Chapter 1

β-Galactosidase enzymes and their galactooligosaccharide products
Introduction

Lactose (β-D-galactopyranosyl-(1→4)-D-glucose) is a disaccharide composed of one galactose molecule linked to a glucose molecule, which can be found in the milk of mammals [1], [2]. The content of lactose in bovine milk ranges from 4.4% to 5.2%, while human milk contains 7% lactose [1]. Human newborns have the ability to produce β-galactosidase (E.C. 3.2.1.23) enzymes to digest lactose, however, adults usually have lost the ability to produce this enzyme [3]. The absence or deficiency of β-galactosidase enzymes can lead to lactose intolerance upon the consumption of dairy products [4], [5]. Lactose can be removed from milk to solve this problem and be used as an ingredient in the food industry. This also provides opportunities for the development of high value-added lactose derivatives such as galactooligosaccharides (GOS). The market price of GOS is 10-12 times higher than that of lactose while the GOS prebiotic effects are widely accepted [6], [7], [8].

The concept of prebiotics was first introduced in 1995 by Gibson and Roberfroid [9], and now is defined as “a non-digestible compound that, through its metabolization by microorganisms in the gut, modulates composition and/or activity of the gut microbiota, thus conferring a beneficial physiological effect on the host” [10]. GOS are important commercial prebiotics which are added into infant formula as alternatives for human milk oligosaccharides (hMOS) [11], [12]. GOS are composed of a number of galactose units linked to a terminal glucose or galactose residue via different glycosidic bonds, with degrees of polymerization (DP) from 2 to 10 units [13]. Nowadays GOS are produced from lactose using microbial β-galactosidase enzymes [14], [15], [16].

Structures of β-galactosidases and reaction mechanism

β-Galactosidase enzymes belong to glycoside hydrolase (GH) families 1, 2, 35, and 42, which belong to the GH-A superfamily (Carbohydrate Active Enzymes
database, [http://www.cazy.org/](http://www.cazy.org/) [17]. β-Galactosidases in GH1, GH2 and GH42 are found predominantly in bacteria and fungi, whereas the enzymes in GH35 have been found in bacteria, fungi, animals and plants [18]. As shown in Table 1, crystal structures are available for several β-galactosidase enzymes. These β-galactosidase enzymes catalyze the hydrolysis of terminal β-galactose residues from various substrates. The β-galactose residues subsequently serve as growth substrates and are used for carbon metabolism, energy generation, and maintaining a normal physiological activity of the organisms. As shown in Figure 1, the domain organization of β-galactosidases from different GH families is quite different. For β-galactosidases in the GH2 family, the catalytic domain is the third domain from the N-terminus, while for β-galactosidases in GH35 and GH42 the catalytic domains are the first domains from the N-terminus [19], [20], [21], [22], [23]. The catalytic domain is a (β/α)_8 TIM barrel that contains one glutamic acid residue as the nucleophile and another glutamic acid residue as the proton donor [17], [24]. The β-galactosidases in the GH1, 2, 35, and 42 families display a retaining mechanism in their reaction. The catalytic nucleophile first attacks the anomeric center of lactose with the assistance of the proton donor, forming a galactosyl-enzyme intermediate while releasing glucose. The second step depends on the identity of the acceptor substrate: if water serves as the acceptor, the

![Diagram of domain organization of β-galactosidases from different Glycoside Hydrolase (GH) families](image)

**Figure 1.** Domain distribution of β-galactosidases from different Glycoside Hydrolase (GH) families [17], [18], [19], [20], [21].
intermediate undergoes hydrolysis and releases galactose; if lactose serves as acceptor substrate, a DP3 GOS (β-D-Galp-(1→x)-β-D-Galp-(1→4)-D-Glc) is

