site, due to the relative high residual activity. The LD-D473A structure had clear electron density for a maltotriose molecule (subsites –1 to –3) and a glucose (subsite +2) (Figure 3.3B). Furthermore, the LD-D473A structure had partial electron-density for \(\beta\)-CD bound in a similar position as in the LD-\(\beta\)-CD structure (PDB entry 2Y4S) (Figure 3.3B), which originates from the first step of purification, i.e. \(\beta\)-CD affinity chromatography. Based on this an additional purification step was included for the LD-D473A-E510S variant prior to crystallisation. The structure of LD-D473A-E510S had clear electron density for two maltotriose molecules (2\(\times\)M3) in the active site (Figure 3.3C), and no electron density indicating the presence of \(\beta\)-CD. As it can be seen from Figure 3.3, the electron density clearly show that the mutations have been introduced in all three LD variants. The only obvious difference between the active site residues of the three LD variants is observed in the backbone position of the carbonyl group of the Ala473 of the LD-D473A-E510S structure (Figure 3.3D).

A comparison of the Met440 of the three LD structures presented in this thesis together with the three published LD structures (PDB entries; 4AIO (no ligand), 2Y4S (\(\beta\)-CD), and 2Y5E (\(\alpha\)-CD)) shows that the methionine is flexible, as the residue can display several different rotamers (Figure 3.4).
Figure 3.3: Overview of the ligands bound in the active site of the three LD variant structures. The electron density is shown at 1.0 $\sigma$, while the ligands and the three catalytic site residues are shown as sticks. The nomenclature for the subsites follows that of Davies et al., 1997. A) LD-E510A with 6$^2$-$\alpha$-maltotriosyl-maltotriose (purple) bound. B) LD-D473A with maltotriose and a glucose molecule (green) bound. The $\beta$-CD (light blue) from the LD:$\beta$-CD structure (PDB entry 2Y4S) is superimposed into the active site too. C) LD-D473A-E510S with two maltotriose molecules (green). D) Superimposition of the catalytic residues of the three LD variants; LD-E510A (green), LD-D473A (orange), and LD-D473A-E510S (light blue). The ligand of the LD-E510A structure is shown as purple sticks. The difference in the back-bone orientation is indicated by a punctured circle.

Figure 3.4: Superimposition of the six different crystal structures; free LD (green, 4AIO), LD:$\alpha$-CD (red, 2Y5E), LD:$\beta$ (blue, 2Y4S), LD:M3-M3 (purple), LD:M3+M3 (orange), and LD:M3+G (cyan). Met440 near subsite –3 is shown as sticks as well as the ligand from the LD:M3-M3 structure.
Table 3.3: Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>LD-E510A</th>
<th>LD-D473A</th>
<th>LD-D473A-E510S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchrotron facility</td>
<td>MaxLab, I911-2</td>
<td>MaxLab, I911-2</td>
<td>MaxLab, I911-2</td>
</tr>
<tr>
<td>and beamline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.041 Å</td>
<td>1.041 Å</td>
<td>1.041 Å</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>22.8-1.67 (1.71-1.67)</td>
<td>24.5-2.00 (2.11-2.00)</td>
<td>28.5-2.40 (2.46-2.40)</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>168.9</td>
<td>84.6</td>
<td>84.5</td>
</tr>
<tr>
<td>b</td>
<td>80.9</td>
<td>93.7</td>
<td>93.9</td>
</tr>
<tr>
<td>c</td>
<td>58.3</td>
<td>115.1</td>
<td>115.1</td>
</tr>
<tr>
<td>β</td>
<td>100.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. observed reflections</td>
<td>420337 (42726)</td>
<td>255697 (36879)</td>
<td>129851 (18834)</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>89214 (10649)</td>
<td>61923 (8859)</td>
<td>35987 (5145)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>15.9</td>
<td>14.2</td>
<td>16.6</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.2 (89.6)</td>
<td>99.0 (98.2)</td>
<td>98.6 (98.2)</td>
</tr>
<tr>
<td>Mean (I)/σ(mean(I))</td>
<td>12.0 (3.7)</td>
<td>7.8 (3.2)</td>
<td>8.0 (3.3)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.7 (4.0)</td>
<td>4.1 (4.2)</td>
<td>3.6 (3.7)</td>
</tr>
<tr>
<td>Rmerge a</td>
<td>0.079 (0.352)</td>
<td>0.158 (0.545)</td>
<td>0.124 (0.365)</td>
</tr>
<tr>
<td>R_pim b</td>
<td>0.039 (0.196)</td>
<td>0.086 (0.287)</td>
<td>0.076 (0.224)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used reflections</td>
<td>84739</td>
<td>58737</td>
<td>34150</td>
</tr>
<tr>
<td>R_cryst/R_free (%)</td>
<td>16.4/19.6</td>
<td>20.6/25.3</td>
<td>19.3/25.3</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>6705</td>
<td>6751</td>
<td>6718</td>
</tr>
<tr>
<td>Calcium ions</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Iodide ions</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Water molecules</td>
<td>378</td>
<td>227</td>
<td>106</td>
</tr>
<tr>
<td>Ligand</td>
<td>6α-maltotriosyl-</td>
<td>maltotriose</td>
<td>2 maltotriose</td>
</tr>
<tr>
<td></td>
<td>maltotriose and</td>
<td>glucose</td>
<td>molecules</td>
</tr>
<tr>
<td>Rmsd values from ideality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.020</td>
<td>0.019</td>
<td>0.015</td>
</tr>
<tr>
<td>Torsion angle (°)</td>
<td>1.994</td>
<td>2.077</td>
<td>1.785</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>0.23</td>
<td>0.58</td>
<td>0.23</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>99.77</td>
<td>99.42</td>
<td>99.77</td>
</tr>
<tr>
<td>MolProbity score c</td>
<td>1.63</td>
<td>2.00</td>
<td>2.05</td>
</tr>
</tbody>
</table>

aRmerge = |∑ₙₖₜₜ (Iₙₖₜₜ) − <Iₙₖₜₜ>)/ ∑ₙₖₜₜ Iₙₖₜₜ, where Iₙₖₜₜ is the intensity of the iⁿ observation of reflection hkl and <Iₙₖₜₜ> is the average over all observations of reflection hkl.

bR_pim is the multiplicity weighted Rmerge (Weiss, 2001)

MolProbity score is a log-weighted combination of a clash-score, percentage Ramachandran not favoured and percentage bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected (Chen et al., 2010)
3.2.4 LD contra \textit{K. pneumoniae} pullulanase

The majority of the residues of the other structure determined GH13,13 enzyme, \textit{KpPUL}, involved in ligand binding have the same position and rotamer as the corresponding residues of LD (Figure 3.5A). But there are a few differences in the active site residues between \textit{KpPUL} and LD: Cys644 of \textit{KpPUL}, which equals Met440, makes a disulphide bond to Cys643. LD has an alanine (Ala439) at that position. The two adjacent cysteines of \textit{KpPUL} are conserved between the pullulanases from bacteria and algae in GH13,13, while the plant enzymes lack the cysteine pair. Furthermore, Thr642 of \textit{KpPUL} makes contact with the glucose unit at 0′, while LD has an alanine (Ala438) at this position. The interaction is to the O-6 of the glucose unit, this atom would be involved in the 1,6-α-bond, which is lacking in the \textit{KpPUL} structure (PDB entry 2FHF). Finally, Lys727 of LD interacts with the glucose unit at subsite −3. \textit{KpPUL} on the other hand has an Asp919 at that position, but it do not interact with the ligand.

The two maltotetraose molecules from the structure of \textit{KpPUL} (PDB entry 2FHF) align quite well with M3-M3 from the LD structure (Figure 3.5B). In \textit{KpPUL} there is no direct contact between the subsite −4 glucose unit and the protein, there is only water interactions. Furthermore the glucose unit in subsite −1′ (subsite not covered in the LD:M3-M3 structure) of the main chain is only in contact with the protein via Ser640 (Mikami et al., 2006). The corresponding residue to Ser640 in LD, Asn436, is orienting away from the ligand, and thus it is not expected to interact with the substrate. The active site architecture of LD and \textit{KpPUL} resembles each other. But there are differences in the topology as shown in Figure 3.5C and D. LD has a more narrow active site cleft as compared with \textit{KpPUL}.

\textbf{Figure 3.5:} A) Superimposition of the residues of LD (green) and \textit{K. pneumoniae} pullulanase (grey; PDB entry 2FHF). The three catalytic residues are indicated by bold letters and the residue labels of \textit{K. pneumoniae} pullulanase are in italic. The general acid/base of LD is in the figure an alanine. B) Superimposition of the $6^2$-α-maltotriosyl-maltotriose ligand (purple) from the LD structure and the two maltotetraose molecules (grey) from the \textit{K. pneumoniae} pullulanase (PDB entry 2FHF). The nomenclature of the subsites follows Davies et al., 1997. C) Overview of the active site cleft of LD with differences in topology as compared with \textit{KpPUL} (D) indicated by punctured circles. The ligands in C and D are the same ligands as in B., and colour code is the same.
3.2.5 Active site architecture: LD contra GH13\_12+14 pullulanases and debranching enzymes from GH13\_11

LD contra streptococci pullulanases

The structures of the two streptococci pullulanases from GH13\_12 in general have a different active site cleft architecture as compared with LD and \( \text{KpPUL} \) (Figure 3.6). The structure of the pullulanase from \( S. \ pneumoniae \) is hard to compare with LD, as the CBM41 of \( \text{SpPUL} \) makes a lid. The CBM41-1 folds back to cover the active site, which creates the subsites –3 and –4 (Figure 3.6B). \( \text{SpPUL} \) has at least eight subsites, through which it recognises and hydrolyses glycogen. Three tryptophan residues of the CBM41-1 appear to be an integral part of substrate-binding at subsites –3 and –4. As it can be observed from Figure 3.6C the two maltotetraose molecules of \( \text{SpPUL} \) (PDB entry 2YA1) are twisted differently as compared with M3-M3 from the LD:M3-M3 structure. The minus subsites nearly sequester the branch from the glycogen branch point that is being removed during hydrolysis. \( \text{SpPUL} \) is unable to hydrolyse branch points of granular glycogen, as the accessibility to the deep active site is limited (van Bueren et al., 2011). One of the specific differences between the active site cleft of LD and \( \text{SpPUL} \), which is not due to the lid formed by the CBM41, is the steric hindrance formed by Phe553 of LD near the + subsites. \( \text{SpPUL} \) has a proline (Pro850) at that position. The \( S. \ agalactiae \) pullulanase has a proline (Pro874) at the position of Phe533 of LD as well. The two structures of \( \text{SaPUL} \) (PDB entries 3FAW and 3FAX) are truncated and lack the CBM41, so it is hard to judge if it would form a lid too. But if the contribution of the lid to the active site architecture is ignored there is still differences in the topology of the active site cleft. Based on a comparison between the structures of both \( \text{SaPUL} \) and \( \text{SpPUL} \) in complex with maltotetraose and the LD structure, it is clear that the glucose units at subsites +2 and +3 are guiding the main chain towards Asp621 of LD, which is a part of the loop insert (Asp600–Leu622) constituting a barrier at the + subsite end of the active site of LD (Figure 3.6, green circle). This loop insert is only found in LD and \( \text{KpPUL} \), i.e. GH13\_13 pullulanases (see multiple alignment, Appendix IV). Furthermore, the Phe514→Arg522 stretch of LD narrows the active site cleft of LD as compared with \( \text{SaPUL} \).

LD contra bacilli pullulanases

The active site cleft of the two bacilli pullulanases from GH13\_14 is wide open at the subsite +2 end (Figure 3.7, green circle) like the active site cleft of \( \text{SaPUL} \). As compared with the structure of \( \text{BaPUL} \) (PDB entry 2WAN), LD has a long insert, Ile318→Asp351, which forms a shoulder outside the active site at the – subsite end (Figure 3.7, orange circle). The active site cleft of \( \text{BsPUL} \) is closed at the – subsite-end of the active site caused by Lys343, which corresponds to Ser407 of LD. In addition, the active site of \( \text{BsPUL} \) is narrowed at subsite –2 due to Val372, which is an alanine in LD (Ala439).
Figure 3.6: Comparison of LD with the GH13_12 pullulanases, SaPUL (A; PDB entry 3FAW) and SpPUL (B; PDB entry 2YA1). The M3-M3 ligand of the LD structure is shown as purple sticks, while the maltotetraose molecules of Sp is shown as light blue sticks. CBM41 of SpPUL forms a lid resulting in a closed active site at the – subsites. The steric hindrance of LD (Asp600–Leu622) at the + subsites is indicated by a green circle, while the more closed active site cleft of the streptococci pullulanases is encircled by a blue circle. C) Closer view of the superimposed ligands from (B).
Figure 3.7: Comparison of LD with the bacilli pullulanases from GH13,14, BsPUL (A; PDB entry 2E8Y) and BaPUL (B; PDB entry 2WAN). The M3-M3 ligand of the LD structure is shown as purple sticks. The steric hindrance of LD (Asp600–Leu622) at the + subsites is indicated by a green circle, while the LD insert Ile318–Asp351 is encircled by a orange circle.

LD contra GH13,11 enzymes

Three enzymes of the GH13 subfamily 11 are structure determined, *i.e.* an isoamylase (ISA) from *Pseudomonas amyloferosa* (PDB entry 1BF2), a glycogen debranching enzyme (GDE) from *E. coli* (PDB entry 2WSK), and a GDE from *Sulfolobus solfataricus* (PDB entries 2VNC, 2VUY, and 2VR5). These enzymes of GH13,11 have debranching activity like the pullulanases, but they have higher activity towards glycogen as compared with the pullulanases. The three GH13,11 enzyme are not very similar, as they have different substrate size preferences (Katsuma et al., 1998; Song et al., 2010; Woo et al., 2008), and oligomerisation is important for the activity of the GDE from *S. solfataricus*, which additionally has both α-1,6-glucosidase and α-1,4-transferase activity (Woo et al., 2008).

The active site cleft of *E. coli* GDE is closed at the position of subsite –3 of LD (Figure 3.8A). A similar topology is seen for GDE from *S. solfataricus* (Figure 3.8B), where the closed active site cleft results in a limitation in the length of the branch to three glucose units. The *E. coli* and *S. solfataricus* GDEs have an open active site at the + subsite end of the active site, and the cleft of *E. coli* GDE is much deeper as compared with LD (Figure 3.8C). In addition, the loop of LD where Phe553 is situated is not present in *E. coli* GDE. Contrary to the two structure determined GDEs, the active site cleft of *P. amyloferosa* ISA is a very long cleft (Figure 3.8C).
Figure 3.8: Comparison of LD with the GH13_11 debranching enzymes: *E. coli* GDE (A; PDB entry 2WSK), *S. solfataricus* GDE (B; PDB entry 2VNC), and *P. amyloderamosa* ISA (C; PDB entry 1BF2). The M3-M3 ligand of the LD structure is shown as purple sticks. The steric hindrance of LD (Asp600–Leu622) at the + subsites is indicated by a green circle, while the closed active site cleft at − subsites of the GH13_11 enzymes is encircled by a yellow circle.
3.2.6 Structure based comparison of GH13 α-1,6-acting enzymes

As mentioned in section 1.1.3, the members of the GH family 13 might only share the overall structure of the catalytic domain, a (β/α)_8-barrel (Kuriki and Imanaka, 1999), and a couple of three-dimensionally as well as sequentially conserved amino acids (Table 1.1), even between the enzymes acting on α-1,6-glucosidic linkages. To compare the structures of the GH13 subfamilies 8–9 and 11–14 a multiple structure based protein sequence alignment was conducted (Figure 3.9 and Appendix IV). As it is seen from Figure 3.9 (and the full alignments in Appendices III and IV), the differences between especially the branching enzymes and the debranching enzymes lead to in differences between the two type of alignments. Not only the topology of the active sites of the branching enzymes and the debranching enzymes in general, but also some of the key active site residues of the debranching enzymes, e.g. the tryptophan from region II, are not shared with the branching enzymes (Figure 3.9B). The branching and partly the GH13.11 debranching enzymes have additional inserted amino acid sequence stretches as compared with the pullulanases and vice versa (Appendix IV).

Besides the branching and debranching enzymes, which process substrates with both 1,4- and 1,6-α-glycosidic linkages, there is the 1,6-α-glucosidases from GH13.31, which hydrolyse the bonds in linear substrates composed of glucose units connected by 1,6-α-glycosidic linkages, i.e. isomaltooligosaccharides (including panose) and the polysaccharide dextran. The glucan 1,6-α-glucosidase from Lactobacillus acidophilus NCFM (LaGH13.31) is an example of this group of enzymes (Møller et al., 2012b, Appendix VII). The GH13.31 enzymes were not included in the multiple sequence alignments, as the catalytic domain architecture differs significantly from the debranching enzymes. The catalytic site residues of LD and LaGH13.31 superimpose very well (Figure 3.10). But the active site of LaGH13.31 is a pocket contrary to the open active site cleft of debranching enzymes (Figure 3.10B and C). The pocket clearly only accommodate linear substrates like the isomaltooligosaccharides (Figure 3.10C).
Figure 3.9: A) Excerpt of a multiple sequence alignment of catalytic domains from enzymes with structures from GH13 subfamilies 8–9 and enzymes from GH11–14, which are enzymatically characterised and/or structure determined. All the eukaryotes defined as characterised in CAZy are included. See Appendix III for the complete alignment and a table with organism information and PDB entries. The residue numbering refers to the active site residues of barley LD. Barley LD is underlined by a dashed green line, and three of the four conserved regions between GH13 enzymes (MacGregor et al., 2001) are labelled II–IV.

B) Excerpt of a structure based multiple protein sequence alignment of the structure determined enzymes from GH13 subfamilies 8–9 and 11–14 generated using PROMALS3D (Pei et al., 2008b). The sequences are coloured according to secondary structure predictions (red: \(\alpha\)-helix, blue: \(\beta\)-strand). The consensus predicted secondary structures are shown indicated by h (\(\alpha\)-helix) or e (\(\beta\)-strand), furthermore the consensus amino acids are shown by the following symbols: conserved residues, bold and uppercase letters; aliphatic residues, l; aromatic residues, @; hydrophobic residues, h; polar residues, p; tiny residues, t; small residues, s; and bulky residues, b.
Figure 3.10: Comparison of LD (green, PDB entry 4AIO) and a representative from GH13.31: the glucan 1,6-α-glucosidase from *L. acidophilus* NCFM (light blue, PDB entry 4AIE). A) Superimposition of the two structures with the ligand from the LD-E510A structure superimposed into the active site (purple sticks), and the catalytic site residues shown as sticks. Some of the sequence segments, which close the active site of the GH13.31 enzyme, are indicated by arrows. B) Topology of LD in the active site cleft area and with ligand shown as purple sticks. C) Structure of the GH13.31 enzyme with the ligand from LD superimposed into the active site pocket. Point of view equals the view in B.
3.3 Discussion

3.3.1 Structures of barley LD with and without ligand

The structure of LD in complex with a branched substrate, \(6^2\)-\(\alpha\)-maltotriosyl-maltotriose (M3-M3), gives not only an insight into the action of barley LD on a natural substrate, but as it is the first structure of an \(\alpha\)-1,6 acting enzyme in complex with a natural substrate it can give additionally insight into the entire group of \(\alpha\)-1,6-debranching GH13 enzymes.

Third catalytic site residue Asp of GH13 members

The active site of the crystal structure of LD in complex with M3-M3 and the active site of un-complexed LD structure (PDB entry 4AIO, (Møller et al., 2012a)) turned out to be essentially identical, but one significant difference was observed: the third catalytic site residue Asp642 changed rotamer upon binding of the substrate. From the LD-(M3-M3) structure it is seen that Asp642 is clearly involved in binding the substrate in subsite +1 (Figure 3.2). The importance of the interaction between the third catalytic site residue of GH13 enzymes and OH-2 and OH-3 of the glucose unit at subsite –1 has been shown by crystal structures of a cyclodextrin glycosyltransferase (CGTase). These structures gave among other things insight into the intermediate of the catalytic reaction of GH13 enzymes. The third catalytic residue participated in substrate distortion (Uitdehaag et al., 1999), but it did not change orientation upon substrate binding, like Asp642 of LD does. A similar movement is seen for the other structure determined GH13_13 enzyme, KpPUL, upon binding of two maltotriose or two maltotetraose molecules oriented in a similar way as if it was a branched substrate (Figure 3.5). As compared with the structure of free KpPUL (PDB entry 2FGZ) the third catalytic residue (Asp834) is also reoriented in these two complex structures (PDB entries 2FHC and 2FHF), while no movement is induced upon binding of glucose (PDB entry 2FH6) or isomaltose (PDB entry 2FH8) at subsite –2 (the reducing end of isomaltose occupied an abnormal position according to Mikami et al., 2006). In addition, the structure of KpPUL in complex with two maltose molecules (PDB entry 2FHB) shows that the third catalytic residue, Asp834, is somewhat in between the free and the substrate bound orientation indicated by the presence of two rotamers of Asp834 in the structure file.

A thorough comparison of the orientation of Asp642 of LD with the corresponding amino acid residue in all other published structures in the GH13 family, i.e. 67 different proteins (only including proteins assigned to a subfamily in CAZy, 22 out of 35 subfamilies are represented), shows that this movement of the third catalytic site residue is restricted to GH13_12 and GH13_13. In all the other structures the aspartic acid is in the "substrate bound" position regardless of whether there is a ligand bound or not. The only exception is the debranching enzyme from *Nostoc punctiforme* assigned to GH13_20 (the neopullulanase subfamily), which shows a third orientation of the third catalytic site residue as compared with all other published GH13 structures. This enzyme is generally an outlier even within its subfamily, which contains cyclomaltodextrinases, maltogenic \(\alpha\)-amylases, and neopullulanases (Stam et al., 2006), because of i) its substrate preferences (Choi et al., 2009), ii) the lack of a N-terminal domain, and iii) the special dimerisation, which is essential for catalytic activity (Dumbrepatil et al., 2010). The debranching enzyme from *Nostoc punctiforme* exhibits activity towards both \(\alpha\)-1,4- and \(\alpha\)-1,6-glycosidic linkages, but prefers \(\alpha\)-1,6-bonds and the activity is highest on substrates occupying at least eight subsites of the active site. The substrate specificity is in the order pullulan > amylopectin > amylopectin > amylose, and it has no specificity towards cyclodextrins (Choi et al., 2009).
3.3.2 Substrate specificity determinants of LD in relation to other α-1,6-acting enzymes

The key to explain the differences in substrate specificity within the three pullulanase GH13 subfamilies 12–14, and between these three subfamilies and the rest of the α-1,6-acting GH13 enzymes, seems to be mainly due to the topology of the active site cleft rather than the specific residues involved in substrate binding, as they are relatively conserved contrary to the topology of the active site, which for instance is influenced by insertion of additional stretches of amino acid residues in to the classical (β/α)₈.

Met440 of LD

Met440 of LD was identified as a possible substrate specificity determinant based on structure superimpositions of maltotetraose molecules from the K. pneumonae pullulanase into the active site of the structure of LD in complex with β-CD, and a bioinformatic analysis including sequences from the bacterial pullulanases, which have higher activities on amylopectin as compared with plant pullulanases. KpPUL has a cysteine (Cys644) at the position equal to Met440 of LD, while the bacterial pullulanases from GH13 have a glycine at the position, and bacterial pullulanases from GH13 subfamilies 13 and 14 have a cysteine. An LD-M440G variant was produced and purified. The kinetics of the LD-M440G variant on pullulan was comparable with the wild type data, whereas with amylopectin as substrate a difference was observed, as the introduced mutation apparently causes a higher affinity and a lower turnover number (Table 3.2). The resulting catalytic efficiency was reduced 2.6 fold, when looking at the values from the classical Michaelis-Menten model. So the mutation, which was expected to result in more space for the large branched substrate has caused a decrease in the efficiency of LD. The question is if introducing a cysteine instead of a glycine could have resulted in a higher activity? It is not likely, as the kinetic values of wild type LD hydrolysis of amylopectin is comparable to the kinetic parameters of KpPUL on potato amylopectin; $K_m$ is 10.1 mg/ml and the turnover number is 14.1 s⁻¹, resulting in a catalytic efficiency ($k_{cat}/K_m$) of 1.4 ml/(mg s) (Yokobayashi et al., 1973). In addition to the results from the kinetic analysis, the additional LD structures solved and presented in this chapter show that Met440 of LD is flexible and can accommodate the binding of substrates with a branch of at least three glucose residues (Figure 3.4). The structure of LD in complex with a small branched substrate together with the recently published bacterial pullulanase structures and structures of other α-1,6-acting GH13 enzymes give now the possibility to identify more differences of not only the active site architecture, but also the accompanying domains, which may explain the differences in substrate preferences.

