Universitat Ramon Llull

DOCTORAL THESIS

Title

STRUCTURAL-FUNCTIONAL ANALYSIS OF LACTO-*N*-BIOSIDASE FROM *Bifidobacterium bifidum*: A POTENTIAL BIOCATALYST FOR THE PRODUCTION OF HUMAN MILK OLIGOSACCHARIDES

Presented by	Cristina Val Cid
Centre	IQS School of Engineering
Department	Bioengineering
Directed by	Dr. Magda Faijes and Dr. Antoni Planas

C. Claravall, 1-3
08022 Barcelona
Tel. 936 022 200
Fax 936 022 249
E-mail: urlsc@sec.url.es
www.url.es

A mis padres

A Jaume

“Quien no haya experimentado la irresistible atracción

de la ciencia, no podrá comprender su tiranía”

Frankenstein o el moderno Prometeo

Mary Shelley

ACKNOWLEDGEMENTS

Quisiera expresar mi agradecimiento a todas las personas que han hecho posible que esté escribiendo estas líneas, tanto a las que me han acompañado en esta etapa, como a las que me inspiraron a empezarla;

Al Dr. Toni Planas y a la Dra. Magda Faijes por llevar a cabo la dirección de este trabajo, y la confianza que depositaron en mí para realizarlo. Toni, gracias por permitirme formar parte de tu equipo de trabajo, por el conocimiento que has compartido y por todo lo que yo he aprendido. Magda, gracias por tu cercanía y tu motivación, tan necesaria en algunos momentos. Tu orientación y tu constancia han hecho de esta tesis un trabajo mejor. En definitiva, gracias por todo tu apoyo!

A Teresa, Dra. Pellicer, mi madre científica, como tú sueles decir. Empecé contigo en esto de la "ciencia", y mira hasta dónde hemos llegado! Porque contigo, y gracias a ti, he crecido tanto profesional como personalmente. Gracias por creer en mí, y, aunque no soy la primera que te lo dice, yo de mayor, quiero ser como tú!

A mis compañeros de bioquímica, los que están y los que estuvieron, ha sido un inmenso placer compartir estos años con vosotros. A los Cristians, Calzada y Benedito, fue un primer año increíble allí abajo, gracias! A Ellen, por tu paciencia, por saber escuchar y por ser tan buena persona. A Montse, Nuria Abajo y Damián, por las tardes de risas. A Patri, por tu ayuda durante estos años. A Edu, a Xavi Turón y, a Xavi Pérez por la introducción al mundo de la analítica. A Marta Palomo y a Amanda por vuestro buen rollo. A Cris Alsina por tu disponibilidad para ayudar. A Laia y Nuria Orive que comienzan ahora esta aventura, mucha suerte! Y por último, pero no menos importante a mis compañeros de fatiga, a Javi, Oscar y Marta V. por compartir esta aventura juntos. Especialmente, a Victoria, un referente del lab, fuiste la primera persona que conocí en bioch y el buen rollo que me trasmitiste ese día, continúa hasta ahora. Admiro la pasión que tienes por la ciencia, y créeme, te llevará lejos! A Sergi, un grande, por tu empatía, tus consejos, tu amistad, has conseguido que en un grupo de enzimas, tus algas tengan un lugar privilegiado! A Carles, por tu ayuda, por nuestras charlas de compañeros de mesa, y porque… qué hubieran sido esos cafés sin tu "aportación"? A Hugo, por tu complicidad, por tu ayuda y por tu amistad, gracias por dejarme conocerte! A Estela, por tu paciencia, optimismo y entrega por los demás, y por esa sonrisa que nos alegra el día.

A Xevi, Dr. Biarnés, por tu inestimable ayuda, tu aportación ha sido clave para esta tesis. Pero además, por tu optimismo, por tu buen humor, y por tu apoyo personal, que, aunque por modestia no reconozcas, yo lo he sentido, y te lo agradezco de corazón.

A Régis, Dr. Fauré, despite only being together for six months, it was a pleasure to work next to you, I learnt a lot. Thanks for your empathy and your wise advice.

A Marta Marszalek, my scientific sole mate (te lo cojo prestado), por tu apoyo, tu complicidad y tu amistad. Lo hemos compartido todo durante estos años, y aunque ahora sea de manera más esporádica, lo mejor es… que lo seguiremos haciendo!

A estas cuatro fantásticas chicas por su colaboración en esta tesis, Annie, Sara, Mireia, ojalá pudieras continuar con este trabajo, y especialmente a Jenny, fuiste mi primer "estudiante" y fue un placer trabajar contigo. Aprendí mucho, y me alegra que, finalmente, nos haya quedado una bonita amistad.

A toda la gente que he conocido en IQS durante estos años, gracias por los buenos momentos. A Cris Castells y a Sejin, por vuestro apoyo y vuestra amistad, y porque, cuan necesario es tener otro punto de vista! A Lourdes, por las largas jornadas de escritura en la biblio, que sin ti, no hubieran sido divertidas!

A la Dra. María José Hernáiz, por acogerme en su laboratorio para realizar la síntesis del sustrato clave de esta tesis, por su cercanía y por su orientación. También quisiera expresar mi agradecimiento a su equipo, en especial a Carlos y a Sara, por su ayuda y por hacerme sentir como en casa.

Al Dr. Andrés Fernández, por confiar en mí, por acogerme en su equipo y por darme la oportunidad de desarrollar mi carrera profesional previa a mi etapa doctoral. Ésta, sin duda, me inspiró para aventurarme en este proyecto.

A todos mis compañeros de Salvat, donde comenzó todo. A todos los que dedicaron parte de su tiempo a enseñar a la que empezó siendo la becaria. Qué buenos recuerdos guardo de aquella época! Y en especial, a las tres buenas amigas que me llevé, Teresa, Carmen López y Carme Alemany. Carmencita, gracias por todo! Por recogerme cada mañana para que no llegara tarde, por nuestras confidencias, por tu apoyo, y en definitiva, por tu amistad!

A mis chicas de Ferrer, Nuria, Vanessa, Wilmar, Clara, Laia, Gemma y Marta N., mis happy deivis! por lo bien que lo pasamos juntas, por vuestro apoyo y vuestra amistad.

A mis amigos, por su cariño, por su apoyo y su paciencia, menudos rollos me habéis aguantado cada sábado… María, Roger, por esos consejos locos!, Albert, Marta, la mejor psicóloga, Dani y en especial, Merca y Arantxa. Anxov eres como una hermana para mí. Mi BFF!, gracias por todos los momentos que hemos pasado juntas a lo largo de todos estos años, por comprenderme, por apoyarme y por estar siempre ahí!

A mi familia política, especialmente, a Lola y a Jaime, por vuestra comprensión y cariño.

A toda mi familia, por todo el cariño y apoyo que recibo de ellos, porque tengo mucha suerte de tener esta familia. A mi hermano Richi, por estar siempre a mi lado, aunque nos separen unos kilómetros, porque tienes el don de, siempre, hacerme reír, y… por esas peleas de bacterias! A mis tíos, Carlos y Ángela, porque sé que me mandáis toda la energía que podéis! A mi prima Tere, por tu apoyo incondicional. Y a mis abuelos, a mi yaya Regu, por preocuparte siempre por mí, por cuidarme y apoyarme, y a mi yayo Marce, que aunque ya no esté con nosotros, siempre me acompaña.

A Jaume, por ser el pilar de mi vida. Por tu apoyo, tu cariño, tu paciencia y tu ayuda. Por animarme a emprender esta aventura, por acompañarme en el camino y por ayudarme a finalizarla. Mi querido Castle, esta tesis también es tuya!

Y mi más profundo agradecimiento es para mis padres, Mª Antonia y Luis. Mamá, papá, no tengo palabras para agradeceros todo lo que habéis hecho por mí. Gracias por creer en mí, por vuestro cariño, vuestro apoyo, por ver siempre lo mejor de mí. Porque esta tesis ha sido posible gracias a vuestro esfuerzo y a la confianza que, día a día, depositáis en mí. Por esto y por mucho más, quiero dedicaros lo más valioso que he conseguido hasta el momento, esta Tesis.

SUMMARY

The potential health benefits of human milk oligosaccharides (HMO) have been studied for many years. It is well known that these oligosaccharides provide a protective barrier and nutritive support that infants with poor access to breast milk do not acquire in the first years of life. Human milk is considered to be unique among mammals in terms of the quantity and complexity of its oligosaccharides. To date, 130 chemical structures within HMO have been identified. No other natural resources provide access to these complex oligosaccharides in such large amounts, and until now, large scale synthesis of HMO has not been possible by any traditional organic chemistry methodology. Enzymatic synthesis is an alternative synthetic tool since enzymes can form the new glycosidic linkage between carbohydrates with high regio- and stereoselectivity.

The objective of this thesis is to evaluate the use of Lacto-*N*-biosidase from *Bifidobacterium bifidum* (LnbB) as an efficient biocatalyst in the following two ways: i) the structural-functional study of LnbB and ii) protein engineering of LnbB to generate biocatalysts able to synthesize the target lacto-*N*-tetraose.

Here, we have analysed the domain organization of GH20 enzymes, and accordingly, have defined two models of domain architectures. Model A, contains at least two domains, a non-catalytic GH20b domain, and the catalytic GH20 which is always accompanied with an extra α-helix. In contrast, Model B consists only of the catalytic GH20 domain. By expressing different truncated forms of LnbB, we have described the structural requirements for functionality of GH20 enzymes, and in particular for LnbB, to obtain a minimal functional unit that retains the enzymatic activity.

With regard to the synthesis of lacto-*N*-tetraose using new engineered LnbB proteins as biocatalysts, we envisage two different enzymatic strategies. First, the glycosynthase strategy, in which the activated donor is the corresponding synthetic sugar oxazoline and the enzyme, a mutant on the assisting residue, is able to transfer the donor to an acceptor without hydrolysis of the product. Second, the enhanced transglycosidase strategy, in which, a new generation of mutants on the acceptor subsites of the enzyme will be able to more favourably accommodate a sugar acceptor instead of water, and thus, increase transglycosylation activity.

SUMARI

Els efectes beneficiosos que els oligosacàrids de la llet materna (OLM) confereixen a la salut dels lactants s'ha estudiat durant anys. Aquests oligosacàrids proporcionen una barrera protectora i un suport nutritiu essencials, als quals no tenen accés els nens que no prenen llet materna. La llet humana és considerada única respecte a la resta de llets de mamífers pel que fa a la quantitat i la complexitat d'oligosacàrids. Actualment, s'han identificat més de 130 estructures químiques diferents de OLM, i no es disposa de cap recurs natural que proporcioni accés a aquestes estructures tan complexes i a bastament. De la mateixa manera, la síntesi química és complicada a causa de l'estructura tan complexa i diversa que presenten els OLM, i de moment, la síntesi en gran escala no ha estat possible. La síntesi enzimàtica, en canvi, es presenta com una eina alternativa de síntesi d'aquestes molècules complexes atès que, en la naturalesa els enzims són els responsables de formar enllaços glicosídics entre carbohidrats amb alta regio- i estéreo-selectivitat.

L'objectiu d'aquesta tesi és avaluar l'ús de l'enzim Lacto-*N*-biosidasa de *Bifidobacterium bifidum* (LnbB) com un biocatalizador eficient des de dues perspectives diferents: i) l'estudi estructural-funcional de LnbB i ii) la generació de biocatalizadors capaços de sintetitzar l'oligosacàrid d'interès (lacto-*N*-tetraosa) mitjançant enginyeria de proteïnes en l'enzim LnbB.

En aquesta tesi, hem analitzat l'organització dels dominis dels enzims de la família GH20, i, en conseqüència, hem definit dos models d'arquitectures. El Model A conté almenys dos dominis, un domini GH20b no catalític i el GH20 catalític, que sempre es presenta acompanyat d'una α-hèlix extra. En canvi, el Model B consisteix únicament en el domini catalític GH20. Mitjançant l'expressió de diferents formes truncades de LnbB, hem descrit els requeriments estructurals per a la funcionalitat dels enzims GH20, i en particular per LnbB, per tal d'obtenir la unitat funcional mínima que conservi l'activitat enzimàtica.

Respecte a la síntesi de la lacto-*N*-tetraosa usant com a biocatalizador noves proteïnes de LnbB obtingudes mitjançant enginyeria de proteïnes, hem contemplat dues estratègies enzimàtiques diferents. En primer lloc, l'estratègia de glicosintasa, en la qual l'enzim (amb mutació en el residu assistent) és capaç de transferir el corresponent donador activat (sucre sintètic derivat d'oxazolina) a un acceptor, sense hidròlisi del producte. En segon lloc, l'estratègia de transglicosilació millorada, en la qual, una nova generació de mutants en els llocs d'unió a l'acceptor seran capaços d'acomodar de manera més favorable un sucre en lloc d'una molècula d'aigua, i d'aquesta manera, augmentar l'activitat de transglicosilació.