<table>
<thead>
<tr>
<th>Protein name</th>
<th>GH family</th>
<th>Organism</th>
<th>PDB code</th>
<th>UniProt code</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BglA</td>
<td>1</td>
<td><em>Thermotoga maritima</em></td>
<td>1OD0</td>
<td>Q08638</td>
<td>[26]</td>
</tr>
<tr>
<td>LacZ</td>
<td>2</td>
<td><em>Arthrobacter sp. C2-2</em></td>
<td>1YQ2</td>
<td>Q8KRF6</td>
<td>[27]</td>
</tr>
<tr>
<td>BgaD</td>
<td>2</td>
<td><em>Bacillus circulans</em> ATCC 31382</td>
<td>4YPJ</td>
<td>E5RWQ2</td>
<td>[28]</td>
</tr>
<tr>
<td>LacZ</td>
<td>2</td>
<td><em>Escherichia coli K12</em></td>
<td>4V40</td>
<td>P00722</td>
<td>[29]</td>
</tr>
<tr>
<td>BgaL</td>
<td>2</td>
<td><em>Paracoccus sp. 32d</em></td>
<td>5EUV</td>
<td>D1LZK0</td>
<td>[30]</td>
</tr>
<tr>
<td>BgaA</td>
<td>2</td>
<td><em>Streptococcus pneumoniae</em> serotype 4</td>
<td>4CU6</td>
<td>A0A0H2UP19</td>
<td>[31]</td>
</tr>
<tr>
<td>Lac4</td>
<td>2</td>
<td><em>Kluyveromyces lactis</em> CBS2359</td>
<td>3OB8</td>
<td>P00723</td>
<td>[32]</td>
</tr>
<tr>
<td>BgaC</td>
<td>35</td>
<td><em>Bacillus circulans</em> ATCC 31382</td>
<td>4MAD</td>
<td>O31341</td>
<td>[33]</td>
</tr>
<tr>
<td>Bgl35A</td>
<td>35</td>
<td><em>Cellovibrio japonicus</em> Ueda107</td>
<td>4D11</td>
<td>B3PBE0</td>
<td>[34]</td>
</tr>
<tr>
<td>BgaC</td>
<td>35</td>
<td><em>Streptococcus pneumoniae</em> TIGR4</td>
<td>4E8C</td>
<td>A0A0H2UN19</td>
<td>[17]</td>
</tr>
<tr>
<td>LacA</td>
<td>35</td>
<td><em>Aspergillus oryzae</em> RIB40</td>
<td>4IUG</td>
<td>Q2UCU3</td>
<td>[35]</td>
</tr>
<tr>
<td>Glb1</td>
<td>35</td>
<td><em>Homo sapiens</em></td>
<td>3THC</td>
<td>P16278</td>
<td>[36]</td>
</tr>
<tr>
<td>LacA</td>
<td>35</td>
<td><em>Penicillium sp.</em></td>
<td>1TG7</td>
<td>Q700S9</td>
<td>[37]</td>
</tr>
<tr>
<td>Tbg4</td>
<td>35</td>
<td><em>Solanum lycopersicum</em></td>
<td>3W5F</td>
<td>O81100</td>
<td>[16]</td>
</tr>
<tr>
<td>Bga1</td>
<td>35</td>
<td><em>Trichoderma reesei</em></td>
<td>3OG2</td>
<td>Q70SY0</td>
<td>[38]</td>
</tr>
<tr>
<td>Bca</td>
<td>42</td>
<td><em>Bacillus circulans</em> subsp. alkalophilus</td>
<td>3TTS</td>
<td></td>
<td>[39]</td>
</tr>
<tr>
<td>Gal42A</td>
<td>42</td>
<td><em>Bifidobacterium animalis</em> subsp. lactis BI-04 ATCC SD5219</td>
<td>4UNI</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td>LacZ2</td>
<td>42</td>
<td><em>Bifidobacterium bifidum</em> S17</td>
<td>4UCF</td>
<td>E3EPA1</td>
<td>[40]</td>
</tr>
<tr>
<td>GanB</td>
<td>42</td>
<td><em>Geobacillus stearothermophilus</em> T-6</td>
<td>4OIF</td>
<td>F8TRX0</td>
<td>[41]</td>
</tr>
<tr>
<td>R-β-Gal</td>
<td>42</td>
<td><em>Rahnella sp. R3</em></td>
<td>5E9A</td>
<td>A0A0B4U815</td>
<td>[42]</td>
</tr>
<tr>
<td>A4- β-Gal</td>
<td>42</td>
<td><em>Thermus sp. A4</em></td>
<td>1KWG</td>
<td>O69315</td>
<td>[43]</td>
</tr>
</tbody>
</table>
formed by transgalactosylation (Figure 2) [25], [26], [27]. This DP3 GOS may serve again as acceptor substrate and undergo another round of transgalactosylation. The transgalactosylation reaction thus results in GOS mixtures containing structures varying in size and in linkage types. Carbohydrates other than lactose can also serve as acceptors in the transgalactosylation reaction. The linkage type and degree of polymerization (DP) of GOS structures strongly depend on the β-galactosidase enzyme origin.