Phe553 of LD

The structure comparisons of LD with the GH13 subfamilies 11, 12, and 14 connected with the structure based multiple alignment suggest that Phe553 of LD could cause steric hindrance at subsite +2. Mikami et al. have suggested that the equivalent residue of KpPUL, Phe746, could be preventing the binding of a second branch chain of amylopectin, since it forms a sharp knob (Mikami et al., 2006). The two phenylalanines occupy the same position in both enzymes (Figure 3.5). The structure based multiple sequence alignment between pullulanases from the three different GH13 subfamilies (Figure 3.9B) shows that a phenylalanine is present at the same position as Phe553 or next to it. But the phenylalanine of the GH13 12 and 14 is oriented away from the active site, somewhat guided by the proline located next to it. The debranching
enzymes from GH13_11 and the branching enzymes (GH13_8 and 9) lack the loop, where the phenylalanine is located resulting in a more open active site.

**LD contra pullulanases in general**

When comparing the specific activities on polysaccharides of pullulanases and limit dextrinases from the three GH13 subfamilies 12–14 it is important to bear in mind that some of the CBMs preceding the catalytic domain are functional (see Section 1.1.3), i.e. they can assist in capturing, orienting and/or retaining the polymeric substrate. From the structure of *SpPUL* in complex with two maltotetraose molecules (PDB entry 2YA1) it is clear that the CBM41 plays a role in binding the substrate at the – subsites (Figure 3.6B). The lid formed by the CBM41 has an impact on the ability of *SpPUL* to bind substrate, but the catalytic efficiency is not affected by removal of the CBM (Gourlay et al., 2009).

In addition to Phe553, another clear difference between LD and the enzymes from GH13 subfamilies 11, 12 and 14 appeared from the structural comparisons and the structure based alignment (Appendix IV), i.e. the insert of 23 amino acid residues (Asp600–Leu622), which causes a narrowed and closed active site cleft after subsite +2. Among the amino acids of this loop seems Asp621 of LD to play a role in obstructing the effective binding of substrates with a main chain longer than four glucose units. This fits with kinetics results on different defined limit dextrins, which showed that the optimal substrate for LD is a 6\(^3\)-α-maltotriosyl-maltotetraose (Jensen, 2004). The optimal length of the branch of three glucose units also agree with the observations from the LD:M3-M3 structure, where there apparently is space for longer branches, but LD only accommodates branches of up to three glucose units.
Chapter 4

Manuscript: Structural rational for regulation of barley limit dextrinase activity by the endogenous proteinaceous limit dextrinase inhibitor

The manuscript, which is going to be submitted to Plant Cell, in the present chapter presents the results from part of the PhD project, which focused on the interaction between LD and LDI. Based on the crystal structure of the complex between LD and LDI, structure guided mutations of both LD and LDI were introduced on one hand to investigate the hot spots of the interaction, and on the other hand to investigate if it is possible to engineer LD to be less sensitive to LDI inhibition. The analyses of the interaction were conducted using surface plasmon resonance analysis. The cited literature as well as the figures and tables referred to in this chapter are present at the end of the manuscript.
Structural rationale for regulation of barley limit dextrinase activity by the endogenous proteinaceous limit dextrinase inhibitor

Marie S. Møller, a,b Malene B. Vester-Christensen, a,b,2 Johanne M. Jensen, a,3 Maher Abou Hachem, a Anette Henriksen, b,1,4 and Birte Svensson a,1

a Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Søltofts Plads Building 224, DK-2800 Kgs. Lyngby, Denmark
b Protein Chemistry Group, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-1799 København V, Denmark
1 Address correspondence to bis@bio.dtu.dk and athx@novonordisk.com

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Birte Svensson (bis@bio.dtu.dk)

2 Present address: Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3, DK-2200 København N, Denmark
3 Present address: Department of Drug Design and Pharmacology, Biostructural Research, University of Copenhagen, Universitetsparken 2, DK-2100 København Ø, Denmark
4 Present address: Protein Structure and Biophysics, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark.

Running title: Barley LD:LDI complex

Estimate of length:
ABSTRACT
Barley limit dextrinase (LD) catalyzes cleavage of α-1,6-glucosidic linkages in limit dextrins produced during starch degradation in germinating seeds. The activity of LD is regulated by its cereal-type protein inhibitor LDI. Proteinaceous inhibitors acting on polysaccharide processing enzymes and proteases are abundant in plant seeds and play crucial roles in the metabolic control and the defense against pests and pathogens. While inhibitors of α-amylases are well studied, molecular aspects of inhibition of debranching enzymes remain unexplored. Here the crystal structure of the LD:LDI complex is solved to 2.7 Å revealing a different mode of inhibition from other cereal-type inhibitors of α-amylases assigned to the same glycoside hydrolase family (GH13) as LD. The kinetics and the energetics of the LD:LDI complex formation are discerned by surface plasmon resonance analysis, while mutational analysis enabled identification of hot spots of the interaction. The potent inhibitory activity of LDI ($K_D\approx 40$ pM) and its exquisite thermostability measured for free LDI and the complex with LDI are discussed to highlight the functional versatility of the conserved structural scaffold of cereal-type inhibitors as mediator of protein-protein interactions in diverse plant regulatory networks.
INTRODUCTION

Starch, the major carbohydrate reserve in plant leaves, tubers and seeds, is composed of two polysaccharides, amylepectin which is an α-1,4-glucan containing α-1,6-linked branches, and amylose an essentially linear α-1,4-glucan (Buléon et al., 1998). In cereals, starch is synthesized and deposited in the seed endosperm during grain filling resulting in semi-crystalline, water-insoluble starch granules (Buléon et al., 1998). In leaves, active in photosynthesis, transient starch is accumulated during the day and mobilized at night to support respiration and growth when photosynthesis is arrested (Buléon, 1998; Geiger and Servaites, 1994). Starch has been identified as a major integrator in the regulation of plant growth supposedly through variations in the regulatory network that balances growth to the carbon supply (Sulpice et al, 2009). Besides its central function in plant physiology, starch is the most important carbohydrate for food and feed (Smith 2008). Breeding towards increased crop yields is imperative bearing the rising demand for food in mind. Overall starch synthesis is a complicated process conducted through the concerted action of a number of enzymes, including starch synthase, branching and debranching enzymes governing the composition and the supramolecular structure of the starch granule (Zeeman et al., 2010). The involvement of isoamylase-type debranching enzymes in starch biosynthesis is well established, whereas the significance of pullulanase-type (limit dextrinase, LD) α-1,6-debranching enzymes is still unclear (Tetlow, 2011). Substantial pullulanase activity has been detected in the developing rice and maize endosperm (Nakamura et al., 1996; Beatty et al., 1999), and it has been proposed that LD in rice (Oryza sativa) and maize (Zea mays) (and Arabidopsis) assists the isoamylase-isozymes in trimming the branched amylepectin precursors in starch biosynthesis (Dinges et al., 2003; Wattelbled et al., 2008; Fujita et al., 2009). Moreover LD was shown to be expressed in barley (Hordeum vulgare) during seed development, the physiological significance, however, remains unclear (Sissons et al., 1993; Burton et al., 1999). By contrast, it is well established that LD is the sole debranching activity relevant for mobilization of stored barley starch during germination, where it hydrolyses α-1,6-glucosidic linkages in branched maltooligosaccharides (limit dextrins) produced by the action of α- and β-amylases (Burton et al., 1999; Kristensen et al., 1999). In addition, LD shows high activity towards pullulan (Vester-Christensen et al., 2010a), a linear polysaccharide of α-1,6-linked maltotriose units, while its activity towards amylepectin is only <1% relative to the activity towards pullulan (Manners and Yellowles, 1971). Cereal-type inhibitors are wide spread in cereals, and implicated in various physiologically important processes, e.g. regulation of endogenous enzyme activities and defense against pathogens and pests, mainly fungi and insects. The cereal-type inhibitors have been found to exhibit distinctly different specificities against mammalian and insect α-amylases (Iulek et al., 2000; O'Donnell and McGeeney, 1976; García-Maroto et al., 1991; Maeda et al., 1982), and some act on dual targets, typically an α-amylase together with either trypsin or chymotrypsin, e.g. the bifunctional α-amylase/trypsin inhibitor from ragi (Eleusine coracana Gaertn.) (RBI).

The endogenous limit dextrinase inhibitor (LDI) from barley seeds acts specifically on LD and effectively suppresses LD activity during germination by complex formation with high affinity (Macri et al., 1993; MacGregor et al., 1994; MacGregor 2004; Jensen et al 2011). Analogously, the α-1,4 acting barley α-amylase isozyme 2 (AMY2), is regulated by a proteinaceous inhibitor, the α-amylase/subtilisin inhibitor
(BASI) (Mundy et al., 1983; Weselake et al., 1983; Nielsen et al., 2004). Structural and mechanistic aspects of the inhibition of AMY2 and several α-amylases by various types of proteinaceous inhibitors have been dissected in detail (Svensson et al., 2004). Noticeably, however the mode of inhibition of the related debranching enzymes was not explored. LDI belongs to the group of cereal-type inhibitors that share a common fold comprising four α-helices connected by irregular loops, and stabilized by four or five disulphide bonds (José-Estanyol et al., 2004). These inhibitors can be extracted from flour by chloroform-methanol, and hence are referred to as CM-proteins (Svensson et al., 2004).

The importance of LDI in seed germination and its regulatory context at the interface of starch synthesis and degradation motivates a better understanding of the molecular features that govern its function. Successful recombinant production of LD (Vester-Christensen et al., 2010a) and LDI (Jensen et al., 2011) makes it possible to generate new knowledge including structural insight about these two proteins and their interaction. Here the crystal structure of the LD:LDI complex is determined and reconciled by a comprehensive analysis of the kinetics and energetics of LD:LDI complex formation as well as hot spots of the protein-protein interaction, examined by mutational analysis. All together, the data bring novel insight into an unexplored facet of cereal-type inhibitors in plants, and highlights structural and energetic details of LD:LDI binding that are required to potently inhibit the open active sites of debranching enzymes.
RESULTS

Bioinformatics Analysis

To outline the evolutionary relationship among cereal-type inhibitors and possibly identify putative inhibitors of LDs and related enzymes, the sequences of barley LDI and three other structurally characterized cereal-type inhibitors, i.e. RBI, the 0.19 α-amylase inhibitor from wheat (0.19 AI) and the corn Hageman factor inhibitor (CHFI) were used in a protein BLAST search querying all sequences from monocots (Poaceae) in the NCBI-database (Figure 1A and see Supplemental Figure 1 online). Only three proteins clustered with LDI; pUP88 (CAA68248) from wheat (Triticum aestivum); a predicted α-amylase/trypsin inhibitor-like protein (XP003561291) from Brachypodium distachyon, which serves as a model organism for cereals; and a putative limit dextrinase inhibitor (ABK34477) from rice (Oryza sativa Indica group). None of these proteins were previously characterized. pUP88 from wheat, the closest non-barley relative of LDI, showed 84% sequence identity, while the protein from Brachypodium distachyon was 55% identical to LDI. The genome of Brachypodium distachyon encodes an uncharacterized protein having 87.6% sequence identity with LD, making it likely that an LDI-like inhibitor is present. Therefore, the regulation of LD activity in other cereals than just barley is likely to be mediated by LDI.

Several clusters in the phylogenetic tree contain one or more known proteins with the same three-dimensional scaffold as LDI, e.g. 0.19 AI-like homodimeric α-amylase inhibitors as well as groups of trypsin inhibitors and α-amylase inhibitors mainly from barley and wheat (Figure 1A). These trypsin and/or α-amylase inhibitors are reported to be heterotetrameric and composed of different cereal-type inhibitor subunits, i.e. CMa–CMe in barley (Grosset et al., 1997) and CM1–CM17 in wheat (Gomez et al., 1989). In addition to the barley and wheat proteins, sequences of small groups of uncharacterized protein from Sorghum bicolour are retrieved in this BLAST analysis. The only residues, conserved in all 45 cereal-type inhibitor sequences, are six cysteines (see Supplemental Figure 1 online). LDI and the putative LD inhibitors from wheat and Brachypodium distachyon contain nine cysteine residues, of which one is a free cysteine (Cys-59 in LDI as shown in the crystal structure, see below). These proteins thus lack one cysteine residue compared to the previously characterized proteinase/α-amylase inhibitors, which contain five disulphide bonds (Figure 1 and see Supplemental Figure 2 online). Noticeably LDI and pUP88 from wheat both have an N-terminal Leu-Glu extension preceding the N-terminal dipeptide sequence Ser-Val in RBI shown to be important for inhibition of yellow meal worm α-amylase (Strobl et al., 1998). It is remarkable that the two proteins from Brachypodium distachyon and rice lack the Ser-Val motif. In summary, the cereal-type inhibitors possess low sequence identity, but share the structural scaffold, which is stabilized by four or five conserved disulfide bonds.
Three-dimensional Structure of the LD:LDI Complex

Overall Structure

The crystal structure of the LD:LDI complex was solved at 2.7 Å resolution (Figure 2A) by molecular replacement using the coordinates of LD (Protein Data Bank [PDB] code: 2Y4S) and the α-helices of RBI (PDB code: 1B1U). Data collection and refinement statistics are provided in Table 1. The asymmetric unit contains two LD:LDI complexes; complex I comprises chains A (LD; residues 2–884) and C (LDI; residues 8–108) and complex II chains B (LD; residues 2–885) and D (LDI; residues 6–107). The two complexes were in the later parts of refinement treated separately and in the final model they could be superimposed with a root-mean-square deviation (RMSD) of 0.3 Å including all atoms. The LDI structures from complex I and II superimposed with an RMSD of 0.2 Å based on Ca and showed the largest differences in the third loop segment (Gly-82–Arg-89) and at the C-terminus (Leu-99–Ser-108), while the two LD structures superimposed with an RMSD of 0.3 Å. Complex I was used in the following analysis.

LDI contains four α-helices (α1–4) connected by long loop segments and is stabilized by four disulphide bonds; C9-C57; C23-C46; C32-C87; and C47-C105 (Figure 2B). These structural features are typical of the cereal-type inhibitors (Figure 1; see Supplemental Figure 1 online). The overall structure of LDI is similar to the three related proteins with known structure; RBI (uncomplexed, PDB code: 1B1U; in complex with α-amylase from yellow meal worm (larvae of Tenebrio molitor; TMA), PDB code: 1TMQ); bifunctional corn Hageman factor/amylase inhibitor (PDB code: 1BEA); and 0.19 α-amylase inhibitor from wheat (PDB code: 1HSS). The RMSD of the Ca atoms between LDI and the three other inhibitors is in the range of 0.78–1.18 Å (Table 3). Differences between the structures are essentially confined to the loop regions and stretches including the N- and C-termini (see Supplemental Figure 2 online).

The overall LD structure in the LD:LDI complex is very similar to that of ligand-free LD (PDB code: 4AIO) (Møller et al., 2012) superimposing with an RMSD of Ca atoms of 0.4 Å. The structure of free LD, however, has three flexible loops in the N-terminal domain which were not traceable in two structures of LD in complex with LDI or with the competitive inhibitors α- or β-cyclodextrin (α- or β-CD) (PDB entries: 2Y5E and 2Y4S, respectively) (Vester-Christensen et al., 2010b). LD in complex with LDI contains two Ca\textsuperscript{2+} ions at the same positions as the two Ca\textsuperscript{2+} ions of the native LD structure (PDB entry: 4AIO) (Møller et al., 2012). LD active site residues, including the catalytic Asp-473, Glu-510, and Asp-642, were essentially superimposable with the counterparts in free LD. Two active site residues, Phe-553 and Arg-697, however, adopted different rotamers in LD:LDI compared to free LD (see Supplemental Figure 3 online). It is likely that LDI residues Leu-41 and Val-42 impose steric constraints on the active site of LD causing a change in the χ2 rotation angle of Phe-553, and forcing Arg-697 into a different rotamer. A number of changes were observed in side chain orientation of LD residues interacting with LDI (distance <4Å) outside the active site region. Asn-551, Arg-582, Glu-726, and Asp-730 thus occurred as different rotamers than in free LD (see Supplemental Figure 3 online). In addition, the side chains of Gln-558 and Phe-620 were poorly defined in the free LD structure (PDB code: 4AIO), but appeared ordered in LD:LDI, possibly due to the complex formation.
The LD:LDI Complex Interface

The active site region of LD is an open, solvent accessible cleft and the total buried surface area in the LD:LDI complex is 2650 Å² (1325 Å² for each of LDI and LD). Residues from two (α1 and α2) of the four α-helices in LDI and parts of the loop regions 1 and 3 (Figure 2), as well as the N-terminal segment before the first α-helix, participate in the interaction (Table 2). Among the LDI residues forming hydrogen bonds and/or salt bridges, Arg-38 seems to play a role in inhibition (Figure 2D), as it interacts with both the catalytic nucleophile (Asp-473) and the general acid/base catalyst (Glu-510) of LD (Table 2). The contact with Asp-473 includes a water molecule. Furthermore, Arg-34 from LDI interacts with LD residues Glu-729 and Asp-730 (Figure 2C) situated at the entrance to the active site. In addition to the two arginines, LDI residues Leu-41 and Val-42 interact with Trp-512, Phe-514, and Phe-553 of LD forming a hydrophobic cluster shared between LD and LDI in the complex. Phe-514 belongs to the conserved loop in glycoside hydrolase family 13 (GH13) carrying the general acid-base (Glu-510) (MacGregor et al., 2001). The hydrophobic cluster residues Leu-41 and Val-42 moreover stabilize the conformation of Arg-38 by packing towards the aliphatic part of its side chain.

Comparison of LD:LDI with the Complex between the α-amylase from Yellow Meal Worm and the Bifunctional α-amylase/Trypsin Inhibitor from Ragi

Superimposition of the inhibitor molecules from LD:LDI and TMA:RBI (PDB code: 1TMQ) reveals distinctly different binding modes of LDI and RBI (Figure 3). In TMA:RBI, the N-terminal segment of RBI (Ser-1–Ala-11) has a key role in the inhibition by blocking the active site of TMA. The RBI N-terminal sequence S⁵VGTS⁶ preceding Cys-6, the first of the conserved cysteine residues, is unstructured in solution (Strobl et al., 1995; Gourinath et al., 2000), but assumes a 3₁₀-helical conformation upon binding to TMA (Strobl et al., 1998). This behavior appears critical for the inhibition mechanism, involving interaction of the N-terminal amino group of Ser-1 with the three acid residues at the catalytic site of TMA (Strobl et al., 1998). As opposed to this mode of contact neither the N-terminal residue nor the N-terminal segment (T¹LESVKDE₈) of LDI contribute to the contact between LDI and LD (Figure 2B). In fact, no significant electron density was observed for the first five N-terminal residues of LDI, indicating that the N-terminal region is not ordered in the LD:LDI complex. Mutations in LDI introducing either insertions or deletions at the N-terminus had only modestly affected the inhibitory activity and confirmed that the N-terminal segment in LDI has no role in the inhibition of LD (Table 4). Besides the N-terminal segment, the Pro-52–Cys-55 stretch of RBI is directly involved in TMA binding (Strobl et al., 1998). Two of the residues are conserved in LDI (Pro-54 and Cys-57) (Figure 1B), but the Pro-54–Cys-57 segment is not involved in the LD:LDI complex formation (Table 2). RBI Arg-61, Val-67–Ser-70, Thr-107–Gly-110 and Leu-115–Leu-117 also interact with TMA (Strobl et al., 1998), and these residues only Arg-61, Thr-68, and Pro-69 are conserved in LDI (Arg-63, Thr-71 and Pro-72).

The trypsin-binding loop of RBI, where Arg-34 and Leu-35 form the scissile peptide bond, is located at the opposite side of its α-amylase-binding site (Figure 3A), (Strobl et al., 1998). LDI has Ser-37 and Arg-38 at
the corresponding positions (Figure 1) consistent with failure of LDI to inhibit porcine pancreatic trypsin (MacGregor et al., 2000).

Biochemical Analysis of the LD:LDI Interaction

Kinetics and Energetics of Complex formation

Maximum affinity of LDI to LD was found to be at pH 6.5 as measured by surface plasmon resonance (SPR) with $K_\text{D} = 27.5 \pm 0.2$ pM (see Supplemental Figure 4 and Supplemental Table 3 online). The $K_\text{D}$ value increased up to 12–15-fold when pH approached 5.0 or 10.0 owing largely to changes in $k_{\text{off}}$. In addition, the $K_\text{D}$ improved by lower temperatures (10–20°C), but increased 6.5-fold at 45°C mainly due to the 16-fold faster $k_{\text{off}}$ (see Supplemental Table 4 online). A van’t Hoff non-linear thermodynamic analysis, which is a tool for understanding the mechanism of recognition between molecules, was performed using SPR data from the temperature range 10–35°C (see Supplemental Table 4 online). According to this analysis, the LD:LDI complex formation was associated with a large decrease in heat capacity ($\Delta C_p = -3.2$ kJ K$^{-1}$ mol$^{-1}$) and driven by a large favorable free energy change, $\Delta G^\circ = -57$ kJ mol$^{-1}$ originating from essentially equally favorable entropy ($\Delta S^\circ = -30$ kJ mol$^{-1}$ corresponding to ~53% of the total free energy) and enthalpy ($\Delta H^\circ = -27$ kJ mol$^{-1}$) changes.

LD Thermal Stability Considerably Increased by LDI Binding

The conformational stability of the free LD and LDI and of the LD:LDI complex was evaluated by using differential scanning calorimetry (DSC) at pH 6.5 (where affinity is highest as indicated by SPR analysis). The thermogram of LD showed a single slightly asymmetrical peak with an assigned unfolding temperature ($T_m$) of 65.9°C (Figure 4), while the thermogram of LDI gave a very broad peak with a $T_m$ of 97.4°C (Figure 4). The unfolding of LD and LDI was irreversible judged by the lack of area recovery after rescanning, which precluded a full thermodynamic analysis. Notably, DSC of LD:LDI revealed two peaks with assigned $T_m$ of 77.4°C and ca. 100°C, respectively. The first peak is ascribed to the dissociation of the complex and unfolding of LD, while the higher temperature transition resembled the unfolding of LDI (Figure 4). Thus, the complex formation provides substantial stabilization to LD, manifested in more than 11°C increase in $T_m$ compared to free LD. The conformational stability of LDI was slightly lower at pH 6.0 and 8.0 with $T_m$ values of 95.2°C and 86.5°C, respectively.

Hot Spots of LD:LDI Interaction

The LD:LDI interface was examined for key residues involved in the complex formation and residues involved in both hydrophilic and hydrophobic interactions was identified. The hydrophobic pair, Leu-41 and Val-42, interacting with LD residues of the active site cleft were selected for site-directed mutagenesis together with Arg-34 and Arg-38 of which Arg-38 interacts with two of the three catalytic residues of LD (Asp-473 and Glu-510). The LDI single mutants R34A, R38A, R38W, L41G, L41W, V42D, and a double mutant L41G-V42D were evaluated for reflecting the contribution of the four selected LDI residues to the interaction with LD (Figure 2). The LDI mutants were purified to homogeneity and their structural integrity
was confirmed using circular dichroism. The LD interaction with the LDI variants was analyzed by SPR at 25°C (Table 4). While $k_{on}$ for the LDI variants and wild type was essentially unchanged, $k_{off}$ increased for all variants except LDI-L41W (Table 4). Modest 12–33-fold affinity decrease occurred with R34A, R38A, and R38W, whereas L41G and V42D, resulted in pronounced loss of affinity (Table 4). Hence, $K_D$ for LDI V42D and LDI L41G-V42D were determined from steady state equilibration data, as $k_{off}$ was too fast to be modeled (see Supplemental Figure 5 online). $K_D$ for the double mutant L41G-V42D increased 4.8·10^5 fold relative to wild type LDI (Table 4), whereas $K_D$ for the corresponding single L41G and V42D variants increased 10^2 and 4.0·10^3 fold, respectively. These data emphasize the critical roles of LDI Leu-41 and Val-42, and their concerted role in formation of the LD:LDI complex.