SUMARIO

Los efectos beneficiosos que los oligosacáridos de la leche materna (OLM) confieren a la salud de los lactantes se han estudiado durante años. Estos oligosacáridos proporcionan una barrera protectora y un soporte nutritivo esenciales, a los que, los niños que no toman leche materna no tienen acceso. La leche humana se considerada única respecto al resto de leches de mamíferos en cuanto a cantidad y complejidad de oligosacáridos. Actualmente, se han identificado más de 130 estructuras químicas diferentes de OLM, y no se dispone de ningún recurso natural que proporcione acceso a estas estructuras tan complejas y en cantidad suficiente. Del mismo modo, la síntesis química es complicada debido a la estructura tan compleja y diversa que presentan los OLM, y por el momento, la síntesis en gran escala no ha sido posible. La síntesis enzimática, en cambio, se presenta como una herramienta alternativa de síntesis de éstas moléculas complejas dado que, en la naturaleza las enzimas son las responsables de formar enlaces glicosídicos entre carbohidratos con alta regio- y estéreo-selectividad.

El objetivo de esta tesis es evaluar el uso del enzima Lacto-*N*-biosidase de *Bifidobacterium bifidum* (LnbB) como un biocatalizador eficiente desde dos perspectivas diferentes: i) el estudio estructural-funcional de LnbB y ii) la generación de biocatalizadores capaces de sintetizar el oligosacárido de interés (lacto-*N*-tetraosa) mediante ingeniería de proteínas en el enzima LnbB.

En esta tesis, hemos analizado la organización de los dominios de enzimas GH20, y, en consecuencia, hemos definido dos modelos de arquitecturas de dominio. El Modelo A contiene al menos dos dominios, un dominio GH20b no catalítico y el GH20 catalítico, que siempre se presenta acompañado de una α-hélice extra. Por el contrario, el Modelo B consiste únicamente en el dominio catalítico GH20. Mediante la expresión de diferentes formas truncadas de LnbB, hemos descrito los requerimientos estructurales para la funcionalidad de las enzimas GH20, y en particular para LnbB, de modo que se obtenga la unidad funcional mínima que conserve la actividad enzimática.

Respecto a la síntesis de la lacto-*N*-tetraosa usando como biocatalizador nuevas proteínas de LnbB obtenidas mediante ingeniería, hemos contemplado dos estrategias enzimáticas diferentes. En primer lugar, la estrategia de glicosintasa, en la que el enzima (un mutante en el residuo asistente) es capaz de transferir el correspondiente dador activado (azúcar sintético derivado de oxazolina) a un aceptor, sin hidrólisis del producto. En segundo lugar, la estrategia de transglicosilación mejorada, en la que, una nueva generación de mutantes en los sitios de unión al aceptor serán capaces de acomodar de manera más favorable un aceptor de azúcar en lugar de una molécula de agua, y de este modo, aumentar la actividad de transglicosilación.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDIX FIGURES

Appendix 1

Appendix 1

Appendix 3

Appendix 4

Appendix 5

LIST OF TABLES

LIST OF ABBREVIATIONS

3D	three dimensional
aa	amino acid
Ac-GlcNAc	peracetylated *N*-Acetylglucosamine
Ac-LNB	peracetylated Lacto-*N*-biose
ACN	Acetonitrile
AcOEt	Ethyl acetate
AcOH	Acetic acid
Ala	Alanine
Apr	Ampicillin resistant
aq	aqueous
arom	aromaromatic
Asn	Asparagine
Asp	Aspartate
BCA	Bicinchoninic acid
β-Gal-3	β-Galactosidase from *Bacillus circulans* ATCC 31382
CAZy	Carbohydrate Active enZymes
CBM	Carbohydrate Binding Modules
COSY	Correlation spectroscopy
Cy	Cyclohexane
d	doublet
DCE	Dichloroethane
DCM	Dichloromethane
dd	doublet of doublets
ddd	doublet of doublet of doublets
DMAP	4-Dimethylaminopyridine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ELSD	Evaporative Light Scattering
ENGase	Endo *N*-acetylglucosaminidase
Et_3N	Trimethylamine
Fuc	Fucose
Gal	Galactose
GH	Glycosyl Hydrolase
Glc	Glucose
GlcNAc	*N*-Acetylglucosamine
Gln	Glutamine
Glu	Glutamate
GT	Glycosyltransferase
HEX	*N*-acetylhexosaminidase
His	Histidine
HPLC	High Performance Liquid Chromatography
HPSEC	High Performance Size-Exclusion Chromatography
HSQC	Heteronuclear Single-Quantum Correlation Spectroscopy
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani
LC/MS-TOF	Liquid Chromatography / Mass Selective– Time Of Flight
Leu	Leucine
LnbB	Lacto-*N*-biosidase from B*ifidobacterium.bifidum*
LNB	Lacto-*N*-biose
LNT	Lacto-*N*-tetraose
m	Multiplet
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization – Time Of Flight
MeOH	Methanol
MS	Mass Spectrometry
Mw	Weight average molecular mass
Neu5Ac	*N*-acetylneuraminic acid
NMR	Nuclear Magnetic Resonance
o/n	overnight
OMe	methoxy = methyl ether
oxa-GlcNAc	1,2-oxazoline derivative of *N*-acetyl-D- glucosamine
oxa-LNB	1,2-oxazoline derivative of 2-Acetamido-2-deoxy-3-O-(β-D-galactopyranosyl)-β-D-glucopyranose
PDB	Protein Data Bank
Phe	Phenylalanine
PMSF	Phenylmethylsulfonyl fluoride
***p*-NP**	4-nitrophenyl
***p*-NP-Gal**	4-nitrophenyl-β-D-galactpyranoside glucosamine

***p*-NP-GlcNAc**	4-nitrophenyl β-*N*-acetyl-D-
***p*-NP-LNB**	4-nitrophenyl 2-acetamido-2-deoxy-3-O-(β-D-galactopyranosyl)-β-D-glucopyranoside
Py	pyridine
RT	Room Temperature
s	singlet
SDS-PAGE	Sodium Dodecyl Sulfate-poliAcrylamide Gel Electrophoresis
t	triplet
TLC	Thin Layer Chromatography
Triflate	Trifluoromethanesulfonate
TMSOTF	Trimethylsilyltrifluoro-methanesulfonate
Tyr	Tyrosine
UV	Ultraviolet
VMD	Visual Molecular Dynamics

CHAPTER 1.
LITERATURE REVIEW

1.1 GLYCOBIOLOGY OF HUMAN MILK

Human milk is considered the gold standard for infant nutrition. It is defined as a very complex mixture that apart from providing complete postnatal nutrition, delivers bioactive components for the growth and development of the intestinal and immune system of the infants. Its composition is dynamic, and varies within a feeding, diurnally, over lactation, and between mothers and populations, but in general terms, it is consist of essential nutrients such as; lactose, fatty acids and proteins, micronutrients such as; nucleotides, vitamins and minerals, and a wide variety of bioactive molecules, which include immunoglobulins, like IgA, antimicrobial agents such as lactoferrin and lysozyme, and human milk glycoconjugates (Garrido *et al*., 2013; Smilowitz *et al*., 2014).

Glycans in milk can be found as free human milk oligosaccharides (HMO), or conjugated to peptides, proteins or lipids. One of the most striking features of breast milk is the diversity and abundance of complex glycans. Large quantities of mucins and glycosaminoglycans are present in its composition, and almost all of the milk proteins are glycosylated (Dallas *et al*., 2012). However, the major glycans in milk are lactose and HMO, representing approximately 6 and 1 % of the total milk, respectively (Figure 1.1). In contrast to other milk components like fatty acids, oligosaccharides do not depend on the nutritional status of the lactating mother, they are controlled by genetics mechanisms (Hennet *et al*., 2014). Several of these glycans are not absorbed by the small intestine and transit through the colon, poorly contributing to infant nutrition. Although their impact and biological activities are not entirely understood, they play a role in the protection and stimulation of development of the nurslings (Newburg, 2013).

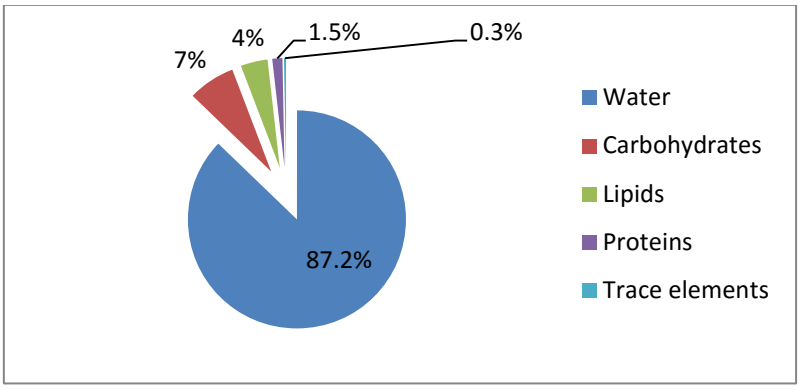


Figure 1.1 Human milk composition.
General molecular classes given in percentage of total mass. Image adapted from Hennet *et al*., 2014.

1.1 HUMAN MILK OLIGOSACCHARIDES (HMO)

Human milk oligosaccharides (HMO) are a group of unconjugated structurally complex glycans that are unique to human milk. HMO are the third largest solid component of human milk, after lactose and lipids, and they are exclusively synthesised in the mammary glands, and only during lactation. Oligosaccharide composition vary between women and over lactation period with levels of >20 g/L in colostrum and 5-14 g/L in mature milk (Bode, 2012; Coppa *et al*., 1999; Kunz *et al*., 2000).

HMO are consider unique among other mammal oligosaccharides with regard to the quantity and the structural complexity exhibit in milk composition (Garrido *et al*., 2012; Kunz *et al*., 2000; Urashima *et al*., 2012a).

1.2.1 HISTORICAL ASPECTS OF HMO

Interest in breast milk benefits started at the beginning of the 20th century when infant mortality rates were as high as 30% in the first year of the infants' life. By this time it was observed that breast-fed infants had 7 times more chances of survival than formula-fed infants (Kunz, 2012).

Research in HMO was driven by scientists with two different perspectives and interests. Chemists were trying to characterize the wide variety of oligosaccharides found in milk, while paediatricians and microbiologists were keen on the benefits that were observed in breast feeding (Bode, 2012).

In 1886, Theodor Escherich published the relationship between intestinal bacteria and the physiology of digestion in infants, being considered the pioneer researcher in this area (Escherich, 1886; Kunz, 2012). Inspired from this fact, by 1900 Herni Tissier and Ernst Moro, independently found differences in the composition of gut flora of breast-fed infants compared to formula-fed ones (Tissier, 1900; Weirich and Hoffmann, 2005). In 1926, Herbert Schönfeld reported that the whey fraction of human milk contains growth-promoting factors for *Lactobacillus bifidus*, later classified as *Bifidobacterium bifidum*, and named it Bifidus factor (Schönfeld, 1926).

At the same time, in 1888, Eschbach was the first who found that human and bovine milks did not contain the same type of lactose. Shortly after, Deniges observed that apart from lactose, human milk also contained an additional unknown carbohydrate fraction (reviewed in Kunz, 2012). Forty years later, Michael Polonowski and Albert Lespagnol established a method to characterize this carbohydrate fraction, which was insoluble in methanol, and they called it Gynolactose (Polonovski and Montreuil, 1954).

It was in 1950, when Paul György and Richard Kuhn established the connection between the Gynolactose and the Bifidus factor, and confirmed that the composition of the Bifidus factor was indeed oligosaccharides containing *N*-acetylglucosamine (Gyorgy *et al*., 1954).

With the development of new analytical techniques, the identification and characterization of new structures of HMO had been achieved. Heinz Egge was one of the leading pioneers in this field (Egge *et al*., 1983). Another milestones in HMO research were done by Akira Kobata, elucidating the ABO (blood type system) and secretor pathway of HMO (Kobata, 2003), and Clemens Kunz, studying the nutritional and biological properties of HMO (Kunz *et al*., 2000).

1.2.2 STRUCTURES AND COMPOSITION OF HMO

HMO consist of a pool of soluble oligosaccharides with a degree of polymerization of 3-15. They are composed of five monosaccharides: glucose (Glc), galactose (Gal), *N*-acetylglucosamine (GlcNAc), fucose (Fuc) and sialic acid (Sia), with the predominant form of *N*-acteylneuraminic (Neu5Ac) (Figure 1.2).

All HMO are characterised by terminal lactose in their reducing end, which can be fucosylated or sialylated to render human milk trisaccharides commonly included as HMO. The modification with fucose can occur at the terminal Gal in α-1,2 linkage to yield 2'-fucosyllactose, or at the reducing end of Glc in a α-1,3 linkage to generate 3'-fucosyllactose. The modifications with sialic acid are done at the terminal Gal in α-2,3 linkage to generate 3'-sialyllactose or in α-2,6 linkage to form 6'-sialyllactose. In addition, lactose can be further elongated and branched with lacto-*N*-biose (LNB; Galβ1-3GlcNAc) or *N*-acetyllactosamine (LacNAc; Galβ1-4GlcNAc) in β-1,3 and β-1,6 linkages to render two of the building blocks of the HMO structures. These core structures are the tetrasaccharides Lacto-*N*-tetraose (LNT; Galβ1-3GlcNAcβ1-3Galβ1-4Glc) core type 1, and Lacto-*N*-neotetraose (LNnT; Galβ1-4GlcNAcβ1-3Galβ1-4Glc) core type 2 (Bode, 2012; Garrido *et al*., 2013; Kobata, 2010; Kunz *et al*., 2000).