![Figure 2. Reaction scheme of β-galactosidase enzymes. This figure has been adapted from Bultema et al [25]. The acid/base catalyst and the nucleophile are both Glutamic acid residues. The hydrolysis reaction uses water as acceptor substrate, while the transgalactosylation reaction uses lactose and other carbohydrates as acceptor substrate.]
Identification of GOS structures produced by β-galactosidases

As mentioned above, the characterized β-galactosidase enzymes adopt the same reaction mechanism. However, they synthesize different GOS products reflecting variations in their crystal structures, especially in their active sites [46],[47]. One of the aims in our work is to understand the relation between the active site structures and the synthesized GOS compounds. This requires the availability and use of reliable techniques to characterize the composition and structures of the GOS compounds. Techniques like high-performance anion-exchange chromatography (HPAEC) coupled with pulsed amperometric detection (PAD), NMR spectroscopy, Mass Spectrometry (MS), and methylation analysis have been widely used in the identification of the GOS structures [48], [49], [50], [51], [52]. This has resulted in the identification of an increasing number of GOS structures. Rodriguez-Colinas et al. identified 5 structures in the GOS mixture produced by β-galactosidase from *Kluyveromyces lactis* [53]. Urrutia et al. found 9 structures in the GOS mixture produced by β-galactosidase from *Aspergillus oryzae* [54]. Yanahira et al. isolated 11 GOS structures from the products of β-galactosidase of *Bacillus circulans* ATCC 31382 [55]. Our laboratory has made big strides to identify GOS compounds, using a series of techniques. We identified 43 structures in the commercial Vivinal GOS produced with β-galactosidase of *B. circulans* ATCC 31382 [56],[57]. Recently van Leeuwen et al. compared 6 commercial GOS products with Vivinal GOS and found 13 new structures [58]. Taken together, a total of 60 structures have been characterized in the GOS produced by various β-galactosidase enzymes (Figure 3). GOS are generally used in the dairy industry, especially in infant formula, as alternatives for hMOS to give beneficial effects to babies [59]. Till now, more than 200 hMOS structures have been revealed [60], [61], [62], and the structural complexity of hMOS is considered as important for their multiple biological functions [63]. Although the GOS structures are less diverse than those of hMOS,
with the development of sensitive and accurate identification techniques and the
discovery and engineering of new \( \beta \)-galactosidases, the potential of discovering
new GOS structures is still growing.

Figure 3. The GOS structures synthesized from lactose in the transgalactosylation
reaction of \( \beta \)-galactosidases. Structure 1 is galactose, structure 2 is glucose, structure 5 is
lactose, all others are GOS structures. The numbers refer to GOS structures identified in
previous studies by van Leeuwen et al [56], [57], [58].
Efforts to improve the GOS yield and change the linkage specificity