The LD variants D730R and D730W probed the electrostatic interaction with LDI at a position distant from the active site. Asp-730 interacts with a positively charged pocket on the surface of LDI via a hydrogen bond (2.72 Å) and a salt bridge (3.97 Å) to Arg-34, and a hydrogen bond to Arg-84 (3.2 Å) (Figure 2C and Figure 2F). Noticeably mutation of Asp-730 (23.1 Å from the catalytic nucleophile) did not affect the kinetics of LD in the hydrolysis of pullulan. The D730R variant, however, showed 10-fold decrease in $k_{on}$, which was the largest effect on $k_{on}$ observed for any single mutant; $k_{on}$ of D730W was only reduced 1.5 fold (Table 4). $K_D$ of D730W and D730R was 8 and 171 fold higher, respectively, than for wild type LD binding to LDI (Table 4).
DISCUSSION

LDI from barley is the only characterized proteinaceous inhibitor of a debranching enzyme even though numerous pullulanase-like enzymes have been found both in seeds and in leaves of different plants; barley (Manners and Yellowlees, 1971; MacGregor et al., 1994); maize (Beatty et al., 1999; Wu et al., 2002; Dinges et al., 2003; Li et al., 2009); mung bean (Morinaga et al., 1997); oat (Dunn and Manners, 1975; Yamada, 1981); pea (Zhu et al., 1998); rice (Yamasaki et al., 2008; Li et al., 2009); sorghum (Hardie et al., 1976); spinach (Ludwig et al., 1984; Henker et al., 1998; Renz et al., 1998); sugar beet (Li et al., 1992); and wheat (Repellin et al., 2008). An early study indicated the presence of proteinaceous inhibitors targeting LDs (pullulanase-type debranching enzymes) from mature cereal grains; oat, wheat, barley and rye since the pullulanase activity in flour samples increased by incubating the samples with sodium dithionite or papain as compared with assay buffer alone (Yamada, 1981). In the same study, oat pullulanase was purified to homogeneity from flour in a buffer with sodium dithionite, resulting in much higher specific activity as compared with oat pullulanase purified previously without any reducing agent present (Dunn and Manners, 1975).

The BLAST search with LDI against all Poaceae sequences revealed putative LDIs only in wheat, Brachypodium distachyon, and rice. MacGregor et al. demonstrated LD inhibitory activity in wheat (MacGregor et al., 1995), and our bioinformatic analysis suggests that the LDI homologue pUP88 from wheat possesses LD inhibitory activity. Furthermore, relatively high levels of LD inhibitory activity were detected in hard red spring wheat, durum wheat, rye and triticale (MacGregor et al., 1995). Low LD inhibition levels were moreover measured in oats, while no LD inhibitory activity was demonstrated in corn, pearl millet, sorghum or rice (MacGregor et al., 1995), but the occurrence of LDI-like inhibitors cannot be excluded as the present sequence analysis indicated a putative LD inhibitor from rice.

Comparison of the LD:LDI Structure and Inhibition Mechanism to other Enzyme:Plant Inhibitor Complexes

The crystal structure of the LD:LDI complex gives not only insight into the inhibition of a starch debranching enzyme by an endogenous inhibitor, but also new surprising knowledge about a protein with a scaffold widely found in nature. The crystal structure of the LD:LDI complex confirmed that the two proteins form a 1:1 complex consistent with earlier data from electrospray time-of-flight mass spectrometry and enzymatic assays (MacGregor et al., 2003; Jensen et al., 2011). The binding orientation of LDI to LD was entirely different from the binding orientation of related α-amylase inhibitors from ragi and wheat in complex with α-amylase. Thus the N-terminus of LDI does not participate in the LD interaction, as opposed to TMA:RBI, where the N-terminal residue (Ser-1), interacts with all three catalytic site residues of TMA (Strobl et al., 1998). A similar binding orientation as seen for the TMA:RBI complex was reported for wheat inhibitor 0.28 AI binding to TMA (Payan, 2004). RBI Ser-1 is conserved in LDI and wheat pUP88 (Figure 1B), but preceded by the tri-peptide TLE. LDI-like proteins from Brachypodium distachyon and rice lack the Ser-1–Val-2 motif. These differences probably ensure that LDI-like proteins do not inhibit α-amylases.
**Inhibition Mechanism of LD by LDI**

The large drop in $C_p (\Delta C_p^\circ = -3.2 \text{ kJ K}^{-1} \text{ mol}^{-1})$ and the favorable entropy change associated with LD:LDI complex formation (ca 50% of total free energy of binding) is consistent with the burial of considerable hydrophobic solvent accessible surface area upon complex formation (Stites, 1997). This is supported by the LD:LDI structure showing Arg-38 that interacts with the catalytic residues being flanked by a hydrophobic patch spanning the active site at the center of the protein interface (Figure 3C). This is markedly different from the interface of TMA:RBI, here the inhibitor has a mainly positively charged surface potential including the N-terminal segment that inserts into the narrower active site cleft of the α-amylase (Figure 3B). No thermodynamic data are available for the TMA:RBI interaction, however the ordering of the N-terminal segment of RBI upon binding to TMA possibly alters the thermodynamic signature of the complex formation by introducing a large entropic penalty and subsequent enthalpic compensation. Remarkably, the binding of the β-trefoil-fold protein barley α-amylase/subtilisin inhibitor (BASI) that inhibits barley AMY2 with a sub-nanomolar affinity is entirely enthalpically driven (Nielsen et al., 2003), owing to an extensive solvent mediated hydrogen bonding network and a fully solvent coordinated Ca$^{2+}$ at the center of the interface (Vallee et al., 1998). The rationale behind the thermodynamic and structural features of the interface in the case of LD:LDI is unclear, but it is tempting to speculate that dehydration accompanying the hydrophobic cluster at the LD:LDI interface is energetically favorable, however the contribution possibly also comes from ordering of flexible LDI loops or ordering of a large number of solvent molecules in the open LD active site cleft that distinguishes debranching enzymes from α-amylases. Electrostatics by contrast are clearly not critical for the LD:LDI interaction as suggested by i) the modest sensitivity to ionic strength and pH for LD:LDI affinity, and ii) the lower magnitude of $k_{on}$ as compared to the AMY2:BASI complex formation (Bønsager et al., 2005). Nonetheless, interface peripheral charged interactions may contribute to electrostatic steering at the encounter complex distance (Sheinerman et al., 2000). Indeed the largest loss in $k_{on}$ occurred due to electrostatic complementarity being abolished for LD-D730R that as suggested by the LD:LDI structure has contact with a positively charged patch at the periphery of LDI (Arg-34) (Figure 2C).

It has been suggested that pivotal hot-spots of protein-protein interactions cluster together at the solvent occluded center of formed interfaces (>70% solvent inaccessible) and that peripheral residues rarely have a large contribution to the binding free energy (Jin and Wells, 1994; Karplus and Sali, 1995; Bogan and Thorn, 1998). This is consistent with our mutational data showing that abolishing the charged interactions between Arg-38 (56% solvent accessible) and the catalytic residues (Asp-473 and Glu-510), and between Arg-34 (56% solvent accessible) and Asp730 in LD have a modest effect on the binding affinity as compared to the dramatic affinity drop observed for the Leu-41 and Val-42 LD variants (Table 4) situated at the center of the protein interface (Figure 3C). However, the free energy change of abolishing a polar/charged interaction (e.g. R38A) by introducing an apolar/hydrophobic residue reflects both the enthalpy loss from the suppressed interaction and the solvation entropy changes, which are favorable and contribute significantly to the $\Delta G$ of complex formation for the LDI-R38A (Table 4). Nevertheless a decrease in $K_D$ is observed due to the loss in the binding energy from the Arg-38 interaction with LD. By comparison, contribution of solvation to $\Delta G$ is minimal for LDI-R34A, and the modest loss in binding energy can be attributed essentially to loss of binding
with Arg-34. In case of LDI-L41G, the solvation effect is highly unfavorable (Table 4), which together with a loss in van der Waal’s binding energy causes a large increase in \( K_D \). L41G is also likely to trap solvent at the interface and may thus violate the solvent occlusion providing dielectric and solvation conditions for the high affinity (Karplus and Sali, 1995). Similarly, introduction of a charged residue in this hydrophobic and solvent inaccessible environment is likely to poise a severe entropic penalty and to destroy the central hydrophobic cluster of the LD:LDI interface.

Hydration and protein-protein interactions effects are reported to contribute to heat capacity, but their relative contribution is not well understood and is not necessarily the same for all proteins. Taking only the hydration term into consideration, a significant decrease in heat capacity is consistent with the removal of hydrophobic surface from water (Baldwin, 1986). The thermodynamic analysis of the interaction between porcine pancreatic \( \alpha \)-amylase and the wheat inhibitor 0.19 AI showed a favorable change in entropy, suggesting that hydrophobic interactions are important in the binding (Oneda et al., 2004). The contribution of both enthalpic and entropic components to the binding free energy has been reported for other protein-protein interactions (Stites, 1997), in particular enzyme-inhibitor pairs, e.g. binding of \( \text{Streptomyces subtilisin} \) inhibitor to subtilisin from \( \text{Bacillus subtilis} \) displays a similar range of free energy change upon binding as LD:LDI (Takahashi and Fukada, 1985).

The affinity of LDI to LD is thus governed by a central hydrophobic and solvent occluded hot-spot of interaction and by avoiding the entropic penalty of ordering the N-terminal segment central in binding to \( \alpha \)-amylases. An independence of ionic strength for the LD:LDI formation (see Supplemental Table 2 online) is consistent with the complexation being mainly driven by apolar residues. \( K_D \), however, depends on pH, possibly due to charge disruption of hydrophilic interactions outside the apolar core.

**LDI Inhibition in vivo and in vitro**

LDI is implicated in the biosynthesis of starch in developing barley grains. Antisense down-regulation of barley LDI thus showed the presence of LDI to influence on starch granule size distribution, starch composition and amylpectin structure. The amylpectin chain length distribution is changed towards fewer long chains (>25 units) and more medium-long chains (10–15 units), when LDI is down-regulated. Furthermore, the ratio between A- and B-type starch granules of 1:20 (A:B) in wild type becomes 1:3 in plants with reduced LDI levels (Stahl et al., 2004). The impact of the down-regulation of LDI seems to support the hypothesis that LD is important in starch biosynthesis barley grains. The role of LD in starch biosynthesis is, however, unclear.

During the first 24 h of germination, LD and LDI are spatially separated, as LD resides mainly in the aleurone layer and LDI in the endosperm (Schroeder and MacGregor, 1998). LD has no classical signal peptide to direct secretion from the aleurone layer, and is only slowly released from aleurone cells (Schroeder and MacGregor, 1998; Burton et al., 1999; Finnie et al., 2011). Possibly, the degradative damage to the aleurone cell walls occurring 24–48 h after the onset of germination allows diffusion of LD into the starchy endosperm. If LDI is encountered at that stage, LD catalyzed hydrolysis of branched dextrins produced in starch degradation catalyzed by \( \alpha \)- and \( \beta \)-amylases is arrested. The full potential of LD will only
be unleashed if simultaneous inactivation of LDI takes place. The bound form of LD in malt extract decreases and the amount of uninhibited free LD rises during germination under wet conditions, while the amount of bound LD is unchanged (Longstaff and Bryce, 1993). The amount of intact LDI is reduced below detection limit concomitant with the reported increase in levels of free LD (Longstaff and Bryce, 1993). The loss of LDI may be due to proteolysis (Longstaff and Bryce, 1993), which probably is facilitated by LDI reduction by Trxh that per se also results in inactivation (Jensen et al., 2011).

In conclusion, the present work illuminates structural and biophysical features related to regulation of pullulanase-type plant debranching enzymes involved in mobilization of seed storage starch during germination. The mode of binding of LDI to LD highlights a novel versatility of cereal-type inhibitors featuring in a variety of regulatory protein-protein interactions in plants. The impressive conformational stability together with possibility for integration with redox regulatory networks, provide a rationale for the structural scaffold being conserved during evolution of CM proteins in spite of their many and functionally divergent roles.
METHODS

Bioinformatics
The sequence of barley LDI (ABB88573), RBI (P01087), 0.19 AI (P01085), and CHFI (P01088), were used in a protein BLAST search to query among all sequences from monocots (Poaceae) in the database provided by the National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine (http://www.ncbi.nlm.nih.gov). Sequences with E-values lower than 8×10⁻¹⁰ from each of the four searches were pooled. Sequences with >95% identity were removed using Skipredundant from the EMBOSS software suite (Rice et al., 2000). The resulting set of 45 sequences, including the four queries, were aligned using MUSCLE from the MEGA version 5 and a neighbor-joining tree was constructed with 1,000 bootstrap steps, expressed as percentiles values using MEGA version 5 (Tamura et al., 2011). The tree was visualized using Dendroscope and the alignment using ESPript (Gouet et al., 1999). Interaction surface area of the LD:LDI was calculated by the PDBePISA server (Protein interfaces, surfaces and assemblies service at European Bioinformatics Institute; Krissinel and Henrick, 2007).

Site-Directed Mutagenesis, Production and Purification of Wild type and of LD and LDI Variants
N-terminal truncation of LDI (TLESV deleted; denoted ΔV⁵LDI) was obtained (QuikChange® Lightning Site-Directed Mutagenesis Kit, Stratagene, La Jolla, CA) using primers LDI-ΔV⁵-Fw and LDI-ΔV⁵-Rv (see Supplemental Table 5 online). An LDI variant elongated at the N-terminus by Glu-Phe (denoted as EF⁵-LDI) was obtained as a result of the cloning procedure, where the EcoRI restriction site was used. The N-terminally truncated LDI (ESV deleted; denoted ΔV⁵LDI) was obtained in the ΔV⁵LDI purification.
LDI variants were produced and purified essentially as described (Jensen et al., 2011) using Pichia pastoris as host and a two step procedure consisting of affinity chromatography on a Ni-NTA column (Qiagen, Düsseldorf, Germany) followed by size exclusion chromatography (Hiload Superdex 75 16/60 column; GE Healthcare). The ΔV⁵LDI variant, however, was purified only on the Ni-NTA column and thereafter buffer-exchanged to 10 mM Bicine/NaOH pH 8.5 (Microcon, 3 kDa cut-off; Millipore). ΔV⁵LDI (770 µg mL⁻¹) gave a single band in SDS-PAGE. A single N-terminal sequence was confirmed by Edman degradation in-house (Procise 494 sequenator; Applied Biosystems, Foster City, CA).
Recombinant LD, wild type and variants, were produced by P. pastoris and purified as described (Vester-Christensen et al., 2010a). Briefly, the proteins were purified by affinity chromatography using β-cyclodextrin (β-CD) conjugated to Sepharose (GE Healthcare, Sweden) followed by gel filtration (Hiload Superdex 200 26/60; GE Healthcare, Sweden).

Circular Dichroism
CD spectra of LDI wild type an variants in 10 mM Bicine/NaOH buffer pH 8.5 were recorded at 20°C (Chirascan™ CD spectrometer, Applied Photophysics, U.K.) using a quartz cuvette with 0.1 mm optical path-length. The spectra were the average of three scans at 195–310 nm at 1.0 nm interval and scan rate 40 nm min⁻¹. The spectra were baseline corrected by subtraction of a buffer spectrum. Secondary structure
distribution was calculated using the program CDnn ver. 2.1 using the 33 base-spectra included in the distribution for deconvolution.

Surface Plasmon Resonance

The LD:LDI interaction was analyzed by SPR (BIAcore® T100; GE Healthcare). Immobilization of LDI wild type and variants on BIAcore CMS sensor chips was performed by amine coupling according to the manufacturer’s protocol using 1–10 µg mL\(^{-1}\) LDI in 5 mM sodium acetate pH 4 to a final chip density of 200–400 response units (RU). The reference cells were treated by the same procedure but without LDI. The standard analysis assay comprised 4 min association, 15 min dissociation, and 2 cycles of regeneration (2 x 60 s) injections of 10 mM glycine/HCl pH 1.5; all steps performed at a flow-rate of 30 µL min\(^{-1}\). The assay was run at 25 °C and seven LD concentrations (0.1–4 nM) in 10 mM Mes/NaOH pH 6.0, 150 mM NaCl, 0.005% P-20 surfactant. Mass transfer limitations were shown to be of no relevance by running the same analysis at 60 µL min\(^{-1}\) and comparison of rate constants.

The effect of ionic strength on the binding kinetics was determined as above at 75 mM – 1 M NaCl, while the pH dependence was evaluated in 10 mM sodium acetate (pH 5.0–5.5); 10 mM Mes/NaOH (pH 6.0–6.5); 10 mM Hapes/NaOH (pH 7.0–7.5); 10 mM Bicine/NaOH (pH 9.5–10.0); and glycine/NaOH (pH 8.0–9.0). Temperature dependence was measured at nine temperatures (10–45°C) for five LD concentrations (0.4–8 nM) using standard running buffer. Two independent data sets were collected for all conditions. All concentrations were analyzed in duplicates, except for 0.4 nM LD that was analyzed in quadruplicates and served as a control to assess the response level changes during the course of the experiment. Sensorgrams from reference cells were subtracted from sample cell sensorgrams to account for refractive index changes due to minor solvent differences and for possible nonspecific LD binding to the cell surface. The reference cell subtracted sensorgrams were corrected by subtraction of averaged blank sensorgrams (buffer injected) to account for drift specific for the sample cell. Doubly corrected sensorgrams from the different binding experiments were analyzed using BIAcore T100 Evaluation Software version 1.1. A 1:1 binding model (eq. 2) (Myszka et al., 1998) also accounting for possible mass-transport limitations was fitted globally using non-linear regression to sensorgrams generated for each set of ligand concentrations to determine the association rate constant, \(k_{\text{on}}\) (M\(^{-1}\) s\(^{-1}\)) and dissociation rate constant, \(k_{\text{off}}\) (s\(^{-1}\)), and hence \(K_D\).

\[
\begin{align*}
LD_0 &\rightarrow LD + LDI \rightleftharpoons LD:LDI \\
k_{\text{on}} &\quad k_{\text{off}}
\end{align*}
\]

Thermodynamic parameters at 25°C and standard conditions were calculated from non-linear van’t Hoff analysis (BIAcore T100 Evaluation Software version 1.1) equation (eq. 2) using kinetic data of the temperature dependence between 10–35 °C.

\[
\text{RT} \ln K_D = \Delta H^\circ - T\Delta S^\circ + \Delta C_p(T - T_0) - T\Delta C_p \ln \left( \frac{T}{T_0} \right)
\]
Differential Scanning Calorimetry

Thermal stability of LD, LDI and the LD:LDI complex was measured on a VP-DSC MicroCalorimeter (MicroCal, Northampton, MA) with a cell volume of 0.5206 mL. LD, LDI, or LD:LDI samples (0.4 or 1 mg/mL; 2 mL) were dialyzed at 4°C in Spectra/Por dialysis tubings 3500 Da cut-off (Spectrum Laboratories, Rancho Dominguez, CA) against 500 volumes of 20 mM sodium phosphate, pH 6.0, pH 6.5, or 8.0, or 20 mM sodium phosphate, pH 6.5, 150 mM NaCl. Protein samples and buffer samples (for baseline scans and reference cell) were degassed (RT, 10 min) prior to scanning between 20 and 120°C at a rate of 1°C/min. The reversibility of LDI unfolding was evaluated by rapid cooling and rescanning of LDI samples. Thermograms were corrected by subtracting reference traces and analyzed by Origin ver. 7 (OriginLab, Northampton, MA).

LD:LDI Protein Complex Formation and Crystallization

To obtain the proper stoichiometry of the complex LD and LDI were mixed in a 1:4 molar ratio (1000 µL of LD, 4.8 mg mL⁻¹ in 50 mM Mes/NaOH pH 6.6, 250 mM NaCl, 0.5 mM CaCl₂; 350 µL of LDI, 7.8 mg mL⁻¹ in 10 mM Bicine/NaOH pH 8.5, 150 mM NaCl) and left at room temperature, and centrifuged (20,000g, 4°C, 5 min) after 30 min. The volume was adjusted to 4 mL with size exclusion chromatography buffer (50 mM Mes/NaOH pH 6.6, 250 mM NaCl, 0.5 mM CaCl₂) before separation on a preequilibrated Hiload Superdex 200 26/60 column (GE Heathcare, Uppsala, Sweden) at a flow-rate of 0.5 mL min⁻¹. Fractions containing the LD:LDI complex were pooled and concentrated (Centricon, 30 kDa cut-off, Millipore, Cork, Ireland) to 550 µL (A₂₈₀=12.1).

Initial crystallization conditions of the LD:LDI complex were 0.05 M KH₂PO₄, 20% (w/v) PEG 8000 using hanging drop vapor diffusion. Seeding with the initial crystal, the crystal conditions were optimized to 24% (w/v) PEG 8000 and 0.05 M KH₂PO₄ with the addition of 0.5 µL 0.1 M NAD to the droplet consisting of 2 µL protein solution added 2 µL reservoir solution. Crystals appeared within 5 days. Crystals were cooled in N₂(l) after addition of 2 µL 25% (w/v) PEG3350, 0.05 M KH₂PO₄, and 10% glycerol to the crystallization droplet followed by removal of 2 µL solution from the droplet. This was done several times before the crystals were mounted in elliptical LithoLoops (Molecular Dimension, Suffolk, state) and flash frozen in N₂(l).

Data collection, Molecular Replacement, and Structure Refinement

X-ray diffraction data of the LD:LDI complex were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble, France), microfocus beamline ID23-2, wavelength 0.873 Å. Diffraction data were collected from four different sections of the crystals to minimize the effect of radiation damage on the data quality. The raw data were processed using MOSFLM (Leslie, 1992) and then merged and scaled using the program SCALA from CCP4i program suite (Potterton et al., 2003; Winn et al., 2011). The resulting structure factors were used for molecular replacement (MR) using Refmac5 from the CCP4i suite and the HvLD-β-CD model (PDB code: 2Y4S) including only the protein moiety and the α-helical parts of the RBI model (PDB code: 1B1U). Manual inspection, rebuilding and addition of water molecules and ions were performed with Coot (Emsley et al., 2010). During the refinement NSC restraints were used, but in the last
refinement rounds the restraints were loosened. In addition to the Coot validation functions, the final analysis of model geometry optimization was performed using the output from PROCHECK and MolProbity (Laskowski et al., 1993; Davis et al., 2007). All visualization of the structures in this paper has been made using “The PyMOL Molecular Graphics System”, version 1.3, Schrödinger, LLC, http://www.pymol.org/. The interactions between LD and LDI were analyzed using the protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (Krissinel and Henrick, 2007).

**Accession Number**
The coordinates and structure factors for crystal structure of LD:LDI have been deposit in the Protein Databank with code XXXX.

**Supplemental Data**
The following materials are available in the online version of this article:

**Supplemental Table 1.** Source organisms and accession numbers of sequences included in the multiple sequence alignment and phylogenetic tree (Figure 1 and see Supplemental Figure 1 online). The different groups of the phylogenetic tree in Figure 1A are mentioned.

**Supplemental Table 2.** Effect of ionic strength on binding kinetics of LD:LDI. Measurements were performed at 25 °C in 10 mM Mes/NaOH pH 6.0, 0.005% P-20. Seven LD concentrations (0.1–4 nM) were used for 75–1000 mM NaCl. $K_D$ is based on independent duplicate experiments.

**Supplemental Table 3.** pH dependence of binding kinetics of LD to LDI.

**Supplemental Table 4.** Temperature dependence of kinetics of binding and dissociation rate constants of the LD:LDI.

**Supplemental Table 5.** Mutagenesis primers for introduction of mutations in LDI and LD. Fw, forward primer; Rv, reverse primer.

**Supplemental Figure 1.** Multiple alignment including 45 sequences from BLAST searches with the sequences of limit dextrinase inhibitor (LDI), bifunctional $\alpha$-amylase/trypsin inhibitor from ragi (RBI), the $\alpha$-amylase inhibitor from wheat (0.19 AI), and the corn Hageman factor inhibitor (CHFI) against monocots.