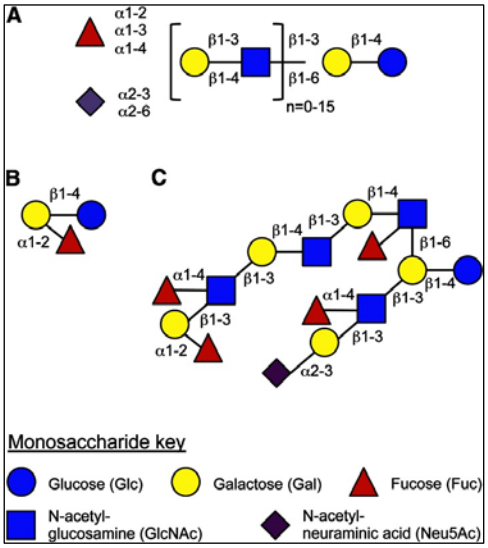


Figure 1.2 An example of the structural diversity of human milk oligosaccharides.
(A) Structural composition of HMO. If n = 0, the lactose backbone is either sialylated or fucosylated to form human milk trisaccharides such as 2'-fucosyllactose (B). If n>0, complex HMO are formed that can be branched and also modified like the sialylated and fucosylated iso-lacto-N-decaose (C). Glycan structures are depicted according to the recommendations of the Consortium of Functional Glycomics (http://www.functionalglycomics.org). Image from Bode and Jantscher-Krenn, 2012.

HMO are structurally classified into 11 more core groups that are built from core type 1 and type 2 (Table 1.1). These backbone structures can be further decorated with Fuc residues in α-1,2, α-1,3 and α-1,4 linkages, or with Neu5Ac in α-2,3 or α-2,6 linkages to achieve the wide complexity and diversity of HMO (Urashima *et al*., 2013).

The β-1,6 linkages introduce branched chains, and branched oligosaccharides are designated as *iso*-HMO. The rest of them generate linear chains, and linear ones are named as *para*-HMO. HMO can be further classified as acidic or neutral oligosaccharides, depending on the presence

of sialic acid or not. The carboxylate group of sialic acid is deprotonated at physiological pH and confers the net negative charge of the oligosaccharide in which is present (Bode, 2012; Garrido *et al*., 2013).

Table 1.1 The 13 core structures that occur in HMO.
(Table from Urashima 2013).

Structure	Name
Gal(β1-4)Glc	Lactose
Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)Glc	Lacto-*N*-tetraose
Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc	Lacto-*N*-neotetraose
Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-hexaose
Gal(β1-4)GlcNAc(β1-6) Gal(β1-4)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-neohexaose
Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc	*para*-Lacto-*N*-hexaose
Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc	*para*-Lacto-*N*-neohexaose
Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-octaose
Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-6) Gal(β1-4)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-neooctaose
Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	*iso*-Lacto-*N*-octaose
Gal(β1-3)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)GlcNAc(β1-3)Gal(β1-4)Glc	*para*-Lacto-*N*-octaose
Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-decaose
Gal(β1-4)GlcNAc(β1-6) Gal(β1-4)GlcNAc(β1-3) ⟩ Gal(β1-4)GlcNAc(β1-6) Gal(β1-3)GlcNAc(β1-3) ⟩ Gal(β1-4)Glc	Lacto-*N*-neodecaose

Due to the nature of chemical structures of HMO, a single molecular weight can be assigned to several structural isomers. Those variations of structures can range from linkage isomers (α-1,2 or α-1,3 or α-1,4) to variations in branching and positional isomers. The number of isomers can be up to 10 in some cases, with many of the isomers in relatively equal abundance (Zivkovic *et al*., 2011). An example is shown in Figure 1.3, in which Wu *et al.* (2011) determined a mass of 1511.6 which corresponded to one Fuc and one NeuAc on a lacto-*N*-hexaose (FS-LNH) or lacto-*N*-neohexaose core (FS-LNnH). This molecular weight was assigned to seven different isomers that were separated into different fractions using standard HPLC. The selected fractions were then analysed by Chip/TOF MS (Figure 1.3 a-g), and their structures were determined using standards (Figure 1.3 c-g) or by tandem MS and targeted exoglycosidase digestion (Figure 1.3 a-b).

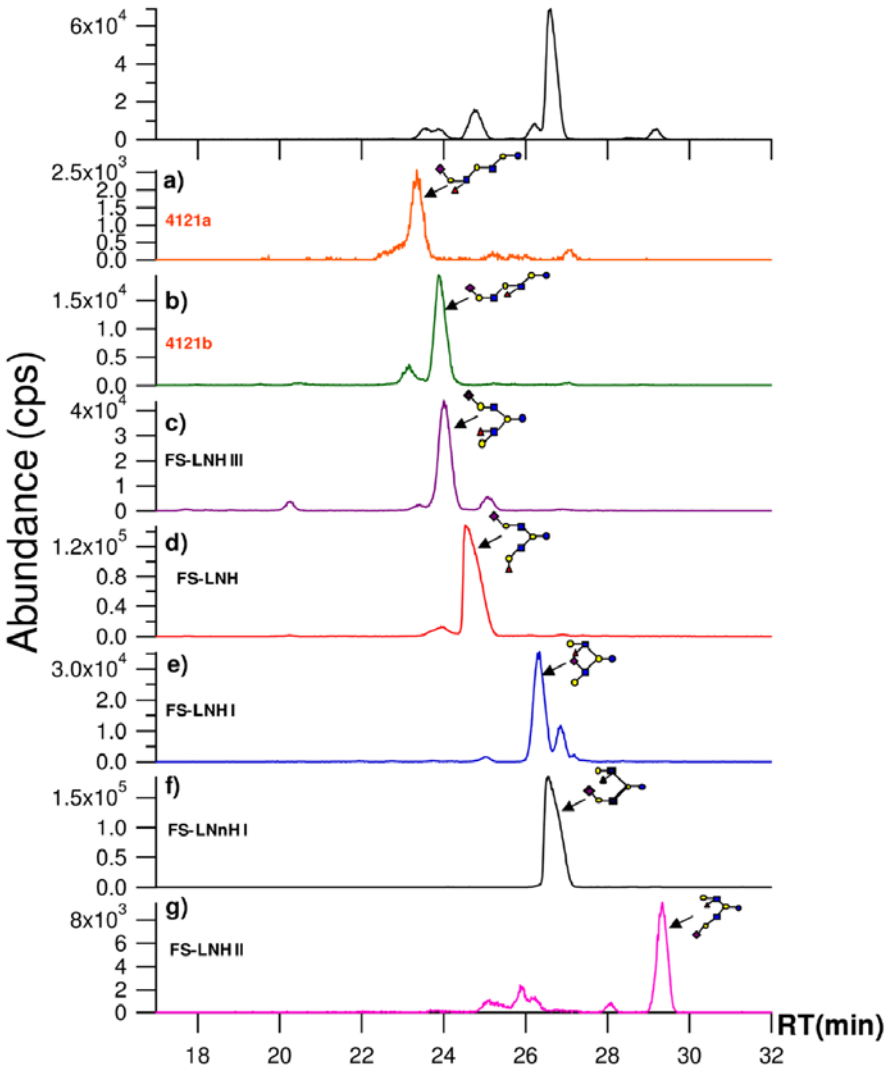


Figure 1.3 Isomeric structures of FS-LNH in human milk.
First image shows the extracted ion chromatogram (EIC) with neutral mass 1511.6 (m/z 756.8 (+2), MS inset). Images a–b show new structures with no assigned nomenclature. Image from Wu *et al*., 2011.

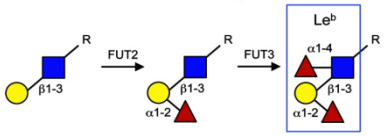
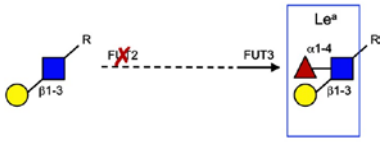
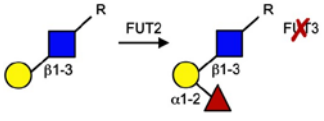
In terms of composition of HMO, there are important intra- and interpersonal variations, although oligosaccharide synthesis follows the cited general rules. In 2003, Kobata published a comparative study on HMO profiles of individual human milk, and revealed the presence of three different patterns that correspond with different blood types of those individuals. Different women, then, synthesise different subsets of oligosaccharides, and this fact depends on women's Secretor and Lewis blood group status (Kobata, 2003).

Lewis group blood types are named because of the fucose containing glycans found on the surfaces of erythrocytes and in glycoproteins of secreted fluids. Lewis motifs are found on the surfaces of cells and have been implicated in many diseases, from cancer to infection. It has been shown, for example, that Lewis b exhibits preferential binding to pathogens, specifically *Helicobacter pylori*. Lewis-related antigens have prognostic value for diseases as diverse as cancer and celiac disease. The Lewis structures are not synthesized in erythroblasts but are believed to be produced in the gut epithelium where they are shed into the digestive tract.

Then they are digested, reabsorbed, and transported as glycolipids into the plasma where they are, then, absorbed onto red blood cells (Zivkovic *et al*., 2011).

The addition of fucose is dependent of at least three different fucosyltransferases in a genetically determined process (Bode and Jantscher-krenn, 2012; Kunz *et al*., 2000). Lewis and Secretor genes have dominant alleles (Le and Se) that codify for two fucosyltransferases, and recessive alleles (se and le) that are not functional. Lewis gene codifies for the fucosyltransferase 3 (FUT3) that adds Fuc to subterminal GlcNAc of Type 1 chains in α-1,4 linkage, generating the Lewis a antigen in non-secretor individuals (Le a^+ b) (Table 1.2). Secretor gene codifies for the fucosyltransferase 2 (FUT2) that connects Fuc to terminal Gal of previously FUT3 modified type 1 chain in α-1,2 linkage. Individuals expressing Se gene are classified as secretors, and generate Lewis b antigens (Le a^- b^+). There is a third type of Lewis phenotypes corresponding to people that are Lewis recessive and can be either secretors or non-secretors (Le a^-b^-). In this case, it has been found that they possess oligosaccharides carrying Fuc in α-1,3 linkage to the reducing end Glc or internal GlcNAc of type 2 chains. Those results suggest that there is a third unknown gene, independent from Se and Le genotype, which codifies for a α-1,3-fucosyltranferase (FUTx).

Table 1.2 The Lewis and Secretor gene-related groups of HMO.
The Le and Se epitopes, which are characteristic for the Le phenotype in red blood cells and in human milk, are synthesized by the listed fucosyltransferases (FUT). The Le and Se genes code for the active FUT in presence of at least 1 functional allele (heterozygous with Lele or Sese, homozygous with LeLe or SeSe). The prevalence of the Le phenotypes is conferred to Europeans. Adapted from Blank *et al*., 2012 and Bode and Jantscher-Krenn, 2012.

Group	Epitope	Lewis phenotype, prevalence	Active fucosyltransferase
Secretor, Lewis positive Se^+Le^+		Le a^-b^+ 72%	FUT 3 α1-3/4 FUT 2 α1-2
Nonsecretor, Lewis positive Se^-Le^+		Le a^+b^- 22%	FUT 3 α1-3/4
Secretor, Lewis negative Se^+Le^-		Le a^-b^- 6%	FUT 2 α1-2
Nonsecretor, Lewis negative Se^-Le^-			-------

Based on the expression of these fucosyltransferases, 4 different HMO groups can be distinguished, as is observed in Table 1.2. According to fucosylation, women with a SeLe phenotype secrete milk with the most complex oligosaccharides, while women with sele phenotype secrete the least complex HMO (Blank *et al*., 2012; Bode and Jantscher-krenn, 2012; Kobata, 2003; Kunz *et al*., 2000).

In terms of sialylation, Tsuchida and coworkers reported that sialylation in α-2,6 of subterminal GlcNAc of oligosaccharides could be facilitated by two novel identified α-2,6-sialyltransferases, called ST6GalNAc V and VI (Tsuchida *et al*., 2003). However, little knowledge has been obtained for sialyltransferases that are responsible for α-2,3 sialylation . Variations in sialylated oligosaccharides among individuals may be due to differential expression and regulation of sialytransferases more than an all-or-nothing expression system (Bode, 2012; Bode and Jantscher-krenn, 2012).