β-Galactosidase enzymes incubated with lactose catalyze two types of reactions: hydrolysis and transgalactosylation. Hydrolysis results in the production of galactose, and transgalactosylation produces GOS mixtures. Due to the hydrolysis reaction, the GOS yield cannot reach 100% in industrial processes (e.g. *B. circulans* β-galactosidase has a GOS yield of ~63.5% with 50% (w/w) lactose, incubated at 60 °C for 20 h [64]). Relatively high lactose concentrations are generally needed to improve the transgalactosylation/hydrolysis ratio, but also results in remaining lactose, in addition to glucose and galactose [65]. There is a clear wish to improve the GOS yield and lower the hydrolysis products and lactose content in the final product mixture. Although GOS are generally used as alternatives for hMOS [66], their structural diversity is much less than that of hMOS. It may be of interest to try and change the linkage specificity of the β-galactosidase enzymes to produce different type of GOS mixtures or to enrich the GOS structure complexity or avoid GOS components implicated in allergenic effects [67]. There are two approaches to improve the GOS yield and/or change the linkage specificity of the products. The first approach is to change the reaction conditions, such as the reaction solvents, temperatures, substrate concentrations and so on; i.e. process engineering. The second approach is to modify the enzymes themselves; i.e. enzyme engineering. It has been shown that bio-solvents derived from dimethylamide and glycerol changed the regioselectivity of Biolacta N5 (commercial available β-galactosidase from *B. circulans*, Daiwa Kasei), and improved the yield of GOS products with (β1→6) linkages [68], [69]. A study of the β-galactosidases from *B. circulans*, *A. oryzae*, *K. lactis*, and *Kluyveromyces fragilis* found that higher temperatures resulted in higher GOS yields, and that high lactose substrate concentrations resulted in even higher GOS yields [70]. Other studies also found that the temperatures, pH values, lactose concentrations and enzyme origins not only contributed to the GOS yield,
but also influenced the composition of the GOS mixtures [71], [72], [73], [74], [75]. Besides, different enzyme immobilization techniques such as adsorption on celite, covalent coupling to chitosan, aggregation by cross-linking, and immobilized on magnetic polysiloxane-polyvinyl also contributed to the improvement of the GOS yield [75], [76], [77],[78], [79]. On the other hand, protein engineering is a powerful tool to optimize the enzyme catalytic efficiency and change the product specificity [46], [80], [81]. For example, deletion mutagenesis showed that removal of 580 amino acids from the C-terminus of the β-galactosidase from *Bifidobacterium bifidum* greatly improved its transgalactosylation ability [82]. The native enzyme only has transgalactosylation activity at 13.7% lactose concentration while the truncated enzyme has a relatively high yield of GOS (39%) at 10% initial lactose [82]. A single mutation (F426Y) in β-glucosidase from *Pyrococcus furiosus* increased the transglycosylation/hydrolysis ratio, increasing the GOS yield from 40% to 45%. A double mutant (F426Y/M424K) improved GOS synthesis at 10% lactose from 18% to 40% due to the increase in the ratio of transglycosylation to hydrolysis [83]. A mutagenesis approach was also applied to the β-galactosidase from *Geobacillus stearothermophilus*; mutation R109W increased the transglycosylation/hydrolysis ratio, and the yield of trisaccharide β-D-Galp-(1→3)-β-D-Galp-(1→4)-D-Glcp enhanced from 2% to 23% at a lactose concentration of 18% [84]. Double mutants F571L/N574S and F571L/N574A of *Thermotoga maritima* β-galactosidase increased the transgalactosylating efficiency of the wild-type enzyme up to 2-fold [85]. In another mutagenesis study of β-galactosidase from *Sulfolobus solfataricus*, the GOS yield was enhanced by 11% by mutating phenylalanine to tyrosine (F441Y), which may be caused by the introduction of new H-bonds [86]. A recent study showed that incubation of the C-terminally truncated β-galactosidase from *B. circulans* with monobodies, synthetic binding peptides which can modulate the catalytic properties of enzymes, altered the enzyme specificity in such a way that it barely
Chapter 1

produced any GOS higher than DP5 [87], whereas in the absence of monobodies this enzyme is capable of producing GOS up to at least DP8 [50].

**The beneficial impact of GOS**

GOS are well-known prebiotics and have been used in infant formula for decades [12], [66], [88]. A study has shown that the administration of highly purified (>95%) short-chain GOS in human subjects improved the lactose digestion and tolerance [89]. Another study showed that the use of GOS in the first 6 months of infant nutrition reduced the total number of infections, and cumulative incidence of infections [90]. The dietary intervention with GOS in the first two years of babies feeding reduced both the manifestation of allergies, as well as infections [91]. GOS produced by whole cells of *Bifidobacterium bifidum* NCIMB 41171 significantly increased the bifidobacterial population in the stool of human adults [92]. Purified GOS greatly inhibited the adhesion of pathogens to the epithelial cell surface [93]. A study found that the consumption of short-chain GOS and long-chain fructooligosaccharides (FOS) in the early life of newborns modulated the microbiota in a similar way as that of hMOS, namely by increasing the number of *Bifidobacterium* [94].