**Supplemental Figure 2.** Structural alignment of the structure-determined cereal-type inhibitors and limit dextrinase inhibitor (orange): (A) bifunctional $\alpha$-amylase/trypsin inhibitor from ragi (RBI; blue; PDB code: 1B1U) and (B) RBI from the complex with $\alpha$-amylase from yellow meal worm (purple; PDB code: 1TMQ) the difference in structure at the N-terminal is encircled (dashed circle); (C) corn Hageman factor inhibitor (green; PDB code: 1BEA); and (D) 0.19 $\alpha$-amylase inhibitor from wheat (red; PDB code: 1HSS). The loop involved in protease inhibition of RBI and corn Hageman factor inhibitor is encircled.
Supplemental Figure 3. Structural alignment of the active site residues from barley limit dextrinase (LD) in complex with LDI and uncomplexed LD (PDB code: 4AIO).
Residues of complexed LD (white sticks), which are in contact with LDI (distance <4.0 Å), are superimposed with the corresponding residues of uncomplexed LD (green sticks). The residues, which adopt different rotamers, are encircled.

Supplemental Figure 4. The kinetic values from the SPR analysis of the pH (A) and temperature (B) dependence of the complex formation of the LD:LDI complex.

Supplemental Figure 5. Representative plots of single SPR datasets.
(A) 1:1 binding model (black line) fitted to the SPR data (orange dashed line) from LD binding to different LDI variants including wild type.
(B) Top: Sensorgram from the SPR analysis of LD binding to the LDI-V42D variant with the points used for the steady-state fit indicated by a cross. Bottom: Steady-state plot of data from a triple determination.
(C) Top: Sensorgram from SPR analysis of LD binding to the LDI-L41GV42D variant. The points used for the steady-state fit are indicated by a cross. Bottom: Steady-state plot of data from a double determination.
(D) 1:1 binding model (black) fitted to SPR data (orange dashed line) from analysis of the binding of the two LD variants to wild type LDI.

ACKNOWLEDGMENTS

Anne Blicher, Technical University of Denmark (DTU), is acknowledged for performing amino acid analysis and N-terminal sequence analysis. Jose A. Cuesta Seijo, Carlsberg Research Center, is acknowledged for the collection of the final dataset at the European Synchrotron Radiation Facility (ESRF). ESRF and ESRF staff as well as MAX-lab, Sweden, and MAX-lab staff are thanked for beam time and assistance. The DANSCTT program from the Danish National Research Council for travel support. Novo Nordisk is thanked for circular dichroism instrument time.

This work was supported by The Carlsberg Foundation, The Danish Council for Independent Research | Natural Sciences (Biacore instrument), Ph.D. stipends from DTU (to MVBC and MSM), and an Oticon foundation M.Sc. scholarship (to JMJ).

AUTHOR CONTRIBUTIONS

Marie S. Møller: Designed experiments, performed research, analyzed data, and wrote the paper
Malene B. Vester-Christensen: Designed experiments, performed research, analyzed data, and wrote the paper
Johanne M. Jensen: Designed experiments, performed research and analyzed data
Maher Abou Hachem: Designed the research, designed experiments, analyzed data and wrote the paper
Anette Henriksen: Designed the experiment, analyzed data, and wrote the paper
Birte Svensson: Designed the research, and wrote the paper
REFERENCES


Table 1. Data Collection and Refinement Statistics for the LD:LDI Complex. Values in parenthesis are referring to the outer resolution shell.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>European Synchrotron Research Facilities beamline</td>
<td>ID23-2</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.873</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>157.76–2.67 (2.85–2.67)</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C222,</td>
<td></td>
</tr>
<tr>
<td>Unit-cell parameters (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>167.0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>168.6</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>157.7</td>
<td></td>
</tr>
<tr>
<td>No. observed reflections</td>
<td>294667 (40824)</td>
<td></td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>60740 (8784)</td>
<td></td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>24.83</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.4 (99.7)</td>
<td></td>
</tr>
<tr>
<td>Mean (I)/σ(mean(I))</td>
<td>8.0 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.9 (4.4)</td>
<td></td>
</tr>
<tr>
<td>( R_{merge} )</td>
<td>0.176 (0.677)</td>
<td></td>
</tr>
<tr>
<td>( R_{pim} )</td>
<td>0.089 (0.348)</td>
<td></td>
</tr>
<tr>
<td>Complex molecules/ asymmetric unit</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

| Refinement       |                         |                         |
| Used reflections | 57661                   |                          |
| \( R_{cryst}/R_{free} \) (%) | 25.7/29.3            |                          |
| Atoms            |                         |                         |
| Amino acid residues | 15126                |                          |
| Calcium ions    | 4                       |                          |
| Water molecules | 99                      |                          |
| Rmsd values from ideality |             |                          |
| Bond length (Å) | 0.006                   |                          |
| Torsion angle (°) | 0.847                |                          |
| Ramachandran plot (%) |             |                          |
| Allowed         | 99.69                   |                          |
| Disallowed      | 0.31                    |                          |
| MolProbity score | 1.46                  |                          |

\( R_{merge} = \frac{1}{|S_{hkl}|} \sum_{i} |I_i(hkl) - \langle I(hkl) \rangle|/\sum_{hkl} \sum_{i} I_i(hkl), \) where \( I_i(hkl) \) is the intensity of the \( i^{th} \) observation of reflection \( hkl \) and \( \langle I(hkl) \rangle \) is the average over all observations of reflection \( hkl \).

\( R_{pim} \) is the multiplicity weighted \( R_{merge} \) (Weiss, 2001).

\(^{1}\)MolProbity score is a log-weighted combination of a clash-score, percentage Ramachandran not favored and percentage bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected (Chen et al., 2010).
Table 2. Contacts and hydrogen bonds between LDI and LD calculated by the PDBePISA interface server (Krissinel and Henrick, 2007). The distances are for the complex formed by chain A and C; values in parenthesis are for the complex formed by chain B and D.

<table>
<thead>
<tr>
<th>Total interface contacts (distance &lt; 4.0 Å) between LDI and LD</th>
<th>Salt bridges/Hydrogen bonds (distance ≤ 3.5 Å) between LDI and LD</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LDI</strong></td>
<td><strong>LD</strong></td>
<td><strong>LDI</strong></td>
</tr>
<tr>
<td>G12</td>
<td>R582</td>
<td>G12 O</td>
</tr>
<tr>
<td>P16</td>
<td>F620</td>
<td>H17 Nε2</td>
</tr>
<tr>
<td>H17</td>
<td>G575, Q574, D579</td>
<td>N18 Nε2</td>
</tr>
<tr>
<td>N18</td>
<td>F572, Y573</td>
<td>T22 Oγ2</td>
</tr>
<tr>
<td>A21</td>
<td>N555, Q558</td>
<td>T25 Oγ2</td>
</tr>
<tr>
<td>T22</td>
<td>N551, F553</td>
<td>R34 Nη2</td>
</tr>
<tr>
<td>T25</td>
<td>F553</td>
<td>D730 Oδ2</td>
</tr>
<tr>
<td>I28</td>
<td>D730</td>
<td>G35 O</td>
</tr>
<tr>
<td>R34</td>
<td>E726, E729, D730</td>
<td>G35 N</td>
</tr>
<tr>
<td>G35</td>
<td>R697, E726, K727</td>
<td>S37 N</td>
</tr>
<tr>
<td>P36</td>
<td>F553, R697, K727</td>
<td>S37 Oγ</td>
</tr>
<tr>
<td>S37</td>
<td>D698, K727</td>
<td>R38 Nη1</td>
</tr>
<tr>
<td>R38</td>
<td>A438, D473, L474, E510</td>
<td>R38 Nη2</td>
</tr>
<tr>
<td>L41</td>
<td>W512, F514</td>
<td></td>
</tr>
<tr>
<td>V42</td>
<td>F553</td>
<td>E44 Oε1</td>
</tr>
<tr>
<td>E44</td>
<td>D513, F514, A515, R519</td>
<td>R45 Nη2</td>
</tr>
<tr>
<td>R45</td>
<td>D545, N551</td>
<td>R84 Nη1</td>
</tr>
<tr>
<td>R45</td>
<td>D545, N551</td>
<td>R85 Nε2</td>
</tr>
<tr>
<td>A52</td>
<td>E621</td>
<td>R85 Nη1</td>
</tr>
<tr>
<td>R84</td>
<td>D730</td>
<td>V77 Y573</td>
</tr>
<tr>
<td>D78</td>
<td>Y573</td>
<td>R85 Nη2</td>
</tr>
</tbody>
</table>

*N.D. = not detected*
Table 3. Structural similarities between core Cα atoms of LDI (chain C; 101 Cα atoms) and related proteins.

<table>
<thead>
<tr>
<th>PDB code</th>
<th>Reference</th>
<th>RMSD (Å)</th>
<th>No. Cα atoms in structure</th>
<th>Sequence identity (%)</th>
<th>Sequence similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B1U</td>
<td>(Gourinath et al., 2000)</td>
<td>0.86</td>
<td>117</td>
<td>48.4</td>
<td>57.8</td>
</tr>
<tr>
<td>1TMQ</td>
<td>(Strobl et al., 1998)</td>
<td>0.83</td>
<td>117</td>
<td>48.4</td>
<td>57.8</td>
</tr>
<tr>
<td>1BEA:chain A</td>
<td>(Behnke et al., 1998)</td>
<td>0.78</td>
<td>116</td>
<td>46.9</td>
<td>53.1</td>
</tr>
<tr>
<td>1HSS:chain A</td>
<td>(Oda et al., 1997)</td>
<td>1.18</td>
<td>111</td>
<td>25.8</td>
<td>37.1</td>
</tr>
</tbody>
</table>
Table 4. Hot Spots of the LD:LDI Interaction Analyzed by SPR.

SPR data – kinetics of binding of LD to wild type LDI and to LDI variants, and binding of LD variants to wild type LDI at 25°C, pH 6.0, 150 mM NaCl. 1Results based on steady state kinetics. 2N-terminal truncated LDI variants. 3Results based on a single experiment. 4Solvation energy gain at complex formation was calculated using the PDBEPIASA server (Krissinel and Henrick, 2007). In silico mutations for these calculations were introduced using Coot (Emsley et al., 2010).

<table>
<thead>
<tr>
<th>LDI variant</th>
<th>$k_{on}$ ($\text{M}^{-1} \text{s}^{-1}$)</th>
<th>$k_{off}$ ($\text{s}^{-1}$)</th>
<th>$K_D$ (M)</th>
<th>Relative $K_D$</th>
<th>$\Delta$G (kJ mol$^{-1}$)</th>
<th>$\Delta$G (kJ mol$^{-1}$)</th>
<th>$\Delta$G$_{solv}$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>(1.5 ± 0.02)·$10^6$</td>
<td>(6.4 ± 0.20)·$10^{-5}$</td>
<td>(4.2 ± 0.20)·$10^{-11}$</td>
<td>1</td>
<td>-59.2</td>
<td>0</td>
<td>-19.7</td>
</tr>
<tr>
<td>R34A</td>
<td>(9.4 ± 0.3)·$10^5$</td>
<td>(8.7 ± 0.11)·$10^{-4}$</td>
<td>(9.2 ± 1.4)·$10^{-10}$</td>
<td>22</td>
<td>-51.5</td>
<td>7.7</td>
<td>-19.3</td>
</tr>
<tr>
<td>R38A</td>
<td>(1.2 ± 0.01)·$10^5$</td>
<td>(1.7 ± 0.09)·$10^{-3}$</td>
<td>(1.4 ± 0.06)·$10^{-9}$</td>
<td>33</td>
<td>-50.5</td>
<td>8.8</td>
<td>-24.8</td>
</tr>
<tr>
<td>R38W</td>
<td>(1.0 ± 0.01)·$10^5$</td>
<td>(5.2 ± 0.11)·$10^{-4}$</td>
<td>(5.2 ± 0.1)·$10^{-10}$</td>
<td>12</td>
<td>-52.9</td>
<td>6.2</td>
<td>-31.5</td>
</tr>
<tr>
<td>L41G</td>
<td>(1.3 ± 0.08)·$10^6$</td>
<td>(5.5 ± 0.38)·$10^{-3}$</td>
<td>(4.2 ± 0.04)·$10^{-9}$</td>
<td>100</td>
<td>-47.8</td>
<td>11.4</td>
<td>-15.5</td>
</tr>
<tr>
<td>L41W</td>
<td>(8.9 ± 1.3)·$10^3$</td>
<td>(6.5 ± 0.5)·$10^{-3}$</td>
<td>(7.4 ± 0.53)·$10^{-11}$</td>
<td>1.8</td>
<td>-57.8</td>
<td>1.4</td>
<td>-21.0</td>
</tr>
<tr>
<td>V42D$^a$</td>
<td>(1.7 ± 0.05)·$10^{-7}$</td>
<td>(2.0 ± 0.09)·$10^{-5}$</td>
<td>(4.0 ± 0.05)·$10^{-11}$</td>
<td>4048</td>
<td>-38.6</td>
<td>20.6</td>
<td>-16.4</td>
</tr>
<tr>
<td>L41G-V42D$^a$</td>
<td>(2.0 ± 0.09)·$10^{-5}$</td>
<td>(2.0 ± 0.09)·$10^{-5}$</td>
<td>(4.7 ± 0.09)·$10^{-10}$</td>
<td>47690</td>
<td>-26.8</td>
<td>32.4</td>
<td>-12.2</td>
</tr>
</tbody>
</table>

| ΔV$^{LDI}$ | (1.0 ± 0.0)·$10^0$ | (5.7 ± 0.2)·$10^{-3}$ | (5.5 ± 0.2)·$10^{-14}$ | 1 | -58.5 | 0.7 |
| ΔE$^{LDI}$ | (9.1 ± 0.1)·$10^5$ | (6.6 ± 0.1)·$10^{-3}$ | (7.2 ± 0.3)·$10^{-11}$ | 1.7 | -57.8 | 1.4 |
| EF-LDI$^{bc}$ | 1.4·$10^6$ | 1.1·$10^{-4}$ | 8.4·$10^{-11}$ | 2 | -57.5 | 1.7 |

LD variant

| D730W       | (1.1±0.04)·$10^0$ | (3.5±0.4)·$10^{-3}$ | (3.3±0.2)·$10^{-10}$ | 7.9 | -54.1 | 5.1 |
| D730R       | (1.5±0.07)·$10^0$ | (1.1±0.04)·$10^{-3}$ | (7.2±0.04)·$10^{-9}$ | 171 | -46.4 | 12.8 |
FIGURE LEGENDS

Figure 1. LDI and Related Cereal-type Inhibitors.
(A) Phylogenetic tree based on a multiple alignment of 45 protein sequences from Poaceae (monocots) related to LDI, RBI, 0.19 AI, and CHFI (the complete multiple sequence alignment is shown in Supplemental Figure 1 online). The tree depicts the clustering of the different types of cereal-type inhibitors annotated based on the characterized cereal-type inhibitors: LDI-like proteins, CMx subunits of tetrameric α-amylase/protease inhibitors, and dimeric α-amylase inhibitors (see Supplemental Table 1 online). The characterized proteins are indicated by stars.
(B) Multiple sequence alignment of eight cereal-type inhibitors including the three structure-determined proteins: LDI (no. 1), pUP88 (no. 2), protein from Brachypodium (no. 3), RBI (no. 4), rice LDI (no. 10), CHFI (no. 42), and dimeric and monomeric wheat α-amylase inhibitors; 0.19 AI (no. 45) and 0.28 AI (no. 43). The secondary structure of LDI is marked at the top of the alignment, while cysteines involved in disulphide bonds are numbered 1–4 and the fifth cysteine pair, which is lacking in LDI, are indicated by an x and punctured red boxes. The residues of LDI analyzed by structure guided mutagenesis are indicated by asterisks, while residues of RBI and CHFI involved in α-amylase binding or trypsin inhibition are indicated by boxes.

Figure 2. The Crystal Structure of the Complex Between LD and LDI.
(A) Overall structure of the LD:LDI complex. LDI is shown in orange. The four LD domains, CBM21-like N-domain (residues 2–124), carbohydrate binding module 48 (CBM48; residues 125–230), catalytic domain (residues 231–774) and C-domain (residues 775–884), are depicted in red, green, grey and blue, respectively. The catalytic triade; Asp-473, Glu-510, and Asp-642, are shown as sticks. Calcium ions are presented as purple spheres.
(B) Closeup of the LDI structure (orange) and the interaction surface with LD (electrostatic surface or white). The amino acid residues of LDI (orange sticks and ribbon), which are subjected to mutational analysis and their interaction with LD (white sticks and ribbon); (C) Arg-34 from loop 1 interacts with Asp-730 and has contact with Glu-729; (D) Arg-38 situated at the transition from loop 1 to helix α2 interacts directly with two catalytic residues, the nucleophile (Asp-473) and the general acid/base (Glu-510); (E) The two hydrophobic residues Leu-41 and Val-42 are in contact with Trp-512 as well as Phe-514 and Phe-553, respectively.
(F) The electrostatic potential of the area of LDI where Asp-730 of LD is in contact with both Arg-34 and Arg-84 of LDI.

Figure 3. LDI Has a Distinctly Different Binding Mode to LD as Compared with the Binding Mode of RBI to TMA.
(A) The two cereal-type inhibitors LDI (orange) and RBI (purple) align very well, but the binding modes of LDI to LD (grey) and RBI to TMA (light blue) are different. The N-terminal serine of RBI (S1, purple stick)
interacts with the catalytic triade of TMA (black sticks), while the N-terminus of LDI has no contact with LD. The catalytic triade of LD is shown in black sticks. The trypsin-binding loop of RBI is encircled.

(B) Binding mode of RBI (purple) to TMA (light blue surface) and the electrostatic potential of the interaction surface. The N-terminus of RBI is encircled.

(C) Binding mode of LDI (orange) to LD (grey surface) and the electrostatic potential of the interaction surface. The hot spot of the interaction is encircled.

**Figure 4. LD Thermal Stability is Considerably Increased by LDI Binding.**

Differential scanning calorimetry thermograms depicting the unfolding of LD (5 µM), LDI (25 µM), and a mixture of LD:LDI (1:5, 5:25 µM) at pH 6.5 and a scan rate of 1 °C/min. The thermograms show major stabilization of LD in the presence of LDI manifested by an increase of 11.5 °C in the unfolding temperature ($T_{m}$).
**Supplemental Data.** Møller et al. 2012. *Structural rationale for regulation of barley limit dextrinase activity by the endogenous proteinaceous limit dextrinase inhibitor*

**Supplemental Table 1.** Source organisms and accession numbers of sequences included in the multiple sequence alignment and phylogenetic tree (Figure 1 and see Supplemental Figure 1 online). The different groups of the phylogenetic tree in Figure 1A are mentioned.

<table>
<thead>
<tr>
<th>#</th>
<th>Group</th>
<th>Accession ID</th>
<th>Organism name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDI-like inhibitors</td>
<td>ABB88573</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>CAA68248</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>XP_003561291</td>
<td><em>Brachypodium distachyon</em></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>ABK34477</td>
<td><em>Oryza sativa Indica gr.</em></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>NP_001147201</td>
<td><em>Zea mays</em></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>XP_002459556</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>NP_001059199</td>
<td><em>Oryza sativa Japonica gr.</em></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>XP_002459322</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>P81367</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>XP_002461685</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>XP_002461687</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>P01088</td>
<td><em>Zea mays</em></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>P01087</td>
<td><em>Eleusine coracana</em></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>XP_002461684</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>XP_002459323.1</td>
<td><em>Sorghum bicolor</em></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>S51811</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>44</td>
<td>CMx subunits of tetrameric α-amylase/protease inhibitors</td>
<td>1312252B</td>
<td><em>Secale cereale</em></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>CAA11030</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>CAA67193</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>CAA11028</td>
<td><em>Hordeum vulgare subsp. spontane</em></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>AAZ67071</td>
<td><em>Secale cereale</em></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>CAA42453</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>P32936</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>P17314</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>CAA49536.1</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>P34951.2</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>P83207</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>P28041</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>P16850</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>P16851</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>NP_001059191</td>
<td><em>Oryza sativa Japonica gr.</em></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>NP_001059192</td>
<td><em>Oryza sativa Japonica gr.</em></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>XP_003561293</td>
<td><em>Brachypodium distachyon</em></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>1208404A</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>Q01881</td>
<td><em>Oryza sativa Japonica gr.</em></td>
</tr>
<tr>
<td>43</td>
<td>0.28 AI monomeric inhibitor</td>
<td>P01083</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>41</td>
<td>Dimeric α-amylase inhibitors</td>
<td>ACP40903</td>
<td><em>Eremopyrum bonaepartis</em></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>ACP40883</td>
<td><em>Triticum timopheevii sups. armeniacum</em></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>ACP40915</td>
<td><em>Secale cereale</em></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>ABI54565</td>
<td><em>Aegilops sharonensis</em></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>ACP40690</td>
<td><em>Triticum dicoccoides</em></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>ACP40801</td>
<td><em>Triticum dicoccoides</em></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>P01085</td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>ACP40906</td>
<td><em>Eremopyrum bonaepartis</em></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>ACP40674</td>
<td><em>Triticum dicoccoides</em></td>
</tr>
</tbody>
</table>
Supplemental Table 2. Effect of ionic strength on binding kinetics of LD:LDI. Measurements were performed at 25 °C in 10 mM Mes/NaOH pH 6.0, 0.005% P-20. Seven LD concentrations (0.1–4 nM) were used for 75–1000 mM NaCl. $K_D$ is based on independent duplicate experiments.

<table>
<thead>
<tr>
<th>NaCl (mM)</th>
<th>$k_{on}$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_{off}$ (s$^{-1}$)</th>
<th>$K_D$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>$(9.0 \pm 0.0) \cdot 10^5$</td>
<td>$(3.4 \pm 1.5) \cdot 10^{-8}$</td>
<td>$(3.7 \pm 1.7) \cdot 10^{-14}$</td>
</tr>
<tr>
<td>150</td>
<td>$(1.5 \pm 0.02) \cdot 10^6$</td>
<td>$(6.4 \pm 0.2) \cdot 10^{-5}$</td>
<td>$(4.2 \pm 0.2) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>300</td>
<td>$(1.7 \pm 0.03) \cdot 10^6$</td>
<td>$(4.7 \pm 0.1) \cdot 10^{-5}$</td>
<td>$(2.7 \pm 0.1) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>1000</td>
<td>$(9.4 \pm 0.1) \cdot 10^5$</td>
<td>$(7.0 \pm 0.1) \cdot 10^{-5}$</td>
<td>$(7.4 \pm 0.1) \cdot 10^{-11}$</td>
</tr>
</tbody>
</table>
**Supplemental Table 3.** pH dependence of binding kinetics of LD to LDI analyzed by SPR. Measurements were performed at 25°C in appropriate buffers (Methods section), containing 150 mM NaCl and 0.005% P-20. Seven LD concentrations (0.1–4 nM) were used. $K_D$ determination is based on independent duplicate experiments.

<table>
<thead>
<tr>
<th>pH</th>
<th>$k_{on}$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_{off}$ (s$^{-1}$)</th>
<th>$K_D$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>(6.1 ± 0.6)·10$^3$</td>
<td>(2.6 ± 0.3)·10$^{-4}$</td>
<td>(4.3 ± 0.12)·10$^{-10}$</td>
</tr>
<tr>
<td>5.5</td>
<td>(9.7 ± 0.1)·10$^3$</td>
<td>(8.6 ± 0.1)·10$^{-5}$</td>
<td>(8.9 ± 0.01)·10$^{-11}$</td>
</tr>
<tr>
<td>6.0</td>
<td>(1.4 ± 0.04)·10$^6$</td>
<td>(5.3 ± 0.1)·10$^{-5}$</td>
<td>(3.7 ± 0.2)·10$^{-11}$</td>
</tr>
<tr>
<td>6.5</td>
<td>(1.9 ± 0.02)·10$^6$</td>
<td>(5.1 ± 0.0)·10$^{-5}$</td>
<td>(2.8 ± 0.02)·10$^{-11}$</td>
</tr>
<tr>
<td>7.0</td>
<td>(1.8 ± 0.2)·10$^6$</td>
<td>(6.2 ± 0.2)·10$^{-5}$</td>
<td>(3.5 ± 0.2)·10$^{-11}$</td>
</tr>
<tr>
<td>7.5</td>
<td>(1.6 ± 0.0)·10$^6$</td>
<td>(8.1 ± 0.3)·10$^{-5}$</td>
<td>(4.9 ± 0.2)·10$^{-11}$</td>
</tr>
<tr>
<td>8.0</td>
<td>(1.7 ± 0.01)·10$^6$</td>
<td>(1.2 ± 0.0)·10$^{-4}$</td>
<td>(7.4 ± 0.04)·10$^{-11}$</td>
</tr>
<tr>
<td>8.5</td>
<td>(1.6 ± 0.01)·10$^6$</td>
<td>(2.2 ± 0.01)·10$^{-4}$</td>
<td>(1.4 ± 0.02)·10$^{-10}$</td>
</tr>
<tr>
<td>9.0</td>
<td>(1.2 ± 0.1)·10$^6$</td>
<td>(2.7 ± 0.05)·10$^{-4}$</td>
<td>(2.2 ± 0.2)·10$^{-10}$</td>
</tr>
<tr>
<td>9.5</td>
<td>(1.4 ± 0.03)·10$^6$</td>
<td>(3.1 ± 0.2)·10$^{-4}$</td>
<td>(2.3 ± 0.2)·10$^{-10}$</td>
</tr>
<tr>
<td>10.0</td>
<td>(1.5 ± 0.2)·10$^6$</td>
<td>(5.0 ± 0.4)·10$^{-4}$</td>
<td>(3.5 ± 0.09)·10$^{-10}$</td>
</tr>
</tbody>
</table>
**Supplemental Table 4.** Temperature dependence of kinetics of binding and dissociation rate constants of the LD:LDI complex formation. Measurements were performed at different temperatures in 10 mM Mes/NaOH pH 6.0, 150 mM NaCl, 0.005% P-20. Five LD concentrations (0.4–8 nM) were used. $K_D$ is based on independent duplicate experiments.