Besides HMO variation among women, there are also differences in different periods of lactation from the same individual. The largest amounts of HMO are observed at early stages and they decrease over the course of lactation (Kunz *et al*., 2000). Furthermore, women delivering preterm have much wider variation (both between women and for a given woman over time) in the percentage of HMOs that contain fucose or sialic acid than women who deliver at term (Underwood *et al*., 2015).

1.2.2.1 The predominance of Type 1 oligosaccharides in human milk

The general trend of quantitative data of HMO reveals specific patterns for human milk and colostrum compare to other mammals. HMO are determined in higher quantities and complexity, with more diversity, and with longer structures than oligosaccharides in other mammals.

In terms of total oligosaccharide amounts, HMO concentrations arise at 12-13 g/L in mature milk, and 22-24 g/L in colostrum. While, for example in bovine milk, only traces of oligosaccharides have been detected (see Table 1.3) (Kunz *et al*., 2000; Ninonuevo *et al*., 2006). The only exception is the amount of oligosaccharides in elephant milk, which is higher than in humans, and may reflect the relative importance of these components in neonatal host defence, in endothelial leucocyte interactions or in brain development (Kunz *et al*., 1999).

Table 1.3 Oligosaccharides detected in human milk and cow's milk.
(Table adapted from Kunz 2000).

	Human	Bovine
Protein (g/L)	8	32
Fat (g/L)	41	37
Lactose (g/L)	70	48
Oligosaccharides (g/L)	5-15	0.05
Number of identified oligosaccharides	>150	~40
% fucosylated	50-80%	~1%
% sialylated	10-20%	~70%

To date, more than 200 different structures of milk oligosaccharides have been identified in a pooled human milk sample (Urashima *et al*., 2012b) containing mostly neutral and fucosylated glycans (Ninonuevo *et al*., 2006). Consensus data estimates that the proportion of fucosylated, sialylated, and nonfucosylated neutral HMOs in term breast milk is 35–50%, 12–14%, and 42–55%, respectively (Smilowitz 2014). It is important to emphasize that the main quantitative component of HMOs includes 20 to 25 oligosaccharides, while the remaining oligosaccharides constitute only minimal amounts (Gabrielli *et al*., 2011). Furthermore, only 50 structures represent 99% of the abundance oligosaccharides in human milk (Smilowitz *et al*., 2014).

Among all glycogonjugates present in human milk, type 1 chains, containing LNT structure, predominate over those chains containing LNnT structure. The predominance of type 1 oligosaccharides is observed in different ethnic population and independently of Le and Se phenotype (Urashima *et al*., 2012a).

Studies in other mammals revealed that their oligosaccharides differ in structure, complexity and quantity compared to HMO (Boehm and Stahl, 2007). On one hand, type 2 oligosaccharides are the most predominant glycans in domestic animals, and in some cases, the only ones detected (Kunz *et al*., 1999; Urashima *et al*., 2001). On the other hand, in the acidic fraction of animal milk oligosaccharides, two forms of sialic acid are present, *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc), while only Neu5Ac is present in human milk. Sialyllactose is the major component detected in caw milk. And finally, in the neutral fraction of oligosaccharides, linkages to Fuc are very rare in oligosaccharides from domestic animals' milk (Kunz *et al*., 2000).

Besides studies in domestic animals, Urashima reported a study with new and old monkeys and primates, in which confirmed that the predominance of type 1 oligosaccharide is a human specific feature rather than a primate or hominoid feature (Urashima *et al*., 2012a). The comparison of milk oligosaccharide profiles across evolution did not exactly match the primate phylogeny showing a non-sequential developmental pattern (Jeong *et al*., 2012)

The identification and structural assignment of HMO has been achieved using a variety of techniques. In 1996, Thurl and coworkers reported the first quantification of representative neutral and acidic HMO. They and other researchers used pH anion exchange chromatography with pulse amperometry detection system for separation of HMO (Coppa *et al*., 1999; Kunz *et al*., 1999; Thurl *et al*., 1996). Nowadays, capillary electrophoresis associated with magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) techniques have been used for chemically structure determination (Kogelberg *et al*., 2004; Wu *et al*., 2011). And the recent development of matrix-assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry (MALDI-FTICR MS) has allowed the accurate identification and quantification of HMO (Barboza *et al*., 2009; LoCascio *et al*., 2007).

Despite these methodological advances, there are some variations in terms of total HMO quantification among published data. One reason may be the different lactation stages at which the samples were obtained. In addition, there may be differences between the samples

from different donors, some of whom may be non-secretors, as well as, variation depending on their ethnic origins. Moreover, these quantitative results are also affected by the methodological differences among the studies as is shown in Table 1.4 (Urashima *et al*., 2012a). There are large differences among the values obtained by these authors. Nevertheless, it is clearly evidenced the predominance of type I oligosaccharides in HMO (Urashima *et al*., 2012a).

Among these complex oligosaccharides, the major secreted ones in human milk include lacto-*N*-tetraose (LNT) and lacto-*N*-neotetraose (LNnT). Lacto-*N*-hexaose (LNH) is usually identified as one of the most abundant. In addition, fucosylated molecules such as fucosyllacto-*N*-hexaoses (LNDFH I) and lacto-*N*-fucopentaoses I, II and III (LNFP I, II and III), which are fucosyllacto-*N*-tetraoses, are among the most abundant neutral HMO, apart from the trisaccharides 2'-fucosyllactose (2'FL), 3'-fucosyllactose (3'FL) (Chen, 2015; Garrido *et al*., 2015).

Finally, sialyl-lacto-*N*-tetraose (LSTa), 3'- sialyllactose (3'SL) and 6'-sialyllactose (6'SL) are the most abundant acidic HMO, although their concentrations do not arrive to 0.3 g/L (Garrido *et al*., 2015; Kunz *et al*., 2000).

Table 1.4 Concentrations of neutral HMO from 5 different studies.
Thurl *et al*. 1996, Kunz *et al*. 1999, and Coppa *et al*. 1999, used high pH anion exchange chromatography with a pulse amperometry detection system for the separation of HMO without any prior derivatization. Chaturdevi *et al*. 2001, determined the concentrations of neutral HMO at several lactation periods using reverse-phase HPLC following O-benzoyl derivatization of the HMO fraction. And, Asakuma *et al*., used reverse-phase HPLC after derivatization with 2-aminopyridine or 1-phenyl-3-methyl-5-pyrazolone. (d) means day of lactation.
2'-FL, Fucα1-2Galβ1-4Glc; 3'-FL, Galβ1-4[Fucα1-3]Glc; LDFT, Fucα1-2Galβ1-4[Fucα1-3]Glc; LNDFH I, Fucα1-2Galβ1-3[Fucα1-4]GlcNAcβ1-3Galβ1-4)Glc; LNDFH II, Galβ1-3[Fucα1-4GlcNAcβ1-3Galβ1-4[Fucα1-3]Glc; LNT, Galβ1-3GlcNAcβ1-3Galβ1-4Glc; LNnT, Galβ1-4GlcNAcβ1-3Galβ1-4Glc; LNFPI, Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc; LNFPII, Galβ1-3[Fucα1-4]GlcNAcβ1-3Galβ1-4Glc; LNFP III, Galβ1-4[Fucα1-3)]GlcNAcβ1-3]Galβ1-4Glc; LNH, Galβ1-3GlcNAcβ1-3[Galβ1-4GlcNAcβ1-6]Galβ1-4Glc. (Table adapted from Urashima 2012).

	Concentration					
Oligosaccharides	Kunz et al.	Chaturvedi et al.	Thurl et al.	Coppa et al. (d 4)	Coppa et al. (d 60)	Asakuma et al. (d 2)
		g/L				
2'-FL	0.45 ± 0.43	2.43 ± 0.26	1.84	3.93 ± 1.11	1.84 ± 0.39	2.01 ± 1.07
3-FL	0.07 ± 0.08	0.86 ± 0.10	0.46	0.34 ± 0.06	0.71 ± 0.07	0.28 ± 0.26
LDFT		0.43 ± 0.04	0.17			0.28 ± 0.30
LNT	1.09 ± 0.47	0.55 ± 0.08	0.86	0.84 ± 0.29	1.56 ± 0.57	1.44 ± 0.70
LNnT	Tr	0.17 ± 0.03	0.11	2.04 ± 0.55	0.95 ± 0.83	0.54 ± 0.14
LNFP I	1.26 ± 1.11	1.14 ± 0.18	0.67	1.36 ± 0.18	0.97 ± 0.61	2.08 ± 1.67
LNFP II			0.20	0.29 ± 0.22	0.29 ± 0.16	
LNFP III			0.28			
LNDFH I		0.50 ± 0.06	0.58	0.79 ± 0.25	1.18 ± 0.22	1.87 ± 1.55
LNDFH II	0.16 ± 0.11	0.09 ± 0.01	0.25			0.020 ± 0.025
LNH			0.13	0.07 ± 0.07	0.09 ± 0.02	

1.2.3 BIOSYNTHESIS OF HMO

As has previously been described, all HMO carry a lactose molecule in the reducing end of their chains. Accordingly, biosynthesis of HMO is suggested to be an extension of lactose synthesis that occurs in the Golgi (Blank *et al*., 2012; Bode, 2012). Lactose synthase, a complex of β-1,4-galactosyltransferase I (β-1,4GalT) and α-lactalbumin (α-LA), catalyses the formation of lactose from uridine diphosphate galactose (UDP-Gal, activated donor) and Glc (acceptor) by transglycosylation (See Figure 1.4) (Castanys-Muñoz *et al*., 2013).

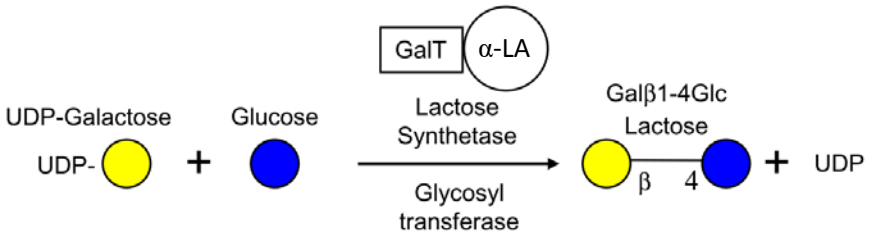


Figure 1.4 Biosynthesis of lactose.
The monosaccharide Uridine-diphosphate galactose (UDP-Gal) is transferred to glucose (Glc) by the action of Lactose synthase activity. Image from Castanys-Muñoz *et al*., 2013.

β-1,4GalT is constitutively expressed to transfer UDP-Gal to terminal GlcNAc during glycoconjugates synthesis, while α-LA is only expressed in the mammary gland of mammals under the expressions of lactation hormones. In the presence of α-LA, β-1,4GalT shifts the acceptor specificity from GlcNAc to Glc and allows the attachment of activated UDP-Gal to Glc to yield lactose. Thus, α-LA is the key to the presence of oligosaccharides in milk, and besides, it is also known that regulates the rate of lactose synthesis (Rudloff, 2006; Urashima *et al*., 2011).

The biosynthetic steps for extending lactose to form LNT and LNnT remain poorly understood. The hypothetical pathway is presented in figure 1.5 and suggests that this extension could be carried out by the action of *N*-acetylglucosaminyltransferases (GlcNAcT) and galactosyltransferases (GalT). They might occur as membrane-bound glycoproteins which act by sequentially addition of single activated (UDP-bound) monosaccharides (Kobata, 2003).

As reported in 1.1.2, HMO fucosylation and sialylation provide further complexity to HMO structures. They involve the addition of fucose and sialic acid sugar via α-1,2/3/4 linkages and α-2,3/6 respectively. The heterogeneity of fucosylated oligosaccharides strongly depends on the expression of specific glycosyltransferases in correlation to the mothers' Lewis blood group and secretor status, while HMO sialylation remains poorly understood.

In general, the study of the biosynthesis of HMO remains incomplete due to the high variation of amount and composition that is observed in human milk. Furthermore, animal models have limited success because the huge differences in oligosaccharide complexity compared to humans, meaning less enzymes implicated in the synthesis. None of the cell lines studied synthesised HMO or lactose, even most of them do not express α-LA enzyme (Bode, 2012).