The main end products of microbial fermentation of GOS are short-chain fatty acids, which can prevent host colon cancer and other intestinal disorders [95]. It is generally considered that GOS benefit the host in the following ways. Firstly, the structures of GOS are similar to some pathogen receptors, thus they act as decoys resulting in pathogen binding and excretion instead of adhesion in the gut [96], [97]. Secondly, GOS inhibit the growth of toxic bacteria such as *Clostridium difficile*, and stimulate the growth of beneficial bacteria such as *Bifidobacterium adolescentis* and *B. bifidum* [98]. Thirdly, GOS modulate the human gut microbiota and increase the percentage of probiotic bacteria in the gut [99], [98]. These probiotic bacteria improve human health in various ways:
inhibitory compounds such as bacteriocins produced by probiotic bacteria inhibit the growth of certain pathogens [100]; probiotic bacteria compete for the limiting nutrients with pathogens, and also produce short-chain fatty acids that result in a lowering of the pH and inhibition of the growth of certain pathogens [100], [96]; probiotic bacteria adhere to the intestinal mucosa and block the adherence of enteropathogens [100], [101]; probiotic bacteria modulate the development of the immune system [102].

**Outline of the thesis**

GOS are generally considered as prebiotics and are widely used in the food industry and pharmacy. They are composed of galactose molecules and one glucose molecule through different glycosidic linkages. The linkage types and degree of polymerization (DP) of GOS structures strongly depend on the enzyme origin. Although there have been many studies of β-galactosidase enzymes and the produced GOS mixtures, there are still many questions that remain unanswered. It is already known that β-galactosidases produce different GOS mixtures, however, the enzyme active site structural details that determine the composition of GOS mixtures has remained unknown. What are the roles of specific amino acid residues in the β-galactosidase active site? What are the structural determinants for the GOS yield and linkage specificity? Can we engineer β-galactosidases to change their GOS linkage specificity and thus change or enrich the GOS structural complexity? Can we use β-galactosidases to synthesize products other than GOS?

In this thesis, we tried to answer these questions. We investigated the active site structural basis for the linkage specificity, to provide insights into the structure-function relationship and provide guidance for the engineering of β-galactosidase enzymes. In chapter 2, a more detailed analysis of GOS profiles of three commercial β-galactosidases from *B. circulans*, *K. lactis* and *A. oryzae* is
reported. The GOS yields, the linkage specificity, and the diversity of GOS produced by these enzymes are clearly different from each other. The presence of the monosaccharides glucose and galactose in the reaction mixture changed the GOS profile and yield, however, the influence of monosaccharides also depended on the enzyme origin.

In chapter 3, residue R484 near the +1 subsite of the C-terminally truncated β-galactosidase from *B. circulans* (BgaD-D) was subjected to site saturation mutagenesis. The mutant enzymes displayed significantly altered enzyme specificity, leading to a GOS mixture with mainly (β1→4) and (β1→3) linkages, while the wild-type enzyme gave a GOS mainly composed of (β1→4) linkages. Besides, the mutant GOS mixtures also contained 14 structures that are not found in the GOS produced by the wild-type enzyme. Chapter 4 investigated the functional roles of selected amino acid residues in the BgaD-D active site using site-directed mutagenesis. A detailed biochemical characterization and product profile analysis was presented, showing that these amino acid residues in the active site were crucial to the enzyme activity, linkage specificity, transgalactosylation versus hydrolysis, acceptor substrate selection.

In chapter 5, lactulose was used as substrate for the BgaD-D wild-type and R484H mutant enzymes to expand the application of β-galactosidase enzymes. The oligosaccharides derived from lactulose were identified and several new-to-nature compounds characterized. They were tested as sole carbon source for growth by Bifidobacteria, reflecting their potential prebiotic properties. Chapter 6 summarizes the results reported in this thesis and gives perspectives for future research.

**References**


of galacto-oligosaccharides at very high lactose concentrations with immobilized β-galactosidases from *Aspergillus oryzae*. Process Biochem. 46, 245–252.


27 Bultema, J. B., Kuipers, B. J. H. and Dijkhuizen, L. (2014) Biochemical characterization of mutants in the active site residues of the β-galactosidase
Chapter 1

enzyme of *Bacillus circulans* ATCC 31382. FEBS Open Bio 4, 1015–1020.


Chapter 1


Chapter 1

Quantification of prebiotics in commercial infant formulas. Food Chem. 194, 6–11.