<table>
<thead>
<tr>
<th>Temp. (°C)</th>
<th>$k_{on}$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_{off}$ (s$^{-1}$)</th>
<th>$K_D$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$(5.6 \pm 0.0) \cdot 10^5$</td>
<td>$(4.9 \pm 0.3) \cdot 10^{-7}$</td>
<td>$(8.8 \pm 0.6) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>15</td>
<td>$(6.5 \pm 0.0) \cdot 10^5$</td>
<td>$(5.7 \pm 0.2) \cdot 10^{-3}$</td>
<td>$(8.8 \pm 0.3) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>20</td>
<td>$(7.6 \pm 0.0) \cdot 10^5$</td>
<td>$(6.6 \pm 0.2) \cdot 10^{-3}$</td>
<td>$(8.6 \pm 0.2) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>25</td>
<td>$(8.8 \pm 0.1) \cdot 10^5$</td>
<td>$(8.3 \pm 0.1) \cdot 10^{-3}$</td>
<td>$(9.5 \pm 0.05) \cdot 10^{-11}$</td>
</tr>
<tr>
<td>30</td>
<td>$(1.0 \pm 0.01) \cdot 10^6$</td>
<td>$(1.2 \pm 0.02) \cdot 10^{-4}$</td>
<td>$(1.1 \pm 0.03) \cdot 10^{-10}$</td>
</tr>
<tr>
<td>35</td>
<td>$(1.1 \pm 0.0) \cdot 10^6$</td>
<td>$(2.0 \pm 0.0) \cdot 10^{-4}$</td>
<td>$(1.8 \pm 0.0) \cdot 10^{-10}$</td>
</tr>
<tr>
<td>37</td>
<td>$(1.2 \pm 0.01) \cdot 10^6$</td>
<td>$(2.6 \pm 0.03) \cdot 10^{-4}$</td>
<td>$(2.3 \pm 0.0) \cdot 10^{-10}$</td>
</tr>
<tr>
<td>40</td>
<td>$(1.3 \pm 0.01) \cdot 10^6$</td>
<td>$(4.0 \pm 0.03) \cdot 10^{-4}$</td>
<td>$(3.2 \pm 0.04) \cdot 10^{-10}$</td>
</tr>
<tr>
<td>45</td>
<td>$(1.4 \pm 0.2) \cdot 10^6$</td>
<td>$(8.0 \pm 0.5) \cdot 10^{-4}$</td>
<td>$(5.8 \pm 1.1) \cdot 10^{-10}$</td>
</tr>
</tbody>
</table>
**Supplemental Table 5.** Mutagenesis primers for introduction of mutations in LDI and LD. Fw, forward primer; Rv, reverse primer.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDI-ΔV5-Fw</td>
<td>GAGAAAAAGAGGGCTGAAGCTAAGGACGAGTGCCAACCAGGGGT</td>
</tr>
<tr>
<td>LDI-ΔV5-Rv</td>
<td>ACCCCCTGGTGGCACAAGCCTGGCTCTTTAGCTTCAGCCTCTCTTTTCTC</td>
</tr>
<tr>
<td>LDI-R34A-Fw</td>
<td>CGGGTCTGCGGGCAGGCCGGTCCAGCGCCGCC</td>
</tr>
<tr>
<td>LDI-R34A-Rv</td>
<td>GGGCCGGCTGGGACCGCGCCGCAGACCGCG</td>
</tr>
<tr>
<td>LDI-R38A-Fw</td>
<td>CGCGGTCCAGCGCCATGCTGGTGAAGGAAGG</td>
</tr>
<tr>
<td>LDI-R38A-Rv</td>
<td>CTTCACCACAGCATGGGCCGGCTGGGACGGCG</td>
</tr>
<tr>
<td>LDI-R38W-Fw</td>
<td>CGCGGTCCAGCTGGCCCATGCTGGTGAAGGAG</td>
</tr>
<tr>
<td>LDI-R38W-Rv</td>
<td>CGCGGTCCAGCTGGCCCATGCTGGTGAAGGAG</td>
</tr>
<tr>
<td>LDI-L41G-Fw</td>
<td>AGCCGGCCCATGGGGGTGAAGGACCGTGCTGC</td>
</tr>
<tr>
<td>LDI-L41G-Rv</td>
<td>AGCCGGCCCATGGGGGTGAAGGACCGTGCTGC</td>
</tr>
<tr>
<td>LDI-L41W-Fw</td>
<td>GCAGACCGCTGCTCTTACACCCCATGGGGCGGT</td>
</tr>
<tr>
<td>LDI-L41W-Rv</td>
<td>GCAGACCGCTGCTCTTACACCCCATGGGGCGGT</td>
</tr>
<tr>
<td>LDI-V42D-Fw</td>
<td>CGGCCCATGCTGGATAGGACCGTGCTGGCC</td>
</tr>
<tr>
<td>LDI-V42D-Rv</td>
<td>CGGCCCATGCTGGATAGGACCGTGCTGGCC</td>
</tr>
<tr>
<td>LDI-L41G-V42D-Fw</td>
<td>AGCCGGCCATGGGGGATAAGGACCGTGCTGGCC</td>
</tr>
<tr>
<td>LDI-L41G-V42D-Rv</td>
<td>CGGCCCATGCTGGATAGGACCGTGCTGGCC</td>
</tr>
<tr>
<td>LD-D730W-Fw</td>
<td>CCAAGTGAAAAGAACGAATGGGAATTGGCCCTGATGAAACC</td>
</tr>
<tr>
<td>LD-D730W-Rv</td>
<td>GGTTCATCACGGCCATTTTCATCTGATTTTCCTTACTTG</td>
</tr>
<tr>
<td>LD-D730R-Fw</td>
<td>CCAAGTGAAAAGAACGAATGGGAATTGGCCCTGATGAAACC</td>
</tr>
<tr>
<td>LD-D730R-Rv</td>
<td>GGTTCATCACGGCCATTTTCATCTGATTTTCCTTACTTG</td>
</tr>
</tbody>
</table>
Figure 4

![Graph showing temperature vs. specific heat capacity (miliCal/°C)](image)

- LD
- LDI
- LD:LDI

Temperature (°C)

Cₚ (miliCal/°C)
Supplemental Figure 3
Supplemental Figure 4

A

\( k_{\text{on}} \) (M\(^{-1}\) s\(^{-1}\))

\( k_{\text{off}} \) (s\(^{-1}\))

\( K_D \) (M)

pH

B

\( k_{\text{on}} \) (M\(^{-1}\) s\(^{-1}\))

\( k_{\text{off}} \) (s\(^{-1}\))

\( K_D \) (M)

Temperature (°C)
Supplemental Figure 5

A

Wild type LDI

LDI-R34A

LDI-R38A

LDI-R38W

LDI-L41G

LDI-L41W

B

LDI-V42D

C

LDI-L41G-V42D

D

LD-D730W

LD-D730R
Chapter 5

Concluding remarks and perspectives

The crystal structure of LD in complex with the branched substrate \(6^2-\alpha\)-maltotriosyl-maltotriose (M3-M3) presented in this thesis represents the first crystal structure of a debranching enzyme in complex with a natural substrate, i.e. a branched \(\alpha\)-glucan (limit dextrin). The structure provided a basis for comparison between LD and other pullulanases from GH13 subfamilies 12–14 both at protein sequence level and at structure level. In addition, the LD:M3-M3 structure was compared with structures of other \(\alpha\)-1,6-acting enzymes. Based on the structure comparisons it is concluded that the topology of the active site accounts for some of the differences in substrate preferences rather than differences in the residues directly involved in substrate binding alone. Met440 of LD was previously suggested to be one of those substrate specificity determinants, since it could cause a steric hindrance for the binding of the polysaccharide amylopectin. But mutational analysis reported in the present thesis showed that the catalytic efficiency of LD hydrolysis of amylopectin was reduced as compared with wild type, when Met440 was substituted with glycine. Based on the structure comparisons, Phe553 of LD is suggested to cause a steric hindrance together with residues from a loop of LD, which is a barrier after the subsite +2. Pullulanases from GH13 subfamilies 12 and 14 have a more open active site cleft at the + subsites. Mutagenesis analysis is, however, needed to confirm this correlation between structural features and substrate specificity.

The present PhD project resulted in the crystal structure of the complex between LD and LDI solved at 2.7 Å, which represents the first insight into the interaction between a debranching enzyme and a cereal-type inhibitor. The structure revealed a novel mode of inhibition distinctly different from that of the interaction between other cereal-type inhibitors and \(\alpha\)-1,4-acting enzymes (\(\alpha\)-amylases) from the same glycoside hydrolase family as LD (GH13) involving the opposite face of LDI compared to the inhibitors of \(\alpha\)-amylases. The loop of LDI, which participates in inhibition of LD, corresponds to the trypsin-binding loop of another cereal-type inhibitor, RBI. Furthermore, structure guided mutagenesis revealed the importance of a hydrophobic pair of LDI, Leu41 and Val42 for the complex formation. In addition two arginines of LDI, Arg34 and Arg38, were shown to be contributing to the high affinity of binding. Arg38 interacted with the catalytic nucleophile and the general acid/base catalyst of LD.

The structures of LD:M3-M3 and LD:LDI can provide a basis for optimisation of the degradation of starch to fermentable sugars, e.g. in relation to brewing. Besides the insight into the complex formation mechanism, the mutational analysis of residues of LD and LDI contributes to search strategies, when analysing different barley cultivars (or “lines”) for their potential for brewing or changes in starch biophysical properties. In brewing it is desirable to have access to LD variants, which are less sensitive to LDI, while maintaining the LD wild-type activity level or
having an even higher activity towards the polysaccharide amylopectin. One could state that the easiest would be to down-regulate LDI production in the barley grains. But this might not be desirable, since studies showed that down-regulation of LDI had an adverse impact on the starch biosynthesis (Stahl et al., 2004). Furthermore the DSC analysis presented in Chapter 4 suggests that the complex formation between LD and LDI enhances the thermostability of LD by 11°C. In the present study Asp730 of LD was identified to be a possible sensitivity modulator, as arginine substitution of Asp730 resulted in a 180-fold reduced affinity without changing the kinetics of LD on pullulan. Alternatively, it could be desirable to identify LD variants for which LDI inhibition could be counteracted by pH or salt adjustments. The SPR analysis of the pH, temperature, and ionic strength dependence of binding and dissociation rates of LD:LDI suggests that by modulating these parameters the release of LD from the complex can be enhanced for a higher debranching activity for biotechnological applications or the LD:LDI dissociation can be postponed in order to protect LD from heat inactivation during processes occurring prior to the need for debranching activity. Lowering the pH to 5 and increasing the temperature to 40°C will cause reduced binding affinity, while the highest affinity would be obtained at pH 6.5 in combination with low temperatures.

Finally, LDI could be a good skeleton for engineering. Proteinaceous inhibitors of carbohydrate-active enzymes have the potential to play a role in i) cereal-based food, ii) the feed sector, and iii) the agriculture sector, where development of crops with increased pathogen resistance is desirable. Furthermore α-amylase inhibitors in cereals might affect human nutrition, e.g. obesity (Juge and Svensson, 2006). The present study has shown that cereal-type inhibitors, which share the overall structure, i.e. four α-helices stabilised by at least four disulfide bonds, can inhibit activity of enzymes with significantly different active site topology, and the property of a few amino acid residues can make the difference for inhibition level. In addition, LDI is very stable as seen from the high melting temperature measured by DSC, and can be produced in high concentrations (Jensen et al., 2011). The only downside is that certain CM-proteins are known to cause both respiratory and food allergies, e.g. baker’s asthma, associated with glycosylations of the proteins (Salcedo et al., 2011).
References


119
Appendix I

Specific activity
Table I.1: Relative specific activities of a number of pullulanases on various substrates.

<table>
<thead>
<tr>
<th>Organism, enzyme name, NCBI Accession no., reference</th>
<th>GH13.12</th>
<th>Hordeum vulgare, barley LD, AAD04189 (Manners and Yellowlees, 1971)</th>
<th>Klebsiella pneumonia (Aerobacter aerogenes), PUL, AAA25124 (Yokobayashi et al., 1973)</th>
<th>Sorgothium bicolour, LD, ABK63595 (Yamasaki et al., 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pullulan</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Amylopectin</td>
<td>55 (potato)</td>
<td>$\ll 1$</td>
<td></td>
<td>34.1</td>
</tr>
<tr>
<td>Limit dextrin of amylopectin</td>
<td>6 (dextrin)</td>
<td>68</td>
<td>68</td>
<td>19.5 ($\beta$-)</td>
</tr>
<tr>
<td>Glycogen</td>
<td>25 (oyster), 24 (bovine muscle)</td>
<td>0</td>
<td>1.7 (oyster), 1.1 (rabbit liver)</td>
<td>6.3</td>
</tr>
<tr>
<td>Phytoglycogen</td>
<td>2.3 (sweet corn)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycogen $\beta$-limit dextrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soluble starch</td>
<td></td>
<td></td>
<td></td>
<td>15.8</td>
</tr>
<tr>
<td>Starch</td>
<td>130 (potato)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylose</td>
<td>28 (potato, DP=17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dextran</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-o-D-glucosyl-maltotriose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-o-maltosyl-maltotriose</td>
<td>13</td>
<td>55</td>
<td>50</td>
<td>220</td>
</tr>
<tr>
<td>6°-o-maltosyl-maltotetraose</td>
<td>123</td>
<td>171</td>
<td>210</td>
<td>300</td>
</tr>
<tr>
<td>6°-o-maltotriol-maltoltriol</td>
<td>119</td>
<td>91</td>
<td>170</td>
<td>250</td>
</tr>
<tr>
<td>6°-o-maltotriol-maltotetraose</td>
<td>243</td>
<td>112</td>
<td>260</td>
<td>400</td>
</tr>
<tr>
<td>α-D-glucosyl-cyclohexa-amylose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-D-maltosyl-cyclohexa-amylose</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>G2-α-cyclodextrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2-β-cyclodextrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table I.2: continued from previous page

<table>
<thead>
<tr>
<th>Organism, enzyme name, NCBI Accession no., reference</th>
<th>GH13,13</th>
<th>GH13,14</th>
<th>No subfamily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinacia oleracea, spinach, PUL (Ludwig et al., 1984)</td>
<td>Spinacia oleracea, spinach, PUL (Renz et al., 1998)</td>
<td>Zea mays, maize, PUL, AAD11599 (Wu et al., 2002)</td>
<td>Bacillus acidopullulyticus, (two active forms F1 and F2), CAC60156 (Kusano et al., 1988)</td>
</tr>
<tr>
<td>Zea mays, maize, PUL, AAD11599, CAC60156 (Kusano et al., 2000)</td>
<td></td>
<td></td>
<td>Desulfurococcus mucosus, PUL type II, recombinant, AAG31003 (Duffner et al., 2000)</td>
</tr>
<tr>
<td>Avena sativa, oat, LD, (Dunn and Manners, 1975)</td>
<td></td>
<td></td>
<td>Clostridium thermohydrosulfuricum, (Thermoanaerobacter thermohydrosulfuricus), PUL, CAZ78897 (Saha et al., 1988)</td>
</tr>
</tbody>
</table>

**Assay conditions:** pH, temperature, substrate concentration

<table>
<thead>
<tr>
<th>Substrate</th>
<th>GH13,13</th>
<th>GH13,14</th>
<th>No subfamily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pullulan 100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Amylopectin 17.5</td>
<td>16 (maize)</td>
<td>20</td>
<td>16/15 (rice)</td>
</tr>
<tr>
<td>Limit dextrin of amylopectin 0</td>
<td>0</td>
<td>0</td>
<td>75 (oyster)</td>
</tr>
<tr>
<td>Glycogen 0</td>
<td>0</td>
<td>2/1</td>
<td>75 (oyster)</td>
</tr>
<tr>
<td>Phytoglycogen 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycogen β-limit dextrin 26/30 (oyster), 32/46 (rabbit liver)</td>
<td></td>
<td>77 (oyster)</td>
<td></td>
</tr>
<tr>
<td>Soluble starch 37</td>
<td>35</td>
<td>10/10</td>
<td>52</td>
</tr>
<tr>
<td>Starch 62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylose 0</td>
<td>0</td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Dextran 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panose 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-α-D-glucosyl-maltotriose 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-α-maltosyl-maltotriose 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-α-maltotetraose 290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-α-maltotriosyl-maltotriose 220</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6°-α-maltotriosyl-maltotetraose 380</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-D-glucosyl-cyclohexa-amylose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α-D-malto-amylose-cyclohexa-amylose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2-α-cyclodextrin 3/3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2-β-cyclodextrin 1/1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table I.3: continued from previous page

<table>
<thead>
<tr>
<th>Organism, enzyme name, NCBI Accession no., reference</th>
<th>Assay conditions: pH, temperature, substrate concentration</th>
<th>No subfamily</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacillus sp. S-1, PUL, (Kim et al., 1993b)</strong></td>
<td>pH 9.0, 50°C, 2% (w/v)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Micrococcus sp. Y-1, PUL, (Kim et al., 1993a)</strong></td>
<td>pH 10.0, 50°C, 1% (w/v)</td>
<td>38</td>
</tr>
<tr>
<td><strong>Pisum sativum L., LD, (Yellowlees, 1980)</strong></td>
<td>pH 6.0, 60°C, 2% (w/v)</td>
<td>35</td>
</tr>
<tr>
<td><strong>Pisum sativum L., developing pea embryos, PUL, (Zhu et al., 1998)</strong></td>
<td>pH 6.0, 37°C, 20 mg/ml</td>
<td>26</td>
</tr>
<tr>
<td><strong>Thermus caldophilus GK-24 PUL-1, (Kim et al., 1996)</strong></td>
<td>pH 8.0, 73°C, 2% (w/v)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Pullulan</strong></td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Amylopectin</strong></td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td><strong>Limit dextrin of amylopectin</strong></td>
<td>34</td>
<td>130</td>
</tr>
<tr>
<td><strong>Glycogen</strong></td>
<td>18.3 (α- and β-limit dextrin)</td>
<td>36.3 (α- and β-)</td>
</tr>
<tr>
<td><strong>Glycogen β-limit dextrin</strong></td>
<td>46.3 (α- and β-limit dextrin)</td>
<td>16.7 (β-)</td>
</tr>
<tr>
<td><strong>Glycogen</strong></td>
<td>13.0 (oyster), 44 (oyster)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Soluble starch</strong></td>
<td>8.6</td>
<td>10.6</td>
</tr>
<tr>
<td><strong>Starch</strong></td>
<td>64</td>
<td>2.4</td>
</tr>
<tr>
<td><strong>Amylose</strong></td>
<td>N.D.</td>
<td>62</td>
</tr>
<tr>
<td><strong>Dextran</strong></td>
<td>N.D.</td>
<td>No detectable activity</td>
</tr>
<tr>
<td><strong>Panose</strong></td>
<td>N.D.</td>
<td>No detectable activity</td>
</tr>
<tr>
<td><strong>6β-α-D-glucosyl-maltotriose</strong></td>
<td>N.D.</td>
<td>No detectable activity</td>
</tr>
<tr>
<td><strong>6β-α-malto-amylose</strong></td>
<td>90</td>
<td></td>
</tr>
<tr>
<td><strong>6β-α-maltosyl-maltotetraose</strong></td>
<td>230</td>
<td></td>
</tr>
<tr>
<td><strong>6β-α-maltotriose-maltotriose</strong></td>
<td>170</td>
<td></td>
</tr>
<tr>
<td><strong>G2-α-cyclohexa-amylose</strong></td>
<td>280</td>
<td></td>
</tr>
<tr>
<td><strong>G2-β-cyclohexa-amylose</strong></td>
<td>123</td>
<td></td>
</tr>
</tbody>
</table>
Appendix II

CM-proteins included in multiple alignment

Table II.1: Overview of the sequences included in the multiple sequence alignment of CM proteins (Figure 1.6).

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Organism</th>
<th>Protein type/name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q00451</td>
<td><em>Solanum lycopersicum</em> (Tomato)</td>
<td>36.4 kDa proline-rich protein (TPRP-F1)</td>
</tr>
<tr>
<td>Q39614</td>
<td><em>Cuscuta reflexa</em> (Southern Asian dodder)</td>
<td>Hybrid proline-rich proteins</td>
</tr>
<tr>
<td>Q41848</td>
<td><em>Zea mays</em> (Maize)</td>
<td>Hybrid proline-rich proteins</td>
</tr>
<tr>
<td>P93346</td>
<td><em>Nicotiana tabacum</em> (Tobacco)</td>
<td>NT16 polypeptide</td>
</tr>
<tr>
<td>P14009</td>
<td><em>Daucus carota</em> (Carrot)</td>
<td>14 kDa proline-rich protein DC2.15</td>
</tr>
<tr>
<td>Q40335</td>
<td><em>Medicago sativa</em> (Alfalfa)</td>
<td>Bimodular protein</td>
</tr>
<tr>
<td>Q01595</td>
<td><em>Zea mays</em> (Maize)</td>
<td>Cortical cell-delineating protein</td>
</tr>
<tr>
<td></td>
<td><em>Medicago sativa</em> (Alfalfa)</td>
<td>/Root-specific protein ZRP3</td>
</tr>
<tr>
<td>AAC62610</td>
<td><em>Arabidopsis thaliana</em> (Thale cress)</td>
<td>AIR1 (Auxin-Induced in Root cultures) related to membrane-cell wall linkers</td>
</tr>
<tr>
<td>AAD37833</td>
<td><em>Glycine max</em> (Soybean)</td>
<td>Hydrophobic seed protein precursor</td>
</tr>
<tr>
<td>P24565</td>
<td><em>Brassica napus</em> (Rape)</td>
<td>Napin-1A</td>
</tr>
<tr>
<td>P30233</td>
<td><em>Capparis masaikai</em> (Mabinlang)</td>
<td>Sweet protein mabinlin-2</td>
</tr>
<tr>
<td>P01086</td>
<td><em>Hordeum vulgare</em> (Barley)</td>
<td>Trpsin inhibitor CMe</td>
</tr>
<tr>
<td>Q2V8X0</td>
<td><em>Hordeum vulgare</em> (Barley)</td>
<td>Limit dextrinase inhibitor (LDI)</td>
</tr>
<tr>
<td>P01087</td>
<td><em>Eleusine coracana</em> (Indian finger millet) (Ragi)</td>
<td>Bifunctional α-amylase and trpsin inhibitor (RBI)</td>
</tr>
<tr>
<td>BAA20139</td>
<td><em>Triticum aestivum</em> (Wheat)</td>
<td>0.19 α-amylase inhibitor (0.19 AI)</td>
</tr>
<tr>
<td>P01088</td>
<td><em>Zea mays</em> (Maize)</td>
<td>Trpsin/factor XIIA inhibitor, corn Hageman factor inhibitor (CHFI)</td>
</tr>
<tr>
<td>P33432</td>
<td><em>Triticum aestivum</em> (Wheat)</td>
<td>Puroindolene-A</td>
</tr>
<tr>
<td>P19656</td>
<td><em>Zea mays</em> (Maize)</td>
<td>Non-specific lipid-transfer protein</td>
</tr>
</tbody>
</table>
## Appendix III

Multiple sequence alignment - catalytic domains of GH13_8–9 and GH13_11–14

### Table III.1: Overview of the sequences included in the multiple sequence alignment of the catalytic domains (Figure III.1).