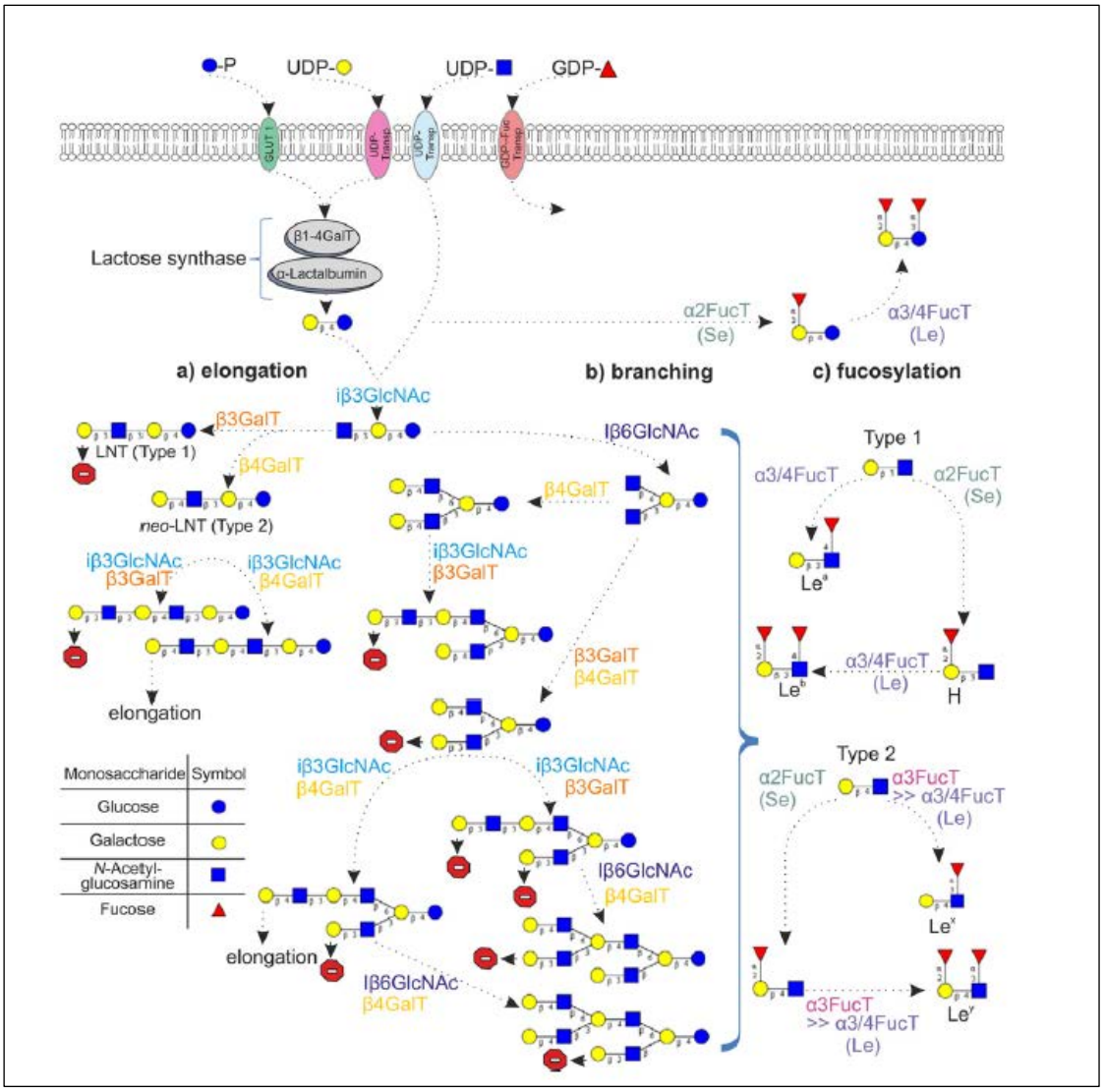


Figure 1.5 Biosynthesis of neutral complex human milk oligosaccharides (HMO).
The assumed biosynthetic pathway starts from the activated monosaccharides. *N*-acetylglucosaminyltransferases (GlcNAcT) attach *N*-acetylglucosamine (GlcNAc) in the β-1,3 position to terminal galactose (Gal), Iβ6GlcNAcT attaches GlcNAc in β-1,6 position to terminal Gal. Galactosyltransferases (GalT): b3GalT attaches Gal in the β-1,3 position to GlcNAc and b4GalT attaches Gal in the β-1,4 position to GlcNAc. Fucosyltransferases (FucT): α2FucT attaches fucose (Fuc) in the α-1,2 position to terminal Gal, secretor (Se) enzyme,α3FucT attaches Fuc in the a1–3 position to GlcNAc, α3/4FucT attaches Fuc in the α1–3/4 position to GlcNAc and in the α1–3 position to Glc of the lactose core, Lewis (Le) enzyme .The no entry signs mean that no further elongation takes place. Fucosylation is indicated exemplarily for terminal type 1 and type 2 chains. Glycan structures are depicted according to the recommendations of the Consortium of Functional Glycomics. Image from Blank *et al*., 2012.

1.2.4 PHYSIOLOGICAL PROPERTIES OF HMO

Lactose is hydrolysed to Glc and Gal by the enzymatic activity of β-1,4-galactosidase lactase (EC 3.2.1.108) which is located in the microvilli in the brush border of the small intestinal epithelial cells. The released monosaccharides are then absorbed to be used as energy source by the infant (Urashima *et al*., 2012a). In contrast, milk oligosaccharides have no nutritional value for humans. HMO tolerate the acidic conditions of the upper gastrointestinal tract, and they do not suffer from hydrolysis in the intestine owing to the lack of locally-expressed glycosidases. They reach the distal small intestine and colon in an intact form, and they are excreted with infants' feces (Engfer *et al*., 2000). However, at least some ingested HMO are absorbed intact

into the circulation and excreted in the urine. Their concentrations in these fluids correlate with levels of the corresponding mothers' milk. Although relative fractions of absorbed HMO are low, these levels have been shown to have biological effects in cell cultures, and could explain some of the postulated benefits of human milk (Goehring *et al*., 2014).

Since the very first beginning of HMO research, these oligosaccharides have been associated with their "bifidogenic" or prebiotic effect. However, since the last two decades, there is striking evidence suggesting a wide spectrum of health benefits for the neonates by these complex oligosaccharides. HMO are antiadhesive antimicrobials that serve as soluble decoy receptors, prevent pathogen attachment to infant mucosal surfaces and lower the risk for viral, bacterial and protozoan parasite infections. In addition, HMO may modulate epithelial and immune cell responses, reduce excessive mucosal leukocyte infiltration and activation, lower the risk for necrotizing enterocolitis, and provide the infant with sialic acid as a potentially essential nutrient for brain development and cognition (reviewed in Lane and Hickey, 2014).

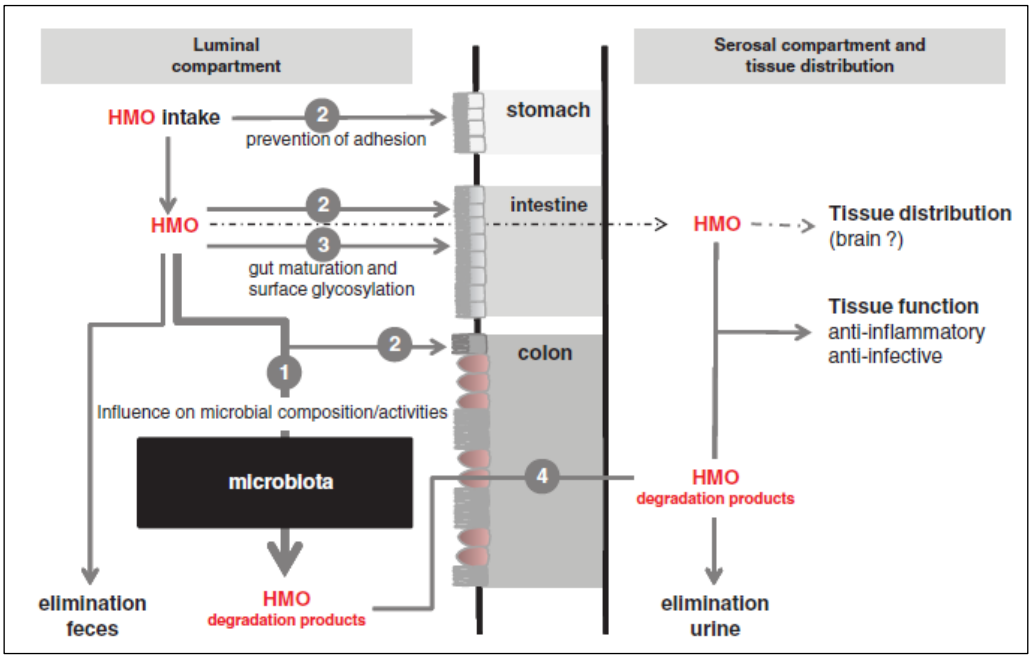


Figure 1.6 Overview of HMO metabolism and potential functions in human milk-fed infants.
The numbers 1 to 4 indicate specific functions, i.e. (1) influence on the microbiota composition and/or activity, (2) prevention of pathogen adhesion, (3) direct effects on epithelial cells and (4) systemic effects. Image from Lane and Hickey, 2014.

1.2.4.1 Prebiotic effect

Although there are some suggestions that infants could be exposed to bacteria prior to birth (Perez *et al*., 2007) is generally accepted that, in normal conditions, the human gastrointestinal tract is sterile at birth (German *et al*., 2008). Several factors can modulate the development of intestinal microflora, like the type of birth, the environmental contamination, sanitary conditions and geographical distributions. However, the type of feeding is the main factor for an optimal development of gut flora (Coppa *et al*., 2006b).
As has already been reported above, human milk is a mixture of interacting factors. In particular, the prebiotic effect has been assigned to a low concentration of proteins and

phosphate, the presence of lactoferrin, lactose and oligosaccharides. However, Table 1.5 clearly shows that oligosaccharides represent the main substances to promote the prebiotic effect.

Table 1.5 Features of milk components proposed as prebiotics.
The analysis of the literature shows that oligosaccharides undoubtedly represent the main substances able to promote a bifidobacteria-dominant microflora in infants. (Adapted from Coppa *et al*., 2006.)

Component	*In vitro* effects	Resistance to digestion	Clinical trials	Prebiotic effect
Phosphates	?	?	-	No
Proteins	±	+	-	Minor
Lactoferrin	?	++	±	Possible
Nucleotides	+	?	±	Uncertain
Lactose	+	±	-	Minor
Oligosaccharides	+++	++	+++	Documented

The most recent definition of prebiotics describes them as "selectively fermented ingredients that allow specific changes, both in the composition and/or activity in the gastrointestinal microbiota that confers benefits upon host well-being and health". This definition requires that prebiotics are resistant to gastric acidity, hydrolysis by host enzymes and gastrointestinal absorption (Roberfroid, 2007). Prebiotics stimulate the growth of beneficial bacteria including bifidobacteria, genus generally predominant in infants' microbiota. In breast-fed infants, lactobacilli and bifidobacteria represent up to 90% of total gut flora, whereas in formula-fed infants these two genera represent 40-60% of total intestinal flora (Coppa *et al*., 2006b).

Since the pioneering research of Herni Tissier and Ernst Moro in 1900, who found different bacterial composition in breast-fed and formula-fed infants' colon, and the research of György in 1954, who first showed that *B. bifidum* was uniquely able to grow on human milk glycan fractions, human milk has been thougth to have a prebiotic effect.

More recently, a large gene cluster, composed of several glycosidases and specific transporters for HMO metabolism in the genome of *B. longum* subsp. *infantis* (Sela *et al*., 2008), has been found. In fact, Marcobal and coworkers reported that *B. longum* subsp. *infantis* can use HMO as a sole source of carbon, whereas other intestinal bacteria were unable to assimilate HMO (Marcobal *et al*., 2011). The localization of HMO degrading enzymes have been identified in some but not all bifidobacterial strains. *B. infantis* consumes HMO entirely over time, including mono- and disaccharide degradation products. *B. bifidum*, in comparison, grows slower on HMO and leaves some monosaccharide degradation products. In contrast, *B. longum* subsp. *longu* and *B. breve* hardly grow on HMO, and metabolize only LNT, but not LNnT (Asakuma *et al*., 2011). The assumption of "bifidogenic" effect of HMO, is therefore incorrect, since not all bifidobacteria are able to grow using HMO.

In addition, a unique sugar metabolic pathway for bifidobacteria has been found. It confirms the ability of some species of this genus to metabolize HMO (Fushinobu, 2010; Wada *et al*., 2008a). This specific pathway is called the GNB/LNB pathway and is reviewed in more detail in

paragraph 1.3.1. The presence of external glycosidases, fucosidases, lacto-*N*-biosidases and specific transporters for LNB explains why bacteria use HMO in different ways. It is understood that the bifidobacterial strains that grow well on HMO have acquired these specific genetic adaptations for select growth on human milk glycans (Sela *et al*., 2008). In addition, supporting the prebiotic concept of HMO, certain HMO species promoted bifidobacterial growth within *in vitro* fecal enrichment assays (Yu *et al*., 2013).

With regard to acidic HMO studies, 3'-sialyllactose, as a single oligosaccharide, has been found to influenced the intestinal microbiota composition of mice (Fuhrer *et al*., 2010).

Furthermore, the growth of beneficial bacteria produces other positive biological effects in the infant. The growth of lactobacilli may inhibit the growth of harmful bacteria, stimulate immune functions and enhance thre absorption of minerals. Bifidobacteria, produce short chain fatty acids (SCFA), acetate and lactate as metabolic end products, which decrease intestinal pH, and contribute to lowering blood cholesterol and ammonia levels, and help to absorb minerals. Both strains produce group B vitamins and act as immunomodulators (Gibson and Roberfroid, 1995). In addition to the direct prebiotic effect, HMO also produced a unfavourable environment for the growth of pathogenic bacteria such us *Shigella sp*, *Escherichia coli*, *Streptococcus faecalis* and *Clostridium sp* (Bode, 2006). Moreover, the SCFA produced by bifidobacteria are absorbed in the large bowel providing nutrition for the neonate. Therefore, most of the energy contained in human milk oligosaccharides is finally available for the breast-fed infant (Miller and McVeagh, 1999).

Although there are is striking evidence of the overall prebiotic effect of HMO, more studies are needed regarding absorption, metabolism and physiological function in infants in order to decide which compound, and in which concentration or combination, it should be used for the development of a desire beneficial gut flora.

1.2.4.2 Antiadhesive antimicrobials

The adhesion of bacterial, many viral and protozoan pathogens involves a ligand-receptor interaction between structures on the pathogen surface and homologous structures on host's mucosal surface. Most commonly, the pathogen adhesion is initiated by lectin-glycan interactions (Bode, 2015; Smilowitz *et al*., 2014; Kunz *et al*., 1999).

Human-milk glycans are often part of the carbohydrate-binding determinants, suggesting that they are innate anti-adhesion agents that protect the breast-fed child by preventing pathogens from adhering to host ligands. HMO have structural homology to host cell receptors and thus can act as "receptor decoys" to prevent the adhesion of such pathogens from binding to the host cell-surface glycans (Morrow *et al*., 2005; Kobata, 2003).

The most extended data on HMO acting as antiadhesive agents has been reported with regard to *Campylobacter jejuni* infections, which are the most common cause of bacterial diarrhoea and infant mortality. *C. jejuni* binds to type 2 H-antigen exposed in the surface of epitheal cells, which is 2'-fucosyllactose structure. α-1,2-fucosylated HMO inhibits *C. jejuni* colonization of mice *in vivo* and human intestinal mucosa *ex vivo* (Ruiz-Palacios *et al*., 2003). Indeed, this antiadhesive effect has been confirmed in a prospective clinical study in which *C. jejuni*

diarrhoea occurred significantly less often in infants whose mothers' milk contain high amounts of 2'-fucosyllactose (Morrow *et al*., 2005).

Besides, some microorganisms express glycans to bind to host's lectins, instead of the vice versa process cited above. Human immunodeficiency virus (HIV) uses a glycoprotein (gp120) to bind to human dendritic cells (DC). A strong gp120-DC interaction is needed for HIV entry through mucosal barriers during HIV mother-to-child transmission via breast-feeding. HMO contain Lewis blood group antigens that compete with gp120 for binding to DC *in vitro*. Thus, a possible explanation for the inefficient transmission of HIV via breast-feeding (10-20% of the infants get infected) would be that mucosal surfaces are covered with high concentrations of HMO, and these might block HIV entry via DC (Bode, 2015 and 2012).

Other examples of *in vitro* assays to determine the antiadhesion effect of HMO have been reported. The inhibition of adhesion of *Escherichia coli, Vibrio cholera and Salmonella fyris* to Caco-2 cells, when HMO were present in the culture media, was revealed (Coppa *et al*., 2006a). Moreover, a reduction of *Entamoeba hystolitica* attachment and cytoxicity to the host's colonic mucosa in HT-29 cells was also reported (Jantscher-Krenn *et al*., 2012a).

The antiadhesive effect of HMO may not only be relevant for enteric infections. Breast-fed infants are less likely to develop otitis media caused by *Streptococcus pneumonia*, *Pseudomonas aeruginosa* or *Haemophilus influenza.* In addition, they are also at lower risk of developing respiratory syncytial virus infection (RSV). These microbial pathogens employ lectin-glycan interactions to initiate infection, and HMO have shown to *in vitro* block their attachment. Human milk covers the mucosal surfaces in the infants' nasopharyngeal regions and occasionally reaches the upper respiratory tract during episodes of aspiration allowing HMO to reach those areas. (Andersson *et al*., 1986). Similarly, HMO are absorbed and excreted with the urine, and they reduce uropathogenic *E. coli*-induced hemagglutination, suggesting that HMO also reduce urinary tract infections (Martín-Sosa *et al*., 2002).

In conclusion, it has been confirmed that some HMO are potential decoy receptors for pathogenic microorganism relevant to infections of the gastrointestinal, urogenital and respiratory tract and reduce the risk of infections (Lane and Hickey, 2014).

1.2.4.3 Modulators of intestinal epithelial cell and immune responses

Different studies have shown that HMO are absorbed intact into the circulation and excreted in the urine. Their concentrations in these fluids correlate with levels of the corresponding mothers' milk. Although the detected levels of HMO were low, these levels have shown to be enough to have biological effects *in vitro*, and could explain some of the postulated benefits of human milk (Goehring *et al*., 2014; Martín-Sosa *et al*., 2002).

It seems likely that HMO may also have systemic effects such as influencing the adhesion of leukocytes to endothelial cells, the interaction of platelets with neutrophils, or influencing systemic infectious, inflammatory and immune processes.

Recent studies have demonstrated the effects of HMO on the glycosylation pattern of epithelial cells, on cell proliferation, differentiation, and apoptosis as well as on cell signalling pathways (Lane and Hickey, 2014). In this sense, Kuntz and coworkers have confirmed that

HMO induce growth inhibition in intestinal cells through two different mechanisms. Firstly, by suppressing cell cycle progression through induction of differentiation and/or, secondly, by influencing apoptosis. As the development and maturation of digestive and absorptive processes depend on differentiation, HMO seem to be effective at influencing various stages of gastrointestinal development *in vitro* (Kuntz *et al*., 2008). Subsequent studies showed the effects of pooled HMO on cell cycle regulation, suggesting that oligosaccharides from human milk inhibited intestinal cell proliferation and altered cell cycle dynamics by affecting corresponding regulator genes and mitogen-activated protein kinase signalling (Kuntz *et al*., 2009). In an *in vitro* study with human intestinal epithelial cell line, a differential expression of glycosylation-related genes and cell surface glycome changes with 3'-sialyllactose were observed. These results suggest that 3'-sialyllactose may influence cell surface glycosylation and thereby reduce the susceptibility of pathogenic bacteria by orally administrated HMO (Angeloni, 2004). This published data strongly indicates that HMO can directly interact with infant epithelial cells, affect gene expression and reprogram the cell cycle as well as cell surface glycosylation. Whether those results are translated into *in vivo* models or, ultimately, into human babies remains to be investigated (Bode, 2015 and 2012).

With regard to immune system modulation, the involvement of intestinal epithelial cells in inflammatory processes of the gastrointestinal tract is widely recognized (Green-Johnson, 2012). The secretion of proinflammatory cytokines in response to bacterial lipopolysaccharides or dietary components is a well-known concept in on-going intestinal inflammation and its progression. Therefore, it is of interest to know to what extent HMO affects the initial step of an inflammatory process in intestinal cells by inhibiting the secretion of proinflammatory cytokines (Lane and Hickey, 2014).

The neutral HMO fraction, LNFP III and LNnT, which are two of the most abundant oligosaccharides in human milk, have been shown to influence peritoneal macrophages, capable of suppressing naive CD4+ Tcell responses. LNFP III also stimulated macrophage activity *in vitro* and increased secretion of prostaglandin E2, IL-10, and TNFα (as reviewed in Lane and Hickey, 2014).

Likewise, sialylated and fucosylated HMO influenced leukocyte infiltration and activation in an *in vitro* flow model with TNFα-activated human umbilical vein endothelial cells and isolated human leukocytes. Moreover, a reduced platelet neutrophil complex formation and neutrophil activation in the presence of sialylated and fucosylated HMO was observed in an *ex vivo* model with fresh human blood (Bode *et al*., 2004).

1.2.4.4 Protection against necrotizing enterocolitis

A special mention is required for HMO protection against necrotizing enterocolitis (NEC). NEC is one of the most common and lethal diseases in preterm infants However, it is still poorly understood. NEC happens when tissue in the small or large intestine is damaged and it is characterised by mucosal or even deeper intestinal necrosis. More than 25% of infants with NEC die, and survivors are often faced with severe neurological complications. The ethology is unknown and the treatment is limited to surgical resection of the necrotic intestine which often comes with long term complications associated with short bowel syndrome. Thus,

prevention appears to be the primary strategy for dealing with this disease (Kosloske and Musemeche, 1989).

Improper establishment of the neonatal gastrointestinal (GI) microbiota has been implicated as one of several contributing factors to necrotizing enterocolitis (NEC) (Hoeflinger *et al*., 2015). Furthermore, the risk of developing NEC is up to 10 times less frequent in breast-fed infants than in formula-fed ones. The significant differences between formula milk and human milk led to the hypothesis that HMO can protect against NEC (Lucas and Cole, 1990). There is still no confirmed evidence of this hypothesis in humans due to the lack of available HMO required for a well-controlled design. However, studies on an established rat model of NEC showed that HMO indeed protect against the disease, and it seemed that a single oligosaccharide, diasyllactose-*N*-tetraose (DSLNT, Neu5Acα-2-3-Galβ-1-3-[α-Neu5Acα-2-6]GlcNAcβ-1-3-Galβ-1-4-Glc), was responsible for NEC protection, suggesting a highly structure-specific host-receptor effect (Jantscher-Krenn *et al*., 2012b).

Further studies are needed to establish by which mechanism sialylated HMO protect infants from NEC. Needless to say, these results seem to be encouraging.

1.2.4.5 Nutrients for brain development

There are some indications that dietary carbohydrates may be important for normal brain composition. Evidence suggests that brain development and cognition in part depend on sialic acid-containing gangliosides and poly-sialic acid containing glycoproteins (reviewed in Wang, B. *et al*., 2009). Post-mortem analysis on human neonates showed that concentrations of ganglioside and proteins bound to sialic acid are significantly higher in the brains of breast-fed infants compared to infants fed with formula. Human milk is a rich source of sialic acid. The amount of sialic acid from HMO is 2 to 3 times higher than the amount from glycoproteins, and sialic acid in glycolipids accounts for only about 1% of the total sialic acid in human milk (Wang, 2009). It remains to be investigated whether sialylated HMO are the primary sialic acid carrier that provides the developing brain with this seemingly essential nutrient and in they contribute to superior developmental scores and intelligence quotients in breast-fed infants (Bode, 2012).

Lactose-derived oligosaccharides, and in particular their Gal moieties, may play a role in the development of the infant brain. Galactocerebroside, with Gal as its polar head group, is the predominant glycolipid in myelin. The liver may not be capable of providing the entire Gal needed by young mammals during this period of myelination and brain development. Taking into account that oligosaccharides are not completely excreted via feces, but are to some extent absorbed in the digestive tract, a possible role of milk oligosaccharides is to provide infants with sufficient Gal levels. This hypothesis is supported by the fact that the oligosaccharide pattern in the milk of elephants is even more complex than that of human milk. Both species show similar patterns of postnatal ontogeny; they both grow slowly, have relatively large and highly developed central nervous systems developing mainly after birth, are highly intelligent, and exhibit a high degree of learned behaviour. The degree of encephalization is considerably higher in humans and elephants than in nonhuman primates, reflecting the differences in milk oligosaccharide concentrations (Kunz *et al*., 2000).

Furthermore, with regard to neutral oligosaccharides, Vazquez and coworkers reported that a diet with 2'-fucosyllactose (2'-FL) increased the expression of a molecular brain function marker in conscious rats and enhanced memory consolidation, spatial learning and associative learning. Dietary 2'-FL exerts a positive effect on learning and memory in rodents while dietary L-fucose showed no measurable effects. Then, 2'-FL could contribute to improved cognitive abilities in breast-fed babies, although the mechanism of action by which this is achieved, has not been elucidated (Vázquez *et al*., 2015).

1.2.5 PRODUCTION OF HMO

Due to the relevant biological effects assigned to HMO, and, given that they are obviously not amenable to large scale production, there is an urgent need for alternative sources from which to obtain a diverse mixture of complex human milk-like oligosaccharides, which resemble as much as possible or even reproduce the variety of complex oligosaccharides in human milk.

1.2.5.1 Recovery from natural sources

The importance of the structure-function relationship of HMO suggests that oligosaccharides from (farmed) animal milk could be the first choice for an acceptable and available source for molecules with structures close to those of human milk oligosaccharides. Both, HMO and MMO (other mammals' milk oligosaccharides) are branched structures that contain the five structural elements: Gal, Glc, Fuc, Sia and GlcNAc. MMO have been studied over the years, but no animal milk matches the high amount and high structural diversity of human milk oligosaccharides. Oligosaccharide concentrations in the milk of most farm animals (including cows, goats, sheep and pigs) are 100-1.000 times lower than in human milk, with a lower number of different oligosaccharides, a higher abundance of sialylated and a lower abundance of fucosylated oligosaccharides (Urashima *et al*., 2001, 2012b).