<table>
<thead>
<tr>
<th>GH13 subfamily</th>
<th>Accession no.</th>
<th>PDB entry</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>BAF20543</td>
<td>3AMK, 3AML</td>
<td><em>Oryza sativa</em> Japonica gr.</td>
</tr>
<tr>
<td>9</td>
<td>AAA23872</td>
<td>1M7X</td>
<td><em>Escherichia coli</em> K-12</td>
</tr>
<tr>
<td></td>
<td>CAA98090</td>
<td>3K1D</td>
<td><em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>11</td>
<td>AAK42273</td>
<td>2VNC, 2VR5, 2VUY</td>
<td><em>Sulfolobus solfataricus</em></td>
</tr>
<tr>
<td></td>
<td>AAC76456</td>
<td>2WSK</td>
<td><em>Escherichia coli</em> K-12</td>
</tr>
<tr>
<td></td>
<td>P10342</td>
<td>1BF2</td>
<td><em>Pseudomonas amylofera</em></td>
</tr>
<tr>
<td>12</td>
<td>AAK74446</td>
<td>2 YA0, 2AY1, 2AY1</td>
<td><em>Streptococcus pneumoniae</em> (SpPUL)</td>
</tr>
<tr>
<td></td>
<td>AAN00998</td>
<td>3FAW, 3FAX</td>
<td><em>Streptococcus agalactiae</em> (SaPUL)</td>
</tr>
<tr>
<td></td>
<td>ACI61883</td>
<td></td>
<td><em>Streptococcus pyogenes</em></td>
</tr>
<tr>
<td></td>
<td>CAR47543</td>
<td></td>
<td><em>Streptococcus suis</em></td>
</tr>
<tr>
<td></td>
<td>AAD04189</td>
<td>2 YA5, 2Y5E, 4A10</td>
<td><em>Hordeum vulgare</em></td>
</tr>
<tr>
<td></td>
<td>AAO00771</td>
<td>2FGZ, 2FH6, 2FH8, 2FHB, 2FHC, 2FHF</td>
<td><em>Klebsiella pneumoniae</em> (KpPUL)</td>
</tr>
<tr>
<td></td>
<td>CCO17805</td>
<td></td>
<td><em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td></td>
<td>ACO64736</td>
<td></td>
<td><em>Bathycoccus prasinos</em></td>
</tr>
<tr>
<td></td>
<td>CAC09471</td>
<td></td>
<td><em>Micromonas sp.</em> RCC299</td>
</tr>
<tr>
<td></td>
<td>BAA28632</td>
<td></td>
<td><em>Oryza sativa</em> Indica gr.</td>
</tr>
<tr>
<td></td>
<td>ABO93829</td>
<td></td>
<td><em>Oryza sativa</em> Japonica gr.</td>
</tr>
<tr>
<td></td>
<td>AAS88886</td>
<td></td>
<td><em>Ostreococcus lucimarinus</em></td>
</tr>
<tr>
<td></td>
<td>ABL84490</td>
<td></td>
<td><em>Ostreococcus tauri</em></td>
</tr>
<tr>
<td></td>
<td>AAD11599</td>
<td></td>
<td><em>Triticum aestivum</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Zea mays</em></td>
</tr>
<tr>
<td>13</td>
<td>AAS47565</td>
<td></td>
<td><em>Anaerobranca gottschalkii</em></td>
</tr>
<tr>
<td></td>
<td>CAC60156</td>
<td>2WAN</td>
<td><em>Bacillus acidopullulyticus</em> (BaPUL)</td>
</tr>
<tr>
<td></td>
<td>AAC00283</td>
<td>2E8Y, 2E8Z, 2E9B</td>
<td><em>Bacillus subtilis</em> subsp. <em>subtilis</em> str. 168 (BsPUL)</td>
</tr>
<tr>
<td></td>
<td>AAD30387</td>
<td></td>
<td><em>Fervidobacterium pennivorans</em> VEN5</td>
</tr>
</tbody>
</table>
Figure III.1: Multiple sequence alignment of catalytic domains from enzymes with structures from GH13.8–9 and enzymes from GH11–14, which are enzymatic characterised and/or structure determined. All the eukaryotes defined as characterised in CAZy are included. See Table III.1 for organism information and PDB entries. Barley LD is underlined by a dashed line. The catalytic residues are indicated by stars, and their numbering refers to barley LD. The four conserved regions between GH13 enzymes (MacGregor et al., 2001) are labelled I–IV.
Appendix IV

Structure based multiple sequence alignment including the structures from GH13_8–9 and GH13_11–14

Figure IV.1: Structure based multiple protein sequence alignment of the structure determined enzymes from GH13 subfamilies 8–9 and 11–14 generated using PROMALS3D (Pei et al., 2008b). Barley LD is underlined by a dashed line. The catalytic residues are indicated by stars, and their numbering refers to barley LD. The four conserved regions between GH13 enzymes (MacGregor et al., 2001) are labelled I–IV. The sequences are coloured according to secondary structure predictions (red: α-helix, blue: β-strand). The consensus predicted secondary structures are shown indicated by h (α-helix) or e (β-strand), furthermore the consensus amino acids are shown by the following symbols: conserved residues, bold and uppercase letters; aliphatic residues, l; aromatic residues, @; hydrophobic residues, h; polar residues, p; tiny residues, t; small residues, s; bulky residues, b; positively charged residues, +; and negatively charged residues, –.

129
Figure IV.2: continued from previous page
Figure IV.3: continued from previous page
Appendix V

List of publications


Møller, M. S., Vester-Christensen, M. B., Jensen, J. M., Abou Hachem, M., Henriksen, A. and Svensson, B.. Structural rational for regulation of barley limit dextrinase activity by the endogenous proteinaceous limit dextrinase inhibitor. Manuscript in preparation, which are included as a chapter in the present thesis. To be submitted to Plant Cell. (Chapter 4)
Appendix VI

Presentations

**Oral presentations** 6th European Symposium of Enzymes in Grain Processing, ESEGP-6, Valby, Denmark. 28\(^{th}\) to 30\(^{th}\) of November 2011. “Barley limit dextrinase and its proteinaceous inhibitor - The complex structure.” (selected from abstract)


**Poster presentations** “Structure, function and protein engineering in starch debranching enzyme systems”. Møller, M. S., Abou Hachem, M., and Svensson, B.. 3\(^{rd}\) Protein.DTU Workshop, Lyngby, Denmark, 16\(^{th}\) March 2010. (Including a 1 min oral presentation)


“Kinetic analysis of a glycoside hydrolase family 13_31 enzyme from the probiotic bacterium *Lactobacillus acidophilus* NCFM reveals broad specificity towards isomaltoooligosaccharides and polymeric dextran”. Møller, M. S., Sørensen, P. G., Abou Hachem, M., and Svensson, B.. 25\(^{th}\) International Carbohydrate Symposium, ICS, Tokyo, Japan, 1\(^{st}\) to 6\(^{th}\) August 2010.

“Isomaltoooligosaccharide catabolism in the probiotic bacterium *Lactobacillus acidophilus* NCFM mediated by 1,6-α-glucosidases from GH13_31”. Møller, M. S., Sørensen, P. G., Abou Hachem, M., and Svensson, B.. 4\(^{th}\) Symposium on the Alpha-Amylase Family, ALAMY_4, Smolenice Castle, Slovakia, 26\(^{th}\) to 30\(^{th}\) September 2010.

“Isomaltoooligosaccharide catabolism in the probiotic bacterium *Lactobacillus acidophilus* NCFM mediated by 1,6-α-glucosidases from GH13_31”. Møller, M. S., Sørensen, P. G., Abou Hachem, M., and Svensson, B.. 4\(^{th}\) Protein.DTU Workshop, Lyngby, Denmark, 12\(^{th}\) November 2010.

“Kinetics of barley limit dextrinase hydrolysis of amylopectin and mutational analysis of a putative -3 subsite residue”. Møller, M. S., Kyasaram, M., Abou Hachem, M., Henriksen, A., and Svensson, B.. 5th Protein.DTU Workshop, Lyngby, Denmark, 9th June 2011. (Including a 1 min presentation)


“Structural and mutational analysis provide a snapshot into the mechanism and energetics of inhibition of starch α-1,6 debranching enzymes”. Møller, M. S., Jensen, J. M., Vester-Christensen, M. B., Hägglund, P., Henriksen, A., Abou Hachem, M. and Svensson, B.. 7th Protein.DTU Workshop, Lyngby, Denmark, 15th May 2012.

Appendix VII

Paper: Enzymology and structure of the GH13.31 glucan 1,6-α-glucosidase that confers isomaltooligosaccharide utilization in the probiotic *Lactobacillus acidophilus* NCFM

Besides the publications which are based on LD and/or LDI experiments performed during my PhD study, I have finalised a manuscript based on my master project. During the time of my PhD study I have done additional bioinformatic analyses and contributed to design of additional experiments. The paper focuses on an enzyme, a glucan 1,6-α-glucosidase, which is a member of the same glycoside hydrolase family as LD, GH13, but belongs to subfamily 31 (GH13.31). It hydrolyses α-1,6 bonds in isomaltooligosaccharides and in the polymeric substrate dextran. The enzyme originates from the probiotic bacterium *Lactobacillus acidophilus* NCFM. The structure of the α-1,6 hydrolysing enzyme is compared with barley LD and the pullulanases in the present thesis (Chapter 3, Section 3.2.6).
Enzymology and Structure of the GH13_31 Glucan 1,6-α-Glucosidase That Confers Isomaltooligosaccharide Utilization in the Probiotic Lactobacillus acidophilus NCFM

Marie S. Möller, Folmer Fredslund, Avishek Majumder, Hiroyuki Nakai, Jens-Christian N. Poulsen, Leila Lo Leggio, Birte Svensson, and Maher Abou Hachem

Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark; and Biophysical Chemistry Group, Department of Chemistry, University of Copenhagen, Copenhagen, Denmark

Isomaltooligosaccharides (IMO) have been suggested as promising prebiotics that stimulate the growth of probiotic bacteria. Genomes of probiotic lactobacilli from the acidophilus group, as represented by Lactobacillus acidophilus NCFM, encode α-1,6-glucosidases of the family GH13_31 (glycoside hydrolase family 13 subfamily 31) that confer degradation of IMO. These genes reside frequently within maltooligosaccharide utilization operons, which include an ATP-binding cassette transporter and α-glucan active enzymes, e.g., maltogenic amylases and maltose phosphorylases, and they also occur separated from any carbohydrate transport or catabolism genes on the genomes of some acidophilus complex members, as in L. acidophilus NCFM. Besides the isolated locus encoding a GH13_31 enzyme, the ABC transporter and another GH13 in the maltooligosaccharide operon were induced in response to IMO or maltotetraose, as determined by reverse transcription-PCR (RT-PCR) transcriptional analysis, suggesting coregulation of α-1,6- and α-1,4-glucosidoligosaccharide utilization loci in L. acidophilus NCFM. The L. acidophilus NCFM GH13_31 (LaGH13_31) was produced recombinantly and shown to be a glucan 1,6-α-glucosidase active on IMO and dextran and product-inhibited by glucose. The catalytic efficiency of LaGH13_31 on dextran and the dextran/panose (trisaccharide) efficiency ratio were the highest reported for this class of enzymes, suggesting higher affinity at distal substrate binding sites. The crystal structure of LaGH13_31 was determined to a resolution of 2.05 Å and revealed additional substrate contacts at the +2 subsite in LaGH13_31 compared to the GH13_31 from Streptococcus mutans (SmGH13_31), providing a possible structural rationale to the relatively high affinity for dextran. A comprehensive phylogenetic and activity motif analysis mapped IMO utilization enzymes from gut microbiota to rationalize preferential utilization of IMO by gut residents.
(Geobacillus thermoglucosidasius), Bacillus coagulans, and Bacillus cereus have been reported, and the structure of the last enzyme has been determined (43, 47, 54, 55).

Lactobacillus acidophilus NCFM is a commercially important probiotic isolated from the human gut and characterized in the 1970s (16). It has since then been widely investigated for its physiologial, biochemical, genetic, and fermentative properties (44). The numerous proteins related to carbohydrate transport and metabolism encoded by L. acidophilus NCFM reflect its capacity to utilize a variety of mono-, di-, oligo-, and polysaccharides (3). Notably, panose was shown to sustain the growth of L. acidophilus NCFM in vitro (32), in line with the suggested prebiotic effect of IMO.

In the present study, the catalobism of IMO is examined in L. acidophilus NCFM as a model organism for probiotic lactobacilli from the gut niche. Genome analysis confirmed the presence of a putative GH13_31 G16G-encoding gene that likely confers IMO hydrolysis. This gene was heterologously expressed in Escherichia coli, and the recombinant enzyme, designated L. acidophilus GH13_31 (LaGH13_31), was biochemically and structurally characterized. Furthermore, an analysis was performed to map IMO utilization loci in probiotic lactobacilli, highlighting commonalities and differences in the organization of IMO utilization genes on the genomes of this group of organisms important for human health.

**MATERIALS AND METHODS**

High-purity chemicals and commercial enzymes were from Sigma-Aldrich, St. Louis, MO, unless otherwise stated. The commercial IMO mix was from Wako Pure Chemical Industries, Osaka, Japan.

**Bioinformatic analysis.** For the sequence alignment, the first 100 protein sequences from BLAST searches using LBA0264 and LBA1872 were retrieved and complemented with lactobacillus sequences from CAZy (8) and the genome database provided by the National Center for Biotechnology Information (NCBI). The organization of 16G- and G16G-encoding genes from various organisms was investigated using the genome database of the National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi).

**Purification of IMO and semiquantitative RT-PCR.** In order to verify the functionality of the genes predicted to be involved in IMO utilization, L. acidophilus NCFM was grown on purified IMO from a commercial mix first treated with rice α-glucosidase to hydrolyze maltooligosaccharides into glucose. The hydrolysate was ultrafiltered (10-kDa Amicon filter; Millipore, Billerica, MA), desalted (Amberlite MB-20; Fluka, Sigma-Aldrich), and filtered (0.45-μm pore size; Frisénette Aps, Kopenhagen, Denmark). Glucose was removed by a high-performance liquid chromatograph ([HPLC] Ultimate 3000; Dionex, Sunnyvale, CA) equipped with a refractive index detector (RI-101; Showa Denko, Kanagawa, Japan) using a TSKgel Amide-80 column (5-μm particle size; 26 by 250 mm with 4.6- by 10 mm guard column [Tosoh, Tokyo, Japan]) at a constant flow rate 1 ml/min of mobile phase (acetoni-trile-water, 70:30 [vol/vol]) at 70°C. The purity was confirmed based on peaks detected with high-performance anion exchange chromatography with peramperometric detection ([HPAEC-PAD] ICS-3000; Dionex) on a CarboPac PA200 anion exchange column (3 by 250 mm and 3- by 50-mm guard column; Dionex) eluted by a linear 0 to 125 mM sodium acetate gradient in 100 mM NaOH (for 35 min at 25°C; flow rate, 0.35 ml/min). The IMO components were identified based on standards: IG2, isomaltotriose (IG3), isomaltotetraose (IG4), isomaltohexaose (IG6), panose, glucose, and maltooligosaccharides from maltose through to maltoheptaose. The purified IMO were essentially glucose free and contained IMO having degrees of polymerization of 2 to 4 (DP 2 to 4).

L. acidophilus NCFM was grown (in duplicate) with 1% of either glucose, maltotetraose, or IMO at DP 2 to 4 (purified as above), under aerobic conditions in a 160-ml baffled culture vessel with a 5°C CTAAGGCTTCATAGTTGGAAAAATGCTGTTG-3' and reverse primer 5'-CCGCTCGAGTCACTTTTGTATTAAAGCCGC-3'. The PCR amplicon (1.632 bp), flanked by Ncol and Xbol restriction sites (in bold), was cloned into pET21a (+) (Novagen, Darmstadt, Germany) and transformed into E. coli TOP10 (Invitrogen, Carlsbad, CA) by heat shock. Transformants were selected on LB-agar plates with 100 μg ml−1 ampicillin, and positive transformants harboring pET21a (+)–LaGH13_31 were verified by restriction analysis and full sequencing. E. coli BL21(DE3) cells (Invitrogen) transformed with pET21a (+)–LaGH13_31 were used to produce the enzyme.

**Production and purification.** LaGH13_31 was produced in a 5-liter bioreactor (Biostat B; B. Braun Biotech International, Melsungen, Germany) according to a fed-batch protocol developed previously for production of other L. acidophilus NCFM recombinant enzymes (14), with the exception that the induction (OD600 of 8.3) was carried out at 16°C with 40 μM isopropyl-β-D-thiogalactopyranoside (IPTG). The fermentation was terminated after 23 h of induction, and 61 g of cell pellet (harvested by centrifugation at 12,200 × g for 10 min at 4°C) was resuspended in 60 ml of a buffer containing 10 mM HEPES, 10 mM imidazole, 10% glycerol, 0.5 M NaCl, and 2 mM CaCl2 (pH 7.5) (buffer A) and disrupted by passage through a French press at 600 × 107 Pa. After Benzonase nuclease (Novagen) treatment, the suspension was centrifuged (at 43,000 × g for 65 min) and sterile filtered (0.22-μm pore size). LaGH13_31 was purified by immobilized metal ion affinity chromatography using a 5-ml HisTrap HP column (GE Healthcare, Uppsala, Sweden) as described elsewhere (14). This purification step was followed by anion exchange chromatography using an 8-ml Mono Q 10/100 GL column (GE Healthcare) equilibrated in 10 mM HEPES, pH 7.0, and 2 mM CaCl2 (GE Healthcare) and installed on an AKTAexplorer chromatograph (GE Healthcare). The LaGH13_31-loaded column was stringently washed (60 ml/h; 0.25 M NaCl, 12 columns volumes [CV]) at the same flow rate using a linear gradient (0.25 to 0.29 M NaCl, 15 CV; 0.29 to 0.5 M NaCl, 2 CV). The fractions containing activity were analyzed by SDS-PAGE, pooled, concentrated (10-kDa Amicon filter; Millipore), and buffer exchanged to 20 mM 2-(N-morpholino)ethanesulfonic acid (MES)-NaOH, pH 6.5, 2 mM CaCl2, and 100 mM NaCl. The protein concentration was determined spectrophotometrically using the molar extinction coefficient ε280 of 138,180 M−1 cm−1 as determined by amino acid analysis (4). The isoelectric point (pI) was determined by focusing on an PhastGel IEF 4 to 6.5 isoelectric focusing gel and a pl marker (2.8 to 6.5) (GE Healthcare) using the PhastSystem (Pharmacia, Uppsala, Sweden).

**Enzyme activity: standard enzyme assay.** A standard assay (50 μl) was performed in 60 mM MES-NaOH, pH 6.0, 2 mM CaCl2, and 0.0005%...
bovine serum albumin (BSA) using 2 mM p-nitrophenyl α-D-glucopyranoside (PNPG) as the substrate and enzyme (2.5 to 6 mM) for 10 min at 37°C. The reaction was stopped by addition of 1 M Na2CO3 (200 μl), and the A405 was measured. The concentration of liberated p-nitrophenol (PNP) was calculated from a PNP standard curve. One unit of enzyme activity (U) was defined as the amount of enzyme required to liberate 1 μmol of PNP per min under the assay conditions.

Optimum pH. The pH optimum of LaGH13_31 (2.8 mM, 200 μl) was determined using the standard assay in 40 mM Britton-Robinson universal buffers pH 2 to 10 (7) and stopped by addition of 1 M Na2CO3 (800 μl).

Temperature optimum and stability. The temperature optimum of LaGH13_31 activity was determined by performing the standard assay at pH 6.0 to 21°C. The rate of irreversible thermal inactivation was also determined at 37°C and 50°C by incubation of the enzyme (28.3 mM) and measurement of residual activity at eight time points (20-μl aliquots, immediately cooled on ice) using the standard assay. The inactivation rate constant k (min−1) was calculated assuming first-order kinetics from the slope of ln(Ai/Ai0) plotted against time (min), where A0 is the activity at time t (min) and A0 is the initial activity; the half-life (t1/2) was calculated from the equation t1/2 = ln(2)/k.

Determination of kinetic parameters of IMO hydrolysis. Reaction mixtures (300 μl) containing LaGH13_31 (1.4 to 9.8 mM) and PNPG (0.05 to 10 mM), IG2 (2.5 to 80 mM), IG3 (isomaltotriose, 1 to 25 mM), IG4 (isomaltotetraose, 2.5 to 30 mM), panose (0.5 to 14 mM), or dextran (1.6 to 40 mg/ml) were used to determine the initial hydrolysis rates by transferring aliquots (50 μl) into 1 M Na2CO3 (200 μl for PNPG) or 2 M Tris-HCl, pH 7.0 (100 μl for other substrates), at 4-min intervals for 16 min. Liberated PNP was measured as described above, whereas glucose liberated from other substrates was quantified by a modified glucose oxidase-peroxidase method (GOPOD) (Megazyme, Bray, Ireland), as follows. First, 200 μl of the GOPOD reagent was preheated for 5 min at 40°C, then 100 μl of the stopped reaction sample was added. After 20 min at 40°C the A405 was measured. Inhibition kinetics of LaGH13_31 by 6 or 8 mM glucose was probed using PNPG as the substrate as described above. The Michaelis-Menten model was fit to the initial rates to determine kcat and Km using Sigma plot, version 9.01 (SYSTAT Software, Inc., Richmond, CA) and also applied to determine the Ki for glucose inhibition.

Crystallization. The purified protein was concentrated to 16 mg/ml in 20 mM MES-NaOH buffer, pH 6.5, 100 mM NaCl, and 2 mM CaCl2 as described above. Initial crystallization conditions were obtained by screening using an Oryx 8 Protein Crystallization Robot (Douglas Instruments, Ltd., United Kingdom) with 96-well trays at room temperature. A JCSG+ screen (Qiagen) was set up in sitting drops consisting of 1:1 (total, 200 nl) and 1:2 (total, 300 nl) protein/reservoir solutions, respectively. Small thin needles were obtained with a reservoir containing 20% glycerol, 16% polyethylene glycol 8000 (PEG 8000), and 0.1 M MES, pH 6.5, and were reproduced in 24-well VDX trays (Hampton Research) in sitting drops consisting of 2 μl of protein and 1 μl of reservoir solution. No extra cryoprotection was used before the crystals were mounted.

Data collection, processing, and refinement. A native data set was collected to 2.05 Å resolution at the I19-1-2 side station of the Cassiopaea beamline MAX-lab, Lund, Sweden. The space group was determined as P212121 with the following cell dimensions: a = 55.8 Å, b = 107.3 Å, and c = 6.6 Å. Processing and scaling of the data were performed with XDS and XSSCALE (21) (data shown in Table 1). Molecular replacement with MOLREP (31) using SmGH13_31 (Protein Data Bank identification [PDB ID], 2ZJC) as a search model yielded a clear solution with one monomer in the asymmetric unit. The model was rebuilt using phenix autobuild (30) and COOT (13) and refined with phenix.refine (2) to an Rwork/Rfree of 0.151/0.196. The final model includes LaGH13_31 residues 2 to 538 (using the native protein numbering) in addition to eight glycerol molecules, three MES molecules (for MES residue 1547 in chain A, only the sulfonic acid moiety is modeled), 475 water molecules, and one calcium ion (Table 1). No electron density was observed for three N-terminal residues and the C-terminal His tag (eight residues). No Ramachandran outliers were observed. A double conformation was modeled for Asn7, and Lys247 was truncated at the C-terminal side chain collided with its symmetry mate. Protein coordinates were represented with PyMOL, version 1.4.1 (Schrodinger, LLC).

Protein structure accession number. Atomic coordinates of LaGH13_31 have been deposited at the Protein Data Bank under accession code 4aie.