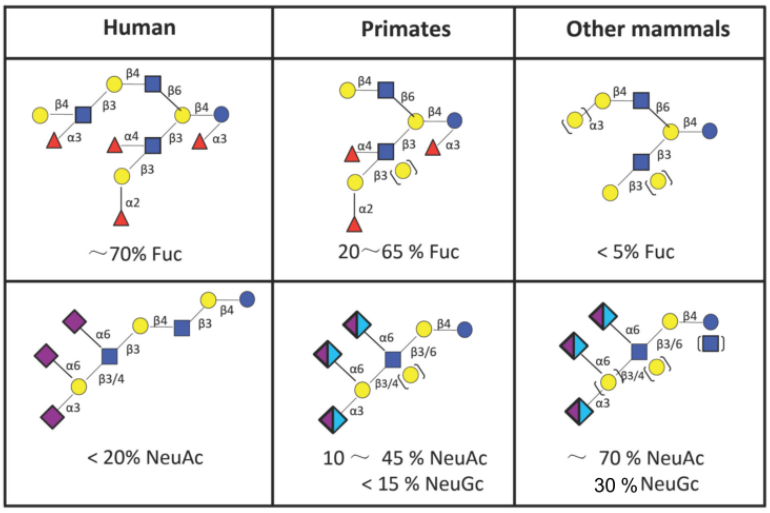


Figure 1.7 The structural features of free milk oligosaccharides among mammals.
Human milk contains large amounts of fucose and highly branched structures, with lower levels of sialic acid in comparison to primates and other mammals. Glycan structures are depicted according to the recommendations of the Consortium of Functional Glycomics (http://www.functionalglycomics.org). Blue square: GlcNAc, yellow circle: Gal, red triangle: Fuc, purple diamond: Neu5Ac and blue diamond: Neu5Gc. Image from Ruhaak and Lebrilla 2012.

Among other milks from domestic animals, goat milk is especially rich in lactose-derived oligosaccharides. Interestingly, it is found that goat milk oligosaccharides (GMO) are present in higher amounts than bovine milk oligosaccharides (BMO), and contain fucosylated species. Goat milk contains a large variety of acidic and galactosyllactose together with *N*-acetylglucosaminyllactose as the major neutral oligosaccharides. The similarities found between human and goats' milk oligosaccharides in terms of the presence of neutral (galactosyl-lactose and lacto-*N*-hexaose) and sialylated (3-, 6-sialyl-lactose and disialyl-lactose) structures suggest that GMO could carry out some of the physiological activities described for HMO (Martinez-Ferez *et al*., 2006).

Until now, among farm animals, there is not one type of milk which has been found to contain the proportion of oligosaccharides resembling that found in human milk, both in quantity and composition. The development of analytical methodologies demonstrates that milk from farmed animals contains a much larger variety of complex oligosaccharides than previously thought. Therefore, farmed animal milk is a highly attractive source of bioactive oligosaccharides, which, once extracted, can be potentially used in medical and functional foods. To justify their use, additional studies on their biological functions are required. The advanced knowledge on their structural composition will certainly help to gain new insights into the complex biochemical pathways underlying their potential health promoting properties (Albrecht *et al*., 2014).

1.2.5.2 Chemical synthesis

The chemical synthesis of long chains of oligosaccharides remains challenging. Conventional chemical methods of synthesis of these compounds require many protection and deprotection steps due to the similar reactivity of the hydroxyl groups of monomer units as well as the requirements for the stereo- and regio-selective control of glycosidic bond formation. Accordingly, there are a lot of synthetic steps which render low productivity and, in some cases, result in the presence of some harmful chemicals in the final product used within the process (Wang and Davis, 2013).

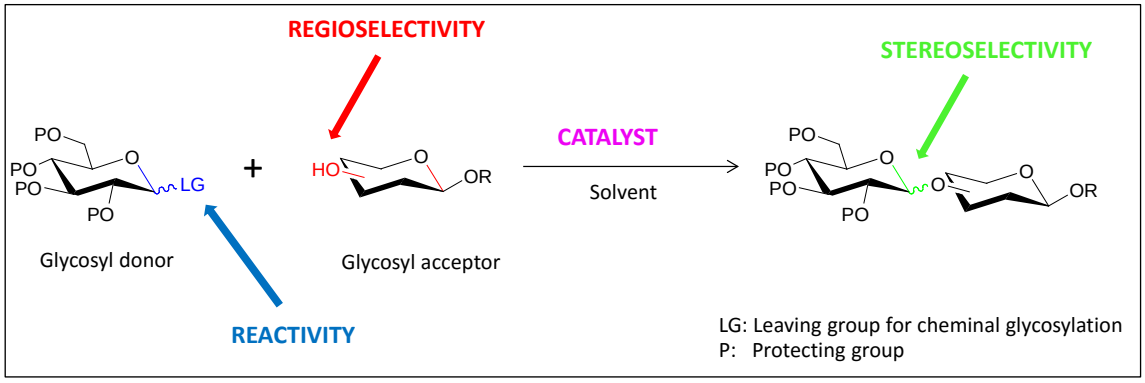


Figure 1.8 General issues in chemically glycoside synthesis.
Image adapted from Wang and Davis, 2013.

A small number of HMO have recently been chemically synthesized and are registered in different patents. See Table 1.6 for detailed information.

Table 1.6 Patented synthesis methods for HMO.

Chemically synthesised HMO	Reference
Lacto-*N*-difucohexaose I (LNDFH I)	WO2013185780 A1
2'-Fucosyllactose (2'FL)	WO 2010/115935
Lacto-*N*-neotetraose (LnNT)	WO 2011/100980
6'-Sialyllactose (6'SL)	WO 2011/100979
2'-FL polymorphs	WO 2011/150939

It is noteworthy that the development in 2001 of automated solid-phase oligosaccharide synthesis (Plante, 2001) enables the fast and efficient production of HMO like sugars. In solid-phase synthesis, the synthesised oligosaccharides are linked to an insoluble material (beads or resins) that allows the fast separation of reaction products from excess reagents, soluble reaction by-products, and solvents. For example, the synthesis of complex oligosaccharides, such as Lewis^y hexasaccharide and Le^y-Le^x nonasaccharide (which are tumor markers that are currently being evaluated as cancer vaccines), took more than 1 year using conventional synthesis. Applying solid-phase methodology, the synthesis was accomplished in 1 day (Castagner and Seeberger, 2007). This methodology opens the gate to the chemical synthesis of some complex HMO. However, achieved yields by solid-phase synthesis are in the mg scale, very far from industrial purposes.

More recently, in 2014, a one-pot procedure to synthesise LNFP I and its α-1,2-fucosylated LNnT analog in high yields was reported. A galactose thioglycoside was coupled chemoselectively to a phthalimide-protected glucosamine thioglycoside. The resulting disaccharide was then coupled to the 3-position in lactose. However, it was not possible to extend the one-pot process to the synthesis of the corresponding 3-linked structures (Arboe Jennum *et al*., 2014).

In conclusion, chemical synthesis has reported the production of 15 different HMOS and derivatives with 3–11 monosaccharide units (Chen, 2015). However, synthesised HMO according to methods mentioned above, may still be contaminated with small residues, organic or protic solvents used in the procedure. In order to use the HMO in nutritional and therapeutic formulations for mammals (especially for humans but particularly for infants), it would be necessary to substantially reduce such residual contaminants (Andreas *et al*., 2013).

1.2.5.3 Enzymatic synthesis

Enzymatic synthesis is an alternative strategy that overcomes synthetic limitations since enzymes offer a perfect control of stereo- and regio-selectivity and usually high efficiency in transformations for glycan assembly in aqueous solution without the need for protecting groups (Faijes and Planas, 2007; Moracci *et al*., 2001; Wang and Davis, 2013).

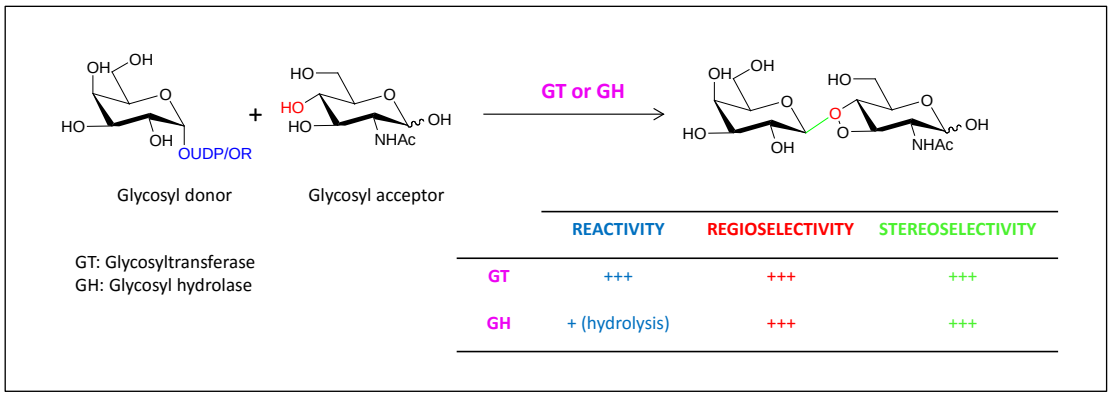


Figure 1.9 General issues in enzymatic glycoside synthesis.
UDP-Gal as donor for glycosyltransferase synthesis, and OR-Gal for glycosidase synthesis. Enzymatic synthesis of carbohydrates offers high regio and stereoselectivity. Image adapted from Wang and Davis, 2013.

Furthermore, the application of enzymes enables the utilization of side-products from the food industry which would provide a double benefit. It generates a product and, at the same time, avoids the disposal of biomass. The use of the enzymes in the production process would ensure that no harsh chemicals are produced and this would ensure safer products for infants.

In nature, condensation reaction can be provided by two types of enzymes: glycosyltransferases and glycosidases. Glycosyltransferase enzymes are responsible for glycoside bond formation in nature, catalysing the transfer of a monosaccharide from a sugar nucleotide donor (in Leloir type) to an acceptor. However, the use of glycosyltransferases for production purposes is still limited since the enzymes are difficult to solubilize and express and also the nucleotide donor is expensive. In contrast, glycosidase enzymes are normally hydrolytic enzymes. However, they can also perform reversal reaction, leading to glycosidic bond formation in specific conditions. Oligosaccharide synthesis may be achieved either by displacing the equilibrium towards glycosidic bond formation (thermodynamically controlled synthesis) or by using activated glycosyl donors (kinetically controlled transglycosylation). In any case, these methodologies only achieve yields of up to 40-50% (Faijes and Planas, 2007; Moracci *et al*., 2001; Wang and Davis, 2013; Wang and Huang, 2009).

Regarding enzymatic synthesis, several HMO have been obtained and is explained as follows.

LNB

LNB production from sucrose and GlcNAc was developed via one-pot enzymatic reaction on an industrial scale and cost-effective process. Four kinds of enzymes, including sucrose phosphorylase (SP) from *Bifidobacterium longum*, UDP-glucose-hexose-1-phosphate uridylyltransferase (GalT), UDP-glucose-4-epimerase (GalE), and lacto-*N*-biose phosphorylase (LNBP) from *B. bifidum* were applied to the LNB production to yield kilograms of LNB with an approximate 83% reaction yield. The overall reaction is described in Figure 1.10.

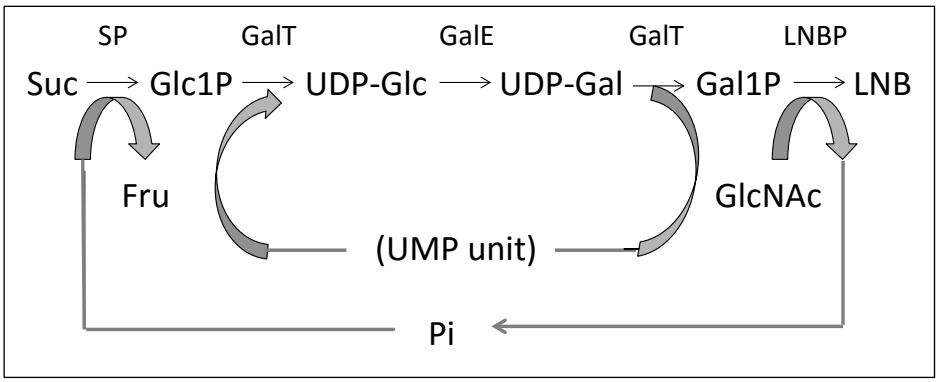


Figure 1.10 Reaction scheme for LNB production.
SP: Sucrose phosphorylase from *Bifidobacterium longum*, GalT: UDP-glucose-hexose-1-phosphate uridylyltransferase, GalE: UDP-glucose-4-epimerase, LNBP: lacto-*N*-biose phosphorylase, Suc: sucrose, Glc1P: α-glucose 1-phosphate, Fru: fructose, LNB: Lacto-*N*-biose. Image from Nishimoto and Kitaoka, 2007.