RESULTS

Bioinformatic analysis. Only a single gene (LBA0264) in the genome of L. acidophilus NCFM was annotated to encode a GH13_31 in CAZY. The amino acid sequence of LBA0264 showed highest identity to glucan 1,6-α glucosidases (G16G) from streptococci (56 to 61%), followed by O16G from bacilli (49 to 53%), whereas clearly lower identities were shared with trehalose-6-phosphate hydrolases from Bacillus subtilis and E. coli, all of which are assigned into GH13 (Table 2). Another GH13 gene, LBA1872 (AAV43672.1), located within the maltotriosaccharide gene cluster (34), was not assigned into subfamily 31 but displayed lower primary structure identities of 30 to 37% to other characterized O16G and other α glucosidases of GH13 (see Table S2 in the supplemental material). The identity between LBA0264 and LBA1872 is 35.4%. A multiple sequence alignment of top hits from BLASTP with LBA0264 and LBA1872 combined with SmGH13_31, characterized O16G, and enzymes from lactobacilli assigned to GH13_31 showed that LBA0264 aligns very well with SmGH13_31 in the conserved regions, which define pivotal ac-

### Table 1: Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value for the parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.038</td>
</tr>
<tr>
<td>High resolution (Å)</td>
<td>2.05</td>
</tr>
<tr>
<td>Space group</td>
<td>P21212</td>
</tr>
<tr>
<td>Unit cell parameter (Å)</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>55.8</td>
</tr>
<tr>
<td>b</td>
<td>107.3</td>
</tr>
<tr>
<td>c</td>
<td>6.6</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>39,496</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30-2.05 (2.10-2.05)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.2 (91.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.3 (4.7)</td>
</tr>
<tr>
<td>Mean fσ(I)</td>
<td>25.35 (9.7)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>5.4 (20.8)</td>
</tr>
<tr>
<td>Rpим (%)</td>
<td>5.9 (25.1)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>4,431</td>
</tr>
<tr>
<td>No. of hetero atoms</td>
<td>77</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>475</td>
</tr>
<tr>
<td>R-factor (%)</td>
<td>13.7</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>18.5</td>
</tr>
<tr>
<td>RMSD value from identity</td>
<td>0.010</td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>1.24</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
</tr>
<tr>
<td>Allowed</td>
<td>100</td>
</tr>
<tr>
<td>Favored</td>
<td>97.2</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* Data were collected using MAX-lab beamline 1912. Values in parentheses are for the outer-resolution shell. RMSD, root mean square deviation.

* Calculated using MolProbity (9).
tive-site residues and region V (Fig. 1), which is the α-1,6 specificity motif in the α-amylase family (clan GH-H according to the CAZy classification) (38). In addition, LBA0264, like SmGH13_31, has the shorter loop between conserved regions II and III, which defines the G16G subspecificity in GH13_31 (43). In contrast, LBA1872 lacked the α-1,6 specificity motif, suggesting that it might have another specificity. In the unrooted phylogenetic tree (Fig. 2) of the enzymes included in the multiple sequence alignment, LBA0264 clusters together with SmGH13_31 and several putative G16Gs from lactobacilli, while LBA1872 segregates in a different cluster together with uncharacterized proteins from a wide range of lactobacilli (see Table S3 in the supplemental material for details). The characterized O16Gs from bacilli form a third cluster together with mainly uncharacterized sequences from bacilli and other Gram-positive bacteria, including a few Lactobacillus sequences, and this cluster is more closely related to the G16G group (Fig. 2). Thus, the analysis supports the annotation of LBA0264 as an IMO active enzyme, whereas LBA1872 seemed not to be involved in IMO hydrolysis based on this analysis.

Transcriptional analysis of genes possibly involved in IMO utilization. Both IMO and maltooligosaccharides sustained the growth of L. acidophilus NCFM, albeit to a lower cell density than glucose. The semiquantitative RT-PCR analysis showed the transcription of LBA0264 to be upregulated on both maltotetraose and the purified IMO mix compared to cells grown on glucose (Fig. 3). Similarly, LBA1872 and the permease (LBA1866) as well as the solute binding protein (LBA1867) components of the ATP-binding cassette (ABC) transport system in the maltooligosaccharide gene cluster were also upregulated on both maltotetraose and IMO.

Production and basic characterization of LBA0264, LaGH13_31. LaGH13_31 was purified to electrophoretic homogeneity.

---

### Table 2: Amino acid sequence comparison of the gene product of LBA0264 to functionally characterized GH13 enzymes

<table>
<thead>
<tr>
<th>Enzyme and organism(s)</th>
<th>UniProtKB/Swiss-Prot accession no.</th>
<th>Identity (%)</th>
<th>Similarity (%)</th>
<th>Gap</th>
<th>Score</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucan-1,6-α-glucosidase Streptococcus mutans</td>
<td>Q99040&lt;sup&gt;b&lt;/sup&gt;</td>
<td>57.6</td>
<td>75.0</td>
<td>2.0</td>
<td>1,697</td>
<td>19, 42, 43</td>
</tr>
<tr>
<td>Oligo-1,6-α-glucosidase Bacillus cereus</td>
<td>P21332&lt;sup&gt;d&lt;/sup&gt;</td>
<td>53.1</td>
<td>70.1</td>
<td>4.5</td>
<td>1,585</td>
<td>48, 54, 56</td>
</tr>
<tr>
<td>Bacillus thermoglucosidasus (Geobacillus thermoglucosidas)</td>
<td>P23094&lt;sup&gt;c&lt;/sup&gt;</td>
<td>51.8</td>
<td>69.3</td>
<td>4.8</td>
<td>1,575</td>
<td>55</td>
</tr>
<tr>
<td>Bacillus coagulans</td>
<td>Q45101&lt;sup&gt;d&lt;/sup&gt;</td>
<td>51.5</td>
<td>66.5</td>
<td>6.4</td>
<td>1,492</td>
<td>47</td>
</tr>
<tr>
<td>Trehalose-6-phosphate hydrolase Bacillus subtilis</td>
<td>P39795&lt;sup&gt;b&lt;/sup&gt;</td>
<td>46.1</td>
<td>63.9</td>
<td>4.3</td>
<td>1,387</td>
<td>18</td>
</tr>
<tr>
<td>Escherichia coli K-12</td>
<td>P28894&lt;sup&gt;c&lt;/sup&gt;</td>
<td>45.8</td>
<td>63.4</td>
<td>6.4</td>
<td>1,307</td>
<td>41</td>
</tr>
</tbody>
</table>

<sup>a</sup> The similarity of amino acid sequences was investigated using the BLASTP program (UniProtKB/Swiss-Prot database). The proteins are listed according to descending percentages of sequence identity and grouped according to specificity.

<sup>b</sup> GH13 subfamily 31.

<sup>c</sup> GH13, no subfamily.

<sup>d</sup> GH13 subfamily 29.
FIG 2 Phylogenetic tree constructed based on the multiple sequence alignment partly shown in Fig. 1 (full alignment is shown in Fig. S1 in the supplemental material). The tree depicts the clustering of three main groups of GH13 enzymes: (i) a glucan 1,6-α-glucosidase (G16G) cluster from mainly acidophilus complex lactobacilli including LgGH13-3, streptococci, enterococci, and others (see Table S3); (ii) an oligo-1,6-α-glucosidase (O16G) cluster containing characterized enzymes from bacilli, together with a variety of uncharacterized sequences from other Gram-positive bacteria including a branch of non-acidophilus complex lactobacilli; (iii) a cluster of uncharacterized sequences homologous to LBA1872 that occurs in the maltooligosaccharide operon in L. acidophilus NCFM and forms a distinct group (Unknown specificity 1), supportive of the lack of α-1,6 sequence motifs and suggestive of a different function. A small group of uncharacterized sequences form a fourth intermediate group (Unknown specificity 2), whereas a single sequence resembling possibly an ancestral O16G segregates alone, likely due to its taxonomic distance to other O16G sequences in the tree. The following Lactobacillus species are represented: L. amylovorans, L. animalis, L. brevis, L. delbrueckii, L. farcininis, L. jensenii, L. helveticus, L. iners, L. pentosus, L. plantarum, L. rhamnosus, L. ruminis, and L. sakei subsp. sakei. Other species are as follows: B. halodurans, Bacillus halodurans; B. sp. 2_A.57_CT2, Bacillus sp. strain 2_A.57_CT2; B. sp. NRRL B-14911, Bacillus sp. strain NRRL B-14911; B. fuscom, Brevibacterium fuscom; L. garvieae, Lactococcus garvieae; C. sp. AT7, Carnobacterium sp. strain AT7; C. sp. 17-4, Carnobacterium sp. strain 17-4; C. sp. DL-VIII, Clostridium sp. strain DL-VIII; C. leptum, Clostridium leptum; C. owensensis, Caldicultiluocapidium owensensis; D. formigenes, Dorea formigenes; E. faecalis, Enterococcus faecalis; E. faecium, Enterococcus faecium; E. italicus, Enterococcus italicus; E. subilacun, Eubacterium subilacun; E. prasunitiszi, Facalidisebacberium prasunitiszi; H. hydrogeniformans, Halanaerobium hydrogeniformans; L. grayi, Listeria grayi; L. pseudomesenteroides, Leuconostoc pseudomesenteroides; O. sp. TW25, Ornithinibacillus sp. strain TW25; P. larvarum, Paenibacillus larvarum; R. sp. 5_1.39B, FAA; R. lactar, Ruminococcus lactar; R. turus, Ruminococcus torus; S. mitis, Streptococcus mitis; T. halophilus, Tetrodosebacberium halophilus; T. thermosaccharolyticum, Thermoanaerobacterium thermosaccharolyticum. Other species are as identified in the text.
geneity, and its pI was determined to be 4.8 in accordance with the theoretically calculated value of 4.9. The yield of LaGH13_31 after two purification steps was 25 mg/liter of culture with a specific activity of 335 U/mg. The pH optimum for LaGH13_31 was determined to be 5.5, and the enzyme retained more than 50% of its maximum activity at pH 4 to 8 (see Fig. S2A in the supplemental material). The temperature optimum at pH 6 was 39°C (see Fig. S2B), and the inactivation rate constants at 37°C and 50°C were 0.001 min⁻¹ and 0.0152 min⁻¹, respectively, corresponding to a $t_{1/2}$ of 4.8 days and 46 min, respectively.

**Substrate preference.** The substrate specificity of LaGH13_31 was investigated by measuring initial reaction rates for eight substrates (1 mM) (Table 3) including IMO (DP 2 to 4), panose that has an α-1,6 linkage, sucrose, and disaccharides containing α-1,1- or α-1,4-linked glucose. Among the natural substrates, the highest rate of hydrolysis was on panose (Table 3), and LaGH13_31 clearly preferred IMO longer than IG2, while having negligible activity on other natural substrates, confirming its function as a glucan 1,6-α-glucosidase.

Kinetic analysis performed on IMO and dextran confirmed this trend and showed higher catalytic efficiencies ($k_{cat}/K_m$) of LaGH13_31 on IG3 and IG4 than isomaltose (Table 4). Although saturation was not reached at the highest dextran concentration used (40 mg/ml), clear curvature of the Michaelis-Menten plot (see Fig. S3 in the supplemental material) allowed unambiguous determination of the kinetic parameters.

Effective inhibition by the product glucose indicates possible feedback regulation of the activity of LaGH13_31. The Michaelis-Menten plots of PNPG hydrolysis showed significant inhibition by 6 or 8 mM glucose (see Fig. S4 in the supplemental material). The competitive inhibition model gave the best fits to the data ($R^2 = 0.999$), and the $K_i$ of glucose was determined to 4.0 ± 0.18 mM.

**Three-dimensional structure of LaGH13_31.** Obtained crystals of LaGH13_31 belonged to space group $P2_12_12$ with one monomer in the asymmetric unit. The crystals diffracted to 2.05-Å resolution, and the structure was solved by molecular replacement using the structure of SmGH13_31 (PDB Z2IC). The overall structure of LaGH13_31 (Fig. 4) shows the classical GH13 architecture of a catalytic (β/α)$_5$-barrel fold domain (domain A) and a C-terminal antiparallel β-sheet domain (domain C). The three catalytic residues conserved in GH13 are situated at the ends of β-strands 4 (Asp198, catalytic nuclease), 5 (Glu240, catalytic acid-base), and 7 (Asp316, transition state stabilizer). Functionally important amino acid residues at the active-site substrate binding subsites −1 through +2 are depicted in Fig. 5A, following the accepted subsite nomenclature, with the minus subsites at the nonreducing end of the substrate and plus subsites at the reducing end and hydrolysis occurring between the −1 and +1 subsites (10). Domain A, comprising the first 465 residues of the protein; LBA1867, multiple sugar binding ABC transporter system; LBA1872, GH13 of unknown function; LBA2071, 16S rRNA gene.

### Table 3 Normalized reaction rate of LaGH13_31 on various substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Reaction rate (s⁻¹)</th>
<th>Relative rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNPG</td>
<td>164.0 ± 5.1</td>
<td>94.3</td>
</tr>
<tr>
<td>IG2</td>
<td>26.6 ± 1.3</td>
<td>15.3</td>
</tr>
<tr>
<td>IG3</td>
<td>41.2 ± 0.9</td>
<td>23.7</td>
</tr>
<tr>
<td>IG4</td>
<td>41.0 ± 0.2</td>
<td>23.6</td>
</tr>
<tr>
<td>Trehalose</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Maltohe</td>
<td>(4.9 ± 0.2) × 10⁻²</td>
<td>0.03</td>
</tr>
<tr>
<td>Sucrose</td>
<td>(10.5 ± 1.5) × 10⁻²</td>
<td>0.06</td>
</tr>
<tr>
<td>Panose</td>
<td>173.9 ± 2.4</td>
<td>100</td>
</tr>
</tbody>
</table>

*Substrates were used at a concentration of 1 mM.*

*The reaction rate was calculated as $V/[E]$, where $V$ is the initial velocity and [E] is the enzyme concentration.*

*Relative to the reaction rate toward panose.*

*ND, not detected.*

### Table 4 Kinetic parameters of LaGH13_31 at 37°C and pH 6.0

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}$ (s⁻¹)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$ (s⁻¹ mM⁻¹)</th>
<th>Relative $k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG2</td>
<td>517 ± 6.3</td>
<td>22.3 ± 0.7</td>
<td>23</td>
<td>14.6</td>
</tr>
<tr>
<td>IG3</td>
<td>475 ± 12.7</td>
<td>14.1 ± 0.8</td>
<td>34</td>
<td>21.7</td>
</tr>
<tr>
<td>IG4</td>
<td>501 ± 12.7</td>
<td>16.1 ± 0.9</td>
<td>31</td>
<td>19.7</td>
</tr>
<tr>
<td>PNPG</td>
<td>597 ± 14.3</td>
<td>2.38 ± 0.16</td>
<td>251</td>
<td>159.9</td>
</tr>
<tr>
<td>Panose</td>
<td>612 ± 12.8</td>
<td>3.9 ± 0.2</td>
<td>157</td>
<td>100</td>
</tr>
<tr>
<td>Dextran</td>
<td>384 ± 17.1</td>
<td>29.2 ± 2.4</td>
<td>13.6</td>
<td>8.7</td>
</tr>
</tbody>
</table>

*Normalized to panose.*

*mg ml⁻¹.*

*s⁻¹ mg⁻¹ ml.*
LaGH13_31, contains two inserts after the fourth and eighth \(\beta\)-strands. The first insert is referred to as domain B (residues 101 to 169) and provides residues responsible for substrate recognition. Domain B’ (residues 359 to 444) is in orange, and domain C (residues 466 to 538) is in red. The catalytic residues (Asp198 and Glu240) are shown as black sticks, and three glycerol molecules in the active site are shown with white carbons. An IG3 molecule from SmGH13_31 (2ZID) is superimposed and shown with green carbons at 60% transparency. A calcium ion (green sphere) is coordinated by a water molecule (red sphere) and residues 20, 22, 24, 26, and 28, and a close-up view of the Ca\(^{2+}\) binding site is shown for clarity.

In addition to structurally ordered water molecules, the electron density revealed the presence of three MES (only the sulfonic acid moiety of MES A 1547 is modeled) and eight glycerol molecules. One of the MES molecules was bound by Tyr150 and Tyr157, also present in SmGH13_31 (Tyr146 and Tyr153) and BcO16G (Tyr151 and Tyr158). Three of the glycerol molecules were found in the active site, and comparison with the SmGH13_31 in complex with isomaltotriose revealed that glycerol 1542 together with waters 2215, 2249, and 2334 are ligands to the same residues as observed at the \(-1\) subsite in SmGH13_31, thus demonstrating the conservation of this pivotal subsite (Fig. 5).

DISCUSSION

The administration of IMO has been reported to result in an increase in bifidobacteria and lactobacilli in humans and rats (23,
This was recently corroborated by a human intervention study showing a significant increase of bifidobacteria and lactobacilli, together with a 3-fold decrease in clostridia, due to IMO administration in elderly constipated humans, highlighting the selectivity of IMO (58). Surprisingly, the genetics and enzymology of IMO utilization remain unexplored in the gut niche. The present study investigates the genetic basis and the enzymology of IMO utilization in L. acidophilus NCFM, which serves as a model for probiotic lactobacilli.

Bioinformatic analysis. A single gene encoding a GH13_31 enzyme (LBA0264, designated LaGH13_31) was found in the L. acidophilus NCFM genome. In addition, a putative enzyme, encoded by the locus LBA1872, displayed low similarity to O16Gs. LaGH13_31 is homologous to the characterized SmGH13_31 (Table 2), as evident from the multiple sequence alignment (Fig. 1) and the phylogenetic tree that depicts the clustering of LaGH13_31 and closely related enzymes mainly from the acidophilus complex lactobacilli together with the G16G from S. mutans (Fig. 2; see also Table S3 in the supplemental material). Characterized GH13_31 O16G from different bacilli formed an adjacent cluster (Fig. 2). Interestingly, although this O16G cluster contained sequences from lactobacilli, none of these belonged to the acidophilus complex (see Table S3), suggesting that the gut niche adaptation has driven the enrichment of the G16G specificity in acidophilus complex Lactobacillus members and/or that shorter IMO are catabolized using different enzymes in acidophilus complex members. Furthermore, LaGH13_31 contained both the QPDLN motif in the conserved region V of the α-amylase family, reported to define the O16G subfamily containing both O16G and G16G (38), and the shorter β-α loop 4 that distinguishes G16G from O16G (43). Taken together, this strongly suggested that LaGH13_31 is a G16G. In contrast, LBA1872 was clearly distinguishable from characterized enzymes with O16G, and its amino acid sequence resembled the intermediate group having a region V MPKLN motif and a conserved histidine in region II (Fig. 1) (38). More importantly, the valine residue following the catalytic nucleophile, which was experimentally identified as a key signature of α-1,6 hydrolytic activity (57), is instead an alanine (shown in boldface) in LBA1872 (GLRLDA;region II). This, together with the segregation of LBA1872 with other homologues from mainly the Lactobacillus genus in a separate cluster (Fig. 2, Unknown specificity 1), is indicative of a different specificity of this group of enzymes tentatively annotated as O16Gs in the NCBI database based exclusively on in silico predictions.

Genetics of IMO utilization in L. acidophilus and other gut bacteria. The organization of genes mediating utilization of FOS and raffinose oligosaccharides in functional operons comprising transport systems, hydrolases, and transcriptional regulators has been reported in L. acidophilus NCFM (5, 6). Interestingly, LBA1872 is located within the maltooligosaccharide utilization operon, which also encodes a complete ABC transport system annotated as a maltose/maltooligosaccharide transporter, a LacI transcriptional regulator, a GH65 maltose phosphorylase (EC 2.4.1.8), and a GH13_20 maltogenic α-amylase (EC 3.2.1.133). GH13_20 enzymes also possess neopullulanase activity (EC 3.2.1.135) and are typically annotated as neopullulanases (34) (Fig. 6A). This organization is also observed in other L. acidophilus strains (e.g., L. acidophilus ATCC 4796) and in closely related acidophilus complex species, e.g., Lactobacillus amylovorus GRL1118 and Lactobacillus crispatus ST1 (Fig. 6B and C). Another similarity between L. acidophilus NCFM and the aforementioned organisms is that genes encoding GH13_31 enzymes which confer IMO hydrolysis are not in proximity to carbohydrate hydrolysis or transport loci (Fig. 6A to C). In contrast, the GH13_31-encoding genes are located in the maltooligosaccharide operon in other species of the acidophilus complex represented by Lactobacillus johnsonii ATCC 33200 and Lactobacillus gasseri JV-V03 (Fig. 6D and E) as well as in other Lactobacillus species, e.g., Lactobacillus casei BL23 (Fig. 6F), and in S. mutans (data not shown), suggesting that the relocation of the GH13_31 genes in L. acidophilus NCFM and closely related organisms is a recent evolutionary event. With respect to IMO transport, the presence of GH13_31-encoding genes in maltooligosaccharide operons in several lactobacilli, including many acidophilus complex members, suggests that IMO may be internalized by the ABC transporters present in these operons. It is unclear if the same is valid for the group represented by L. acidophilus NCFM, where the GH13_31 resides on a separate locus. It cannot be ruled out, however, that a shorter IMO, e.g., IG2, occurs via other types of transporters, e.g., phosphoenolpyruvate-dependent phosphotransferase system (PTS), as demonstrated in the case of maltose/maltooligosaccharide utilization in other Gram-positive bacteria, where maltose is internalized via a PTS transporter and where longer maltooligosaccharides are transported via an ABC transporter (1, 45). Phosphorylated maltose and maltotriose internalized via a PTS are recognized by a specific GH4 6-phospho-α-glucosidase (EC 3.2.1.122) (45). Noticeably, GH4-encoding genes occur frequently in acidophilus complex lactobacilli including L. acidophilus NCFM (LBA1689). Bifidobacteria encode GH13 α-glucosidases that are distinctly related to GH13_31 but resemble GH13_31 O16Gs with respect to activity on shorter IMO (39, 52) and have the important Val residue following the nucleophile, consistent with α-1,6 activity. Genes encoding GH13_31 enzymes are also present in enterococci, e.g., Enterococcus faecalis OGI1F (Fig. 6G). On the other hand, commensals from the genus Bacteroides lack GH13_31 enzymes. Only three gut-adapted Clostridia difficile strains possess putative GH13_31 enzymes. The specificity of these putative enzymes toward IMO has not been demonstrated, and a closer analysis showed that they lack the valine residue presented above as an important α-1,6 specificity signature. This is consistent with the observed selectivity of IMO on human gut microbiota, manifested in a 3-fold reduction in clostridia counts following IMO intake (58).

The present transcriptional analysis suggests that IMO induces the expression of LBA0264 and the ABC transport system as well as LBA1872 in the maltooligosaccharide gene cluster. Remarkably, the same loci are also induced by maltotetraose, suggesting coregulation of genes associated with utilization of α-1,4- and α-1,6-glucosidocholigosaccharides. The inducer of the maltose/maltoigosaccharide operon in Streptococcus pneumoniae (belongs to the same order as lactobacilli) has been identified as maltose (36). This disaccharide is the product of panose (abundant in commercial IMO preparations) hydrolysis by LaGH13_31-like enzymes or is the product of maltooligosaccharide hydrolysis by GH13_20 maltogenic α-amylases (EC 3.2.1.133) encoded within the maltooligosaccharide operon discussed above (Fig. 6A). Thus, maltose is a common degradation product of the panose fraction of IMO and maltooligosaccharides, which provides a possible rationale for the coregulation and the colocalization of hydrodrases of these two pathways in some acidophilus complex members.
Activity and structure of LaGH13_31. The activity profile of LaGH13_31 confirmed the bioinformatic analysis predicting its G16G specificity. Interestingly, this enzyme displays a 43-fold higher dextran/panose catalytic efficiency ratio (Table 4) than SmGH13_31 (43). This suggests higher affinity in the distal aglycone substrate subsites of the active-site cleft of LaGH13_31. A structural alignment of LaGH13_31 and SmGH13_31 indicates similar binding modes at the /H11001 and /H11002 subsites but large differences at subsite /H11001, where the residues Arg212, Glu213, and Asn243 in LaGH13_31 correspond to Val208, Ser209, and Gly239, respectively, in SmGH13_31 (43). Superposing the two structures (using superpose in COOT [13]) shows that Arg212 NH1 is 2.9 Å from Glc-879 O3 (glucosyl moiety at subsite /H11001) and that Asn243 OD1 is 2.8 Å from Glc-879 O4 at the /H11001 subsite (Fig. 5), thus supporting a larger number of interactions at subsite /H11001 in LaGH13_31 relative to SmGH13_31. It is intriguing that additional interactions at subsite +2 would have this profound effect on the activity on dextran. One hypothesis is that these additional
contacts at subsite +2 result in better anchoring of the substrate polysaccharide chain, which can act in concert with weak surface binding sites to elicit the observed difference. This synergy between the active site and surface binding sites has been reported in other GH13 enzymes (35) but remains currently unidentified in GH13_31. The conservation of some aromatic residues on the surface of GH13_31 enzymes, e.g., Tyr150 and Tyr157 in LaGH13_31, merits further studies to assess their role in activity on polymeric substrates.

Product inhibition by glucose seems to be relevant for the regulation of LaGH13_31 activity as the measured $K_M$ for glucose (4 mM) is in the same range or lower than the $K_M$ for both natural and synthetic substrates (Table 4). This inhibition has not been shown for GH13_31 enzymes displaying G16G specificity, but it was reported for the O16G from *B. coagulans*, albeit slightly less effectively ($K/K_M$ ratio of 4.7 compared to 1.7 for LaGH13_31) (47).

**Conclusions.** Genomes of probiotic lactobacilli from the acidocellulase complex continuously encode glucan 1,6-$\alpha$-glucosidase genes located either in a maltoligosaccharide operon encompassing an ABC transport system and maltoligosaccharide active enzymes, as in *L. johnsonii* and *L. gasseri* members, or located on separate loci, as in *L. acidophilus* NCFM and closely related species and strains. Similar transcriptional regulation of the GH13_31-encoding gene and genes in the maltoligosaccharide operon including the ABC transporter suggests that the pathways for $\alpha$-1,4 and $\alpha$-1,6 glucan catabolism are linked, possibly through maltose, which is a common hydrolysis product. Bioinformatic analysis of one of the conserved GH13 $\alpha$-glucan operon genes (LBA1872 in *L. acidophilus* NCFM) showed that this gene is most probably mis-annotated as an O16G as it lacks motif signatures shown to be pivotal for $\alpha$-1,6 activity and as it segregates in a separate cluster from canonical GH13_31 IMO active enzymes. In contrast, the recombinant LaGH13_31 was shown to be an active glucan 1,6-$\alpha$-glucosidase catalyzing hydrolysis of longer IMO and product-inhibited by glucose. The crystallographic structure of LaGH13_31 shows a conserved subsite +1 but several additional substrate contacts at subsite +2 compared to related enzymes, which may explain the high affinity of this enzyme for dextran.

**ACKNOWLEDGMENTS.** Mette Pries is acknowledged for technical assistance and Anne Blischer for performing the amino acid analysis. Haruhide Mori from Hokkaido University is acknowledged for valuable discussions on the determinants of 1,6-$\alpha$-glucosidase activity. Joakim Mark Andersen is acknowledged for performing the amino acid analysis. Haruhide Mori from Hokkaido University is acknowledged for help with crystallization. MAX-lab and MAX-lab staff are thanked for beam time and assistance as well as the DANSCTT program from the Danish National Research Council for travel support.

This work was supported by a grant from the Danish Strategic Research Council, Committee of Health and Nutrition, to the project Gene Discovery and Molecular Interactions in Prebiotics/Probiotics Systems: Focus on Carbohydrate Prebiotics.

**REFERENCES.**

Appendix VIII

Paper: Efficient secretory expression of functional barley limit dextrinase inhibitor by high cell-density fermentation of *Pichia pastoris*

I have contributed with an assay of the relative inhibition of LD by different LDI molecular equivalents to this paper.
Efficient secretory expression of functional barley limit dextrinase inhibitor by high cell-density fermentation of *Pichia pastoris*

Johanne Mørch Jensen\(^a\), Malene Bech Vester-Christensen\(^a\), Marie Sofie Møller\(^a\), Birgit C. Bønsager\(^a\), Hans Erik Mølager Christensen\(^b\), Maher Abou Hachem\(^a\), Birte Svensson\(^a,\*\)

\(^a\) Enzyme and Protein Chemistry, Department of Systems Biology, Saltsøs Plads Building 224, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

\(^b\) Metalloprotein Chemistry and Engineering, Department of Chemistry, Kemitorvet Building 207, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

**A R T I C L E   I N F O**

Article history:
Received 22 March 2011
and in revised form 11 April 2011
Available online 27 April 2011

Keywords:
Limit dextrinase inhibitor
Pullulan
Fed-batch bioreactor fermentation
Electrospray ionization mass spectrometry
High stability

**A B S T R A C T**

The limit dextrinase inhibitor (LDI) from barley seeds acts specifically on limit dextrinase (LD), an endogenous starch debranching enzyme. LDI is a 14 kDa hydrophobic protein containing four disulfide bonds and one unpaired thiol group previously found to be either glutathionylated or cysteinylated. It is a member of the so-called CM-protein family that includes \(\alpha\)-amylase and serine protease inhibitors, which have been extremely challenging to produce recombinantly in functional form and in good yields. Here, LDI is produced in very high yields by secretory expression by *Pichia pastoris* applying high cell-density fermentation in a 5 L fed-batch bioreactor. Thus about 200 mg of LDI, which showed twofold higher inhibitory activity towards LD than LDI from barley seeds, was purified from 1 L of culture supernatant by His-tag affinity chromatography and gel filtration. Electrospray ionization mass spectrometry verified the identity of the produced glutathionylated LDI-His\(_6\). At a 1:1 M ratio the recombinant LDI completely inhibited hydrolysis of pullulan catalyzed by 5–10 nM LD. LDI retained stability in the pH 2–12 range and at pH 6.5 displayed a half-life of 53 and 33 min at 90 and 93 \(^\circ\)C, respectively. The efficient heterologous production of LDI suggests secretory expression by *P. pastoris* to be a promising strategy to obtain other recombinant CM-proteins.

© 2011 Elsevier Inc. All rights reserved.

---

**Introduction**

Plant seeds contain a large number of proteinaceous inhibitors that target hydrolytic enzymes and have important roles in plant defense and starch mobilization [1]. Inhibitors belonging to the CM-protein family, also known as cereal-type inhibitors are small, hydrophobic proteins of 110–160 amino acid residues containing four or five disulfide bonds [1]. They are abundant in cereal seed endosperm and include besides inhibitors acting on exogenous \(\alpha\)-amylases, trypsin- and chymotrypsin-like proteases in defense against bacteria, fungi, and insects [2], the endogenous inhibitor of barley limit dextrinase and several allergens [3].

Barley limit dextrinase inhibitor (LDI) inhibits the debranching enzyme limit dextrinase (LD) that specifically hydrolyses \(\alpha\)-1,6 glucosyl linkages with preference for short \(\alpha\)-limit dextrins rather than the longer branches present in amylopectin. LD, assigned into glycoside hydrolase family 13 subfamily 13 (GH13_13) ([www.cazy.org](http://www.cazy.org)) [4], is the sole starch debranching activity in the germinating seed [5,6] and it occurs in a free active form as well as a bound inactive form, possibly in complex with LDI [7,8]. The ratio of free to bound LD increases during germination probably reflecting the de novo synthesis of LD and inactivation of LDI [7] due to disulfide reduction by thioredoxin h [9] and/or hydrolysis by cysteine proteases [7]. LDI is synthesized in the developing seed [10], together with small amounts of LD, proposed to play a role in starch biosynthesis [11,12]. Thus antisense down-regulation of LDI in transgenic barley resulted in increased LD activity in developing and germinating grains [13]. LD is important in malting and mashing and its inhibition by LDI leads to branched unfermentable dextrans in the beer [8]. LDI contains 114 amino acid residues, four disulfide bonds and one cysteine residue (presumably Cys59[14]), which was reported to be present as a mixed disulfide with either glutathione or cysteine [15]. These two LDI forms of pi 6.7 and 7.2 were purified in small amounts from barley seeds and shown to share the same high inhibition potency towards LD [15]. The LD/LDI complex has a 1:1 stoichiometry as determined by mass spectrometry [16]; generally, however, very little is known on the properties and structure of this complex as opposed to several \(\alpha\)-amylase inhibitors and their target enzymes [17,18]. LD and \(\alpha\)-amylases both belong to GH13, but LD and other starch debranching enzymes possess a more open active site topology compatible with accommodation of branched substrates with \(\alpha\)-1,4 linked main-chain and \(\alpha\)-1,6 branches [19–22]. Access to recombinant LD

---

\* Corresponding author. Fax: +45 4588 6307.

E-mail address: bios@bio.dtu.dk (B. Svensson).
and LDI is a prerequisite to gain detailed insight into their biologically important interaction because only very small amounts of LD and LDI can be obtained from natural sources [5,15]. Functional LD was recently produced by secretory expression in *Pichia pastoris* [23], whereas various attempts to produce LDI in *Escherichia coli* have failed. CM-proteins are challenging to produce recombinantly, and despite their interesting roles in defense against pathogens and pests and in allergies, only small amounts of a few functional CM-proteins; e.g., the corn inhibitor of activated Hageman factor (CHFI) [25]; wheat monomeric inhibitor 0.28 (WMAI-1) [26]; and wheat CM16 protein from *Triticum aestivum* [24], were obtained by expression in *E. coli* and transformed into an amylase inhibitor (BIII) from rye [24]; corn inhibitor of activated Hageman factor (CHFI) [25]; wheat monomeric inhibitor 0.28 (WMAI-1) [26]; and wheat CM16 protein [27] were obtained by expression in *E. coli* as inclusion bodies followed by refolding. Besides enabling biochemical, structural, and mutational analyses of the LD/LDI complex, the present effective production of recombinant LDI by secretory expression from *P. pastoris* demonstrates the potential for successful heterologous expression of other CM-proteins in this host.

**Materials and methods**

**cDNA isolation and construction of the LDI expression plasmid**

RNA was extracted (RNEasy Plant Mini kit; Qiagen, Düsseldorf, Germany) from frozen, dehusked, homogenized (mortar cooled by liquid nitrogen) barley seeds (cultivar Morex). cDNA encoding residues 25–138 of LD (NCBI database Accession No. ABB88573), identical to the identified protein sequence [3], was obtained by RT-PCR (One-Step RT-PCR Kit, Qiagen) with primers 5'-GGATCCACCTGAGCCACGTCAAGACCG-3' (sense) and 5'-GGATCCCTTATCCCCGCTCTGAGGACCA-3' (antisense). The PCR was cloned into pCR 2.1-TOPO (Invitrogen, Carlsbad, CA) in an insert and BamHI sites (in bold) and transformed into *E. coli* TOP10 competent cells (Invitrogen). This construct served as PCR template with primers 5'-AAAAAAAGATACCTCAGAGGAGCG-3' (sense) and 5'-TGAT TAACTGCTTTCAATGAGCAGTAGATCCTGAGGCTCCTG-3' (antisense) in a hot-start PCR at 94 °C (3 min), 3 cycles of 94 °C (30 s), 50 °C (30 s), and 72 °C (45 s), followed by 22 cycles of 94 °C (30 s), 58 °C (30 s), and 72 °C (45 s) and a final elongation at 72 °C (7 min) to give an ampiclon encoding LDI with a C-terminal His6-tag (in italics), which was cloned in-frame with the *α*-factor secretion signal of pPIC ZA (Invitrogen) using EcoRI and KpnI sites (in bold). The EcoRI site caused an extension of the LDI wild-type sequence by Glu-Phe according to the guidelines for methanol feeding of the Mut+ phenotype. The glycerol batch phase (15 h) and glycerol fed-batch phase (10 h per column volume) were added to the fermentor (B. Braun Biotech International, Melsungen, Germany) with a methanol gradient of 0.5% methanol (v/v) every 24 h and centrifuged (12,000 *g*, 30 min) after 72 h. Supernatants were analyzed by SDS–PAGE and for LD activity and the transformant secreting LDI in highest amount was used for large scale fermentation in 5 L Biostat B bioreactor (*B. Braun Biotech International, Melsungen, Germany*) with automated control of pH, temperature, foam, and dissolved oxygen tension. The transformant was grown (in BMGY in 300 mL shake flask, 18 h, 30 °C, harvested (1,500 g, RT, 5 min), resuspended in BMGY to OD600 = 52, and used (180 mL) to inoculate 3 L fermentation basal salt medium supplemented with PTM1 trace salts [29], 1% (w/v) casamino acids, and 1 × 10-5 (w/v) biotin in the bioreactor. Glycerol (50% w/v) and methanol (100% v/v), both added PTM1 trace salts and 2 × 10-5 (w/v) biotin, were used as carbon sources during fed-batch glycerol and methanol phases, respectively. The glycerol batch phase (15 h) and glycerol fed-batch phase (10 h per column volume) were added to the fermentor (B. Braun Biotech International, Melsungen, Germany) with a methanol gradient of 0.8–1.9 g/L (h) glycerol, which generated biomass under substrate limited conditions and derepressed the alcohol oxidase promoter (AOX1)1, were maintained at 30 °C. This was followed by a methanol fed-batch phase (72 h) initiated by a temperature decrease to 20 °C, and onset of a methanol gradient of 0.8–1.9 g/L (h) in 18 h to insure adaptation to methanol metabolism, according to the guidelines for methanol feeding of the Mut+ phenotype [29,30]. The pH was maintained at 5.5 by addition of 28% (w/v) ammonia.

**Purification of LDI**

The culture supernatant (3.3 L) was recovered by centrifugation (14000g, 4 °C, 60 min), imidazole and NaCl were added to 10 and 500 mM, respectively, and pH was adjusted to 7.4 by K2HPO4(aq). The filtered (0.45 μm) supernatant was applied to 120 mL/φ Ni-NTA superflow resin (Qiagen; 10 mL packed in XK 16 column; GH Healthcare, Uppsala, Sweden) equilibrated with 20 mM sodium phosphate, pH 7.4, 10 mM imidazole, 500 mM NaCl (buffer A). After washing with buffer A for 6 column volumes, a more

---

**Notes:**

1. Abbreviations used: AOX, alcohol oxidase; CWW, cell wall weight; ESI-MS, electrospray ionization mass spectrometry; GH13, glycoside hydrolase family 13; IEF, isoelectric focusing; LD, limit dextrinase; LDI, limit dextrinase inhibitor; nanoES, nanoelectrospray; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate.
stringent wash in the presence of 23 mM imidazole was performed for 5 column volumes by mixing buffer A and 2.5% buffer B (i.e. buffer A containing 500 mM imidazole). Thereafter, LDI was eluted (60 mL/h) by a two step gradient reaching 100% buffer B in 22 column volumes. LDI-containing fractions were identified (SDS-PAGE), pooled, concentrated, buffer-exchanged to 10 mM bicine-NaOH, pH 8.5, 150 mM NaCl (Centricron 10 kDa cut-off, Millipore, Cork, Ireland), and subjected to gel filtration in four portions (Hi-load Superdex 75 16/60 column; GE Healthcare) using the above buffer including 0.5 mM EDTA (60 mL/h). LDI-containing fractions were pooled, concentrated, and stored at 4 °C. Chromatographic steps were performed using an ÄKTApurifier chromatograph (GE Healthcare) at 4 °C.

Protein characterization

Protein was analyzed by the BCA method (Pierce, Rockford, IL) using BSA as standard. Concentration of pure LDI was determined spectrophotometrically using ε280 = 5500 M⁻¹ cm⁻¹, determined by aid of amino acid analysis [31]. N-terminal sequencing (Proics 494 sequenator, Applied Biosystems, Foster City, CA) was carried as recommended by the manufacturer. Reducing SDS–PAGE (Nu-PAGE® Novex Bis-Tris 4–12%) and isoelectric focusing (Novex® pH 3–10) were performed according to the manufacturer’s recommendations (Invitrogen) and stained with Coomassie Brilliant Blue G-250 [32].

Electrospray ionization mass spectrometry

LDI was analyzed on an LCT Premier mass spectrometer (Waters, Milford, MA) with a nanoelectrospray (nanoES) ionization source. Data was collected in positive ion mode and the instrument was calibrated using 100 mg/mL CsI in 50% (v/v) isopropanol. LDI (100 µM) was exchanged into 600 mM ammonium acetate/ammonium hydroxide, pH 8.5 (Micro Bio-Spin P-6 size exclusion columns; Bio-Rad, Hercules, CA) and sprayed from nanoES capillaries (ES380; Proxeon, Odense, Denmark) using the following parameters; capillary voltage: 900–1500 V; sample cone voltage: 60 V; source temperature: 30 °C; and cone gas flow: 2 L/h (N2). LDI was equilibrated at 37 °C (10 min), and the reaction was initiated by addition of LD (4.3–8.7 nM)/LDI (0–23.2 nM) mixtures (110 µL). Aliquots (100 µL) were removed at 3 min intervals (0–12 min) and mixed with developing buffer (500 µL) and Milli-Q water (400 µL) [23]. Initial rates of hydrolysis at LD/LDI molar ratios of 0–2.7 were normalized to the reaction rate in the absence of LDI to give the relative inhibition (%). The experiment was carried out in triplicate and the results were presented as means ± standard deviations of relative inhibition as a function of the LD/LDI molar ratio.

pH and temperature stability of LDI

LDI (3 µM) was incubated (18 d; 4 °C) at pH 2–12 in Britton and Robinson buffers [34] and residual inhibitory activity was measured (using the Limit-Dextrizyme assay, see above). Rates of thermal inactivation of LDI (3 µM) were determined at 90 and 93 °C in 100 mM sodium phosphate pH 6.5, from residual LDI activity (using the Limit-Dextrizyme assay) in aliquots removed at appropriate intervals and transferred to ice cold assay buffer. The inactivation rate constant k (min⁻¹) was determined from the linear regression of the data (R² = 0.9983 at 90 °C and R² = 0.9967 at 93 °C) assuming first order kinetics (ln(A0/A) = −kt); the half-life was calculated as t½ = ln2/k. The experiments were made in duplicate; standard deviations were <5%.

Results

High cell-density fermentation and LDI purification

The LDI inhibitory activity in the culture supernatant increased throughout the methanol induction to a final value of 28700 U/L (Fig. 2). The low activity at the start (0–18 h) of the induction reflected derepression of the AOX1 promoter after substrate limited growth during the glycerol fed-batch phase. The cell wet weight (CWW) increased from 187 g/L after the glycerol fed-batch phase to 232 g/L at the end of the methanol induction with an initial slight decrease due to adaptation to methanol metabolism (Fig. 2).

LDI was the predominant protein in the culture supernatant (Fig. 3A, lane 2). His-tag affinity chromatography (3.3 L supernatant) on a Ni-NTA column (10 mL) resulted in 240 mg LDI of

![Fig. 2. Progress of secreted LDI inhibitory activity (C) and CWW (cell wet weight) (O) during the methanol fed-batch phase. LDI is efficiently secreted in active form, which is reflected by the accumulation of LDI inhibitory activity throughout the induction phase of the fermentation.](image-url)
specific activity of 71.4 U/mg (Table 1) suggesting a capacity of around 24 mg LDI/mL Ni–NTA resin. SDS–PAGE showed LDI of an apparent molecular mass of 14 kDa (Fig. 3A, lane 3), in agreement with the theoretic value of 13,455 Da. The Ni–NTA eluate was purified by gel filtration in four consecutive rounds (60 mg protein/round), which removed small amounts of LDI dimer and resulted in 138 mg LDI of 82.1 U/mg with an overall 3.0-fold purification (Table 1; Fig. 3A, lane 4). The flow-through from the overloaded Ni–NTA column contained large amounts of LDI (Fig. 3A, lane 5), and based on the specific activity of purified LDI, 1.15 g LDI was present in the culture supernatant (3.3 L). Based on the above estimated capacity of the resin, a 60 mL Ni–NTA column would be suitable to purify all LDI from the culture supernatant in a single step. From this Ni–NTA eluate a total of 660 mg LDI could be purified by consecutive rounds of gel filtration (60 mg portions) assuming the same overall yield of 57% obtained for the gel filtration step.

Characterization of recombinant LDI

In addition to the major LDI component of pI 7.7 (theoretic pI 7.05) (www.expasy.org/tools/pi_tool.html), a few minor ones appeared in a smear of slightly lower pl (Fig. 3B). Electrospray ionization mass spectrometry (ESI-MS) showed LDI charge states +6,+7, and +8 (Fig. 4A). The main component of 13,752.8 Da (Fig. 4B) matched the calculated mass of glutathionylated LDI-His6 (13,752.3 Da). No peak was found corresponding to LDI-His5 with cysteine bound. A component of 13,615.4 Da assumed to be LDI-His5, while three minor peaks matched LDI-His4, LDI-His3, and LDI-His2 (peaks 5, 3, and 1, respectively, Fig. 4B). Probably these forms with less than 6 C-terminal histidine residues are found in the smear seen in IEF (Fig. 3B). ESI-MS furthermore showed two small peaks of 13,410.7 and 13,272.7 Da (peaks 4 and 2, Fig. 4B), corresponding to cleavage of the N-terminal tripeptide TLE from the LDI-His6 and LDI-His5, respectively.

Inhibitory activity and stability of LDI

The strong inhibition of LD (5–10 nM) at varying LDI/LD molar ratio in the range of 0–2.7 demonstrated the very potent activity of LDI (Fig. 5). Thus, LD was fully inhibited by equimolar amounts of LDI, consistent with an expected high affinity and the 1:1 stoichiometry found by ESI-MS for the complex of barley LDI and LD [16]. The data in Fig. 5 prove that the recombinant LDI is fully functional and binds to recombinant LD with the same molar stoichiometry (i.e. 1:1) as found for LDI and LD purified from barley seeds and malt, respectively.

Table 1

| Purification of LDI from 3.3 L P. pastoris culture supernatant by one round of Ni–NTA affinity chromatography followed by four consecutive rounds of Hiload Superdex 75 gel filtration applying portions of 60 mg protein (see Methods and Materials). |
|---|---|---|---|---|
| Total protein (mg) | Total activity (U) | Specific activity (U/mg) | Purification (fold) | Yield (%) |
| Supernatant | 3500 | 94,700 | 27.1 | 1.0 | 100.0 |
| Affinity chromatography | 240 | 17,130 | 71.4 | 2.6 | 20.9 |
| Gel filtration (4x) | 138 | 11,330 | 82.1 | 3.0 | 12.0 |

The estimated overall purification yield for repeated purification of the Ni–NTA flow-through followed by gel filtration is 57% which would result in 660 mg pure LDI from the whole fermentation.
LDI was extremely stable at pH 2–11 and 4°C, and at pH 12 it lost only about 30% activity after 18 d (not shown). Moreover, irreversible inactivation kinetics of LDI (Fig. 6) were suggested to be pivotal for LD binding [14]. By contrast, especially Ser4 and Val5 in the N-terminal part of LDI, were suggested to be crucial for LD interaction with LDI [36]. Similar to native barley LDI [14], recombinant LDI showed high affinity for LD proposed for native barley LDI [36] similar to rags [16], suggesting that the C-terminal region is not crucial for LDI inhibition. Finally, especially Ser4 and Val5 in the N-terminal part of LDI, were suggested to be pivotal for LD binding [14], based on interactions seen for the corresponding residues in two crystal structures of α-amylase/α-amylase inhibitor complexes [17,35]. The complete inhibition of LD by LDI at a 1:1 M ratio, attested that LDI is fully functional, and also confirmed the 1:1 stoichiometry demonstrated previously of the barley LD/LDI complex by ESI-MS [16]. This inhibitory capacity is consistent with high affinity for LD proposed for native barley LDI [36] similar to, of 15 and 57 nM reported for ragi α-amylase/trypsin inhibitor (RATI), and wheat α-amylase inhibitor 0.19, respectively, inhibiting porcine pancreatic α-amylase catalyzed hydrolysis of p-nitrophenyl-α-D-maltoside [37,38]. Finally, the very high stability of LDI at a broad pH range and high temperature reflects its rigid compact conformation maintained by the four disulfide bonds, which is in accordance with known structures for related inhibitors and their high binding affinities.

A very efficient and simple LDI expression and purification protocol was established, which enables future biochemical and structural analyses of the LD/LDI interaction. Furthermore it allows mutational analysis of LDI to unravel the structural basis for LD inhibition. Finally, P. pastoris may be used as host for heterologous production of other CM-proteins resulting in excellent yields and avoiding inclusion body formation reported using E. coli as a host for this family of proteins.

Acknowledgments

Anne Blicher is gratefully acknowledged for performing amino acid analysis. This work was supported by the Danish Research Councils for Independent Research/Natural Sciences and Technology & Production Sciences, the Carlsberg Foundation, a DTU Ph.D. fellowship (MBV), and an Oticon Foundation M.Sc. thesis stipend (JM).

References