Sucrose is phosphorolysed into α-glucose 1-phosphate (Glc1P) and fructose by SP. Glc1P is converted into UDP-Glc concomitantly with the conversion of UDP-Gal into Gal1P by GalT. UDP-Glc is converted into UDP-Gal by GalE, and the resulting UDP-Gal is consumed in the reaction of GalT. Finally, LNB is synthesized from Gal1P and GlcNAc by LNBP. Phosphate and UDP-Glc are recycled during the reaction system (Nishimoto and Kitaoka, 2007).

In the same way, GNB was easily produced using a one-pot enzymatic reaction, but in this case, the starting material was sucrose and GalNAc (Nishimoto and Kitaoka, 2009).

As an alternative process, LNB was synthesized from *p*-NP-β-Gal and GlcNAc by a β-galactosidase from *Bacillus circulans*, β-Gal-3. The transglycosylation was achieved in mg scale in the presence of 1-butyl-3-methylimidazolium hexafluorophosphate ([Bmim][PF6]), an ionic solvent with a 90% of yield (Bayón *et al*., 2013). This methodology was used in this project for the production of LNB, and it is further explained in chapter 2.

LNT

The enzymatic synthesis of lacto-*N*-tetraose (type 1) was first described by Murata *et al*., 1999 (Figure 1.11). It consisted of consecutive additions of *N*-acetylglucosamine and galactose to lactose. Firstly, the β-1,3-*N*-acetylglucosaminyltransferase from bovine serum transferred *N*-acetylglucosamine from UDP-GlcNAc to lactose to obtain the trisaccharide lacto-N-triose. In the next step, the β-galactosidase from *B. circulans* ATCC31382 transglycosylated *o*-nitrophenyl galactoside to the trisaccharide to achieve the lacto-*N*-tetraose derivative (regiosomer β-1,3) and the regiosomer β-1,6 with an overall yield of 19% (Murata *et al*., 1999).

In addition, Murata and coworkers assayed the transglycosylation reaction with partially purified lacto-*N*-biosidase from *Aureobacterium sp.* strain L-101. This enzyme was able to condense *p*-nitrophenyl-lacto-*N*-biose as the donor and lactose as the acceptor yielding the lacto-*N*-tetraose with a 3.7% yield. Some years later, Wada and coworkers demonstrated that LnbB from *Bifidobacterium bifidum* was also capable of preforming the same transglycosylation reaction. The catalytic efficacy was also very low, <4% of yield. However, these results suggest that LnbB is a good candidate for enzymatic synthesis of LNT (Murata *et al*., 1999; Wada *et al*., 2008a).

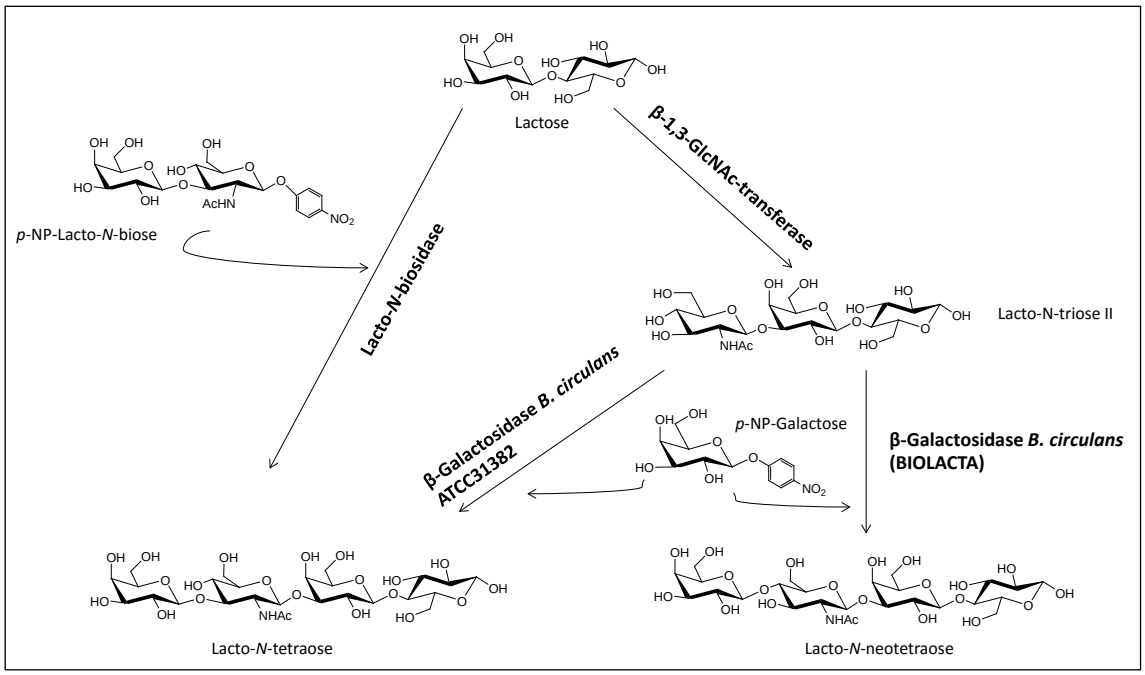


Figure 1.11 Enzymatic conversion of lactose into LNT and LNnT.
Image adapted from Murata *et al*., 1999.

LNT-II and LNnT

Recently, Lacto-*N*-triose II (LNT-II) and Lacto-*N*-neotetraose (LNnT) were successfully synthesised in large quantities using sequential "onepot" multienzyme (OPME) systems (Figure 1.12). In these systems, free monosaccharides were added one by one to each of the "one-pot" systems containing multiple enzymes responsible for catalyzing monosaccharide activation, followed by transfer processes. The overall yields of both HMO were 95% (1.36 g) for LNT-II, and 92% (1.16 g) for LNnT (Yu *et al*., 2014).

LNT-II and LNnT were synthesized from the inexpensive disaccharide lactose and the monosaccharide GlcNAc by using a one-pot four-enzyme GlcNAc activation and transfer system containing *Bifidobacterium longum N*-acetylhexosamine-1-kinase (NahK), *Pasteurella multocida N*-acetylglucosamine uridyltransferase (PmGlmU), *Pasteurella multocida* inorganic pyrophosphatase (PmPpA), and *Neisseria meningitidis* β-1,3-*N*-acetylglucosaminyltransferase (NmLgtA) (Figure 1.12).

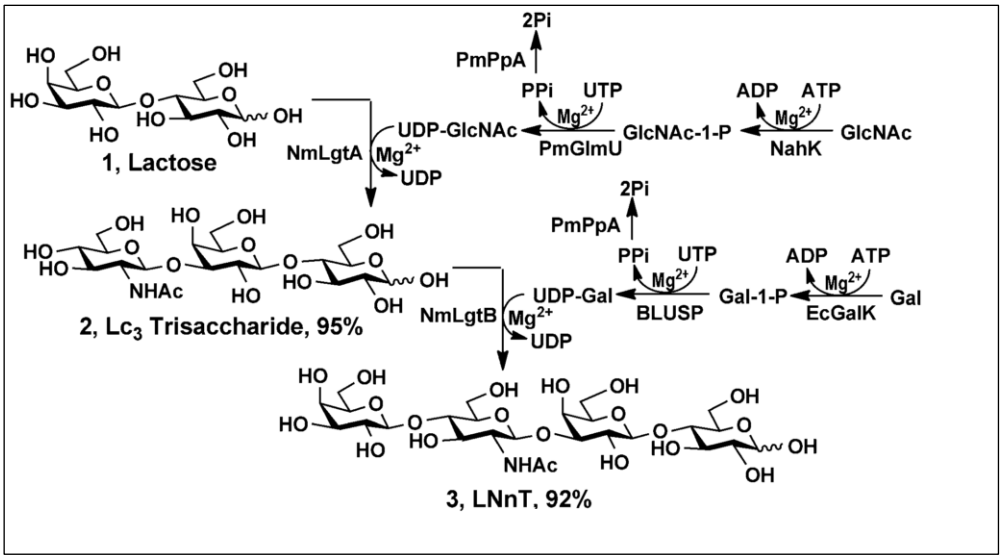


Figure 1.12 Sequential one-pot multienzyme (OPME) synthesis of Lacto-*N*-triose (LNT-II) and lacto-*N*-neotetraose (LNnT).
Image from Yu *et al*., 2014.

Sialylated HMO

For the enzymatic synthesis of sialyllactose, sequential enzymatic reactions and a nucleotides recycle system were performed with substrates in a one-pot reaction (Figure 1.13). The process started with the production of NeuAc from GlcNAc, GlcNAc 2-epimerase and Neu5Ac aldolase. GlcNAc 2-epimerase converts GlcNAc to *N*-acetyl-D-mannosamine (ManNAc), which then reacts with pyruvate to form Neu5Ac through the enzymatic action of Neu5Ac aldolase. To date, mammalian GlcNAc 2- epimerase and mammalian and *E. coli* Neu5Ac aldolases have been used to produce Neu5Ac (Li *et al*., 2008).

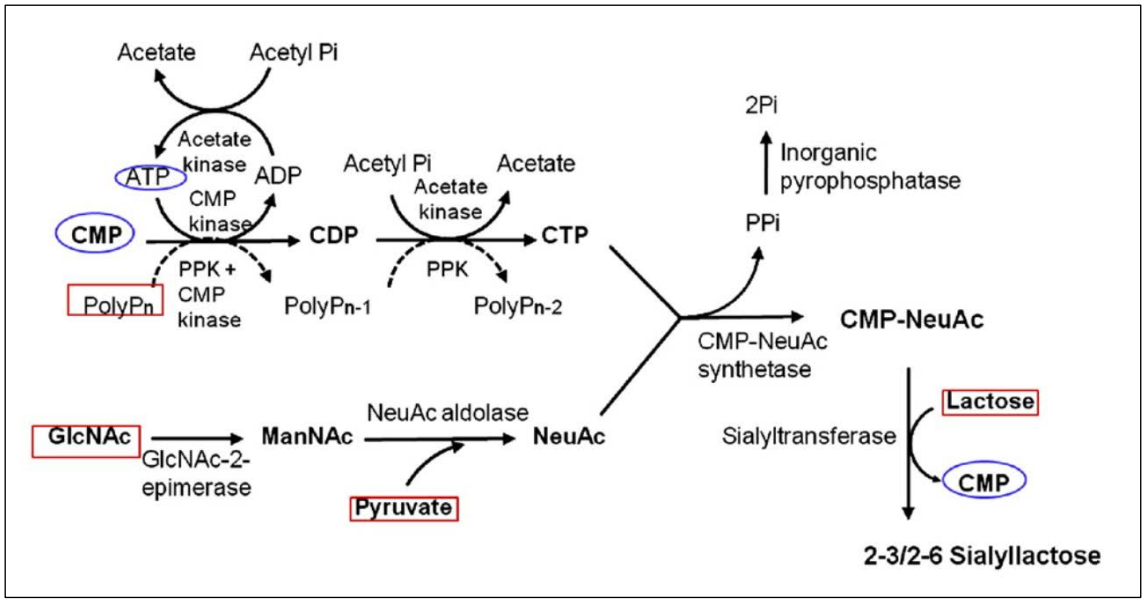


Figure 1.13 Scheme for the coupled enzymatic synthesis of sialyllactose.
Substrate molecules are in the box. The oval frame represents the intermediate molecules to be recycled. Discontinuing arrows represent the alternative cytidine triphosphate (CTP) recycle pathway. PPK, polyphosphate kinase; PolyPn, polyphosphate. Image from Han *et al*., 2012.

Then NeuAc was converted to CMP-NeuAc *via* the cytidine triphosphate (CTP) regeneration system using acetate kinase, cytidine monophosphate (CMP) kinase, and polyphosphate kinase enzymes. CTP was synthesized from an inexpensive substrate CMP in a high yield of more than 97% via cytidine diphosphate (CDP) by CMP-kinase and acetate kinase (Woo *et al*., 2014).

A fusion protein of NmCSS and *Neisseria meningitides* α-2,3-sialyltransferase (Nmα2–3ST) was used in a sugar nucleotide regeneration reaction for the synthesis of 3'-SL (68 g in a partially purified solid form, 68% yield) from lactose, Neu5Ac, phosphoenolpyruvate, and catalytic amounts of ATP and CMP (Sallomons *et al*., 2013).

Both 6'SL and 6'SLN were synthesized from CMP-Neu5Ac and lactose using a *Pasteurella dagmatis* α-2,3-sialyltransferase double mutant that was completely switched to an α-2,6-sialyltransferase (Schmölzer *et al*., 2015). Although they are promising results, this approach is still in laboratory scale evaluation (72% yield, 0.72 mM).

A sequential 2-step synthesis of sialyl T-antigens and derivatives, including α-2,3-sialylated LNB, was also reported. During the first step, galactose was enzymatically converted to LNB via a one-pot reaction with galactokinase and LNB phosphorylase. The resulting LNB was successively transformed to an α-2,3-sialylated LNB using a one-pot 3-enzyme sialylation process employing *Pasteurella multocida* sialic acid aldolase, *N. meningitidis* CMP-sialic acid synthetase, and *P. multocida* α-2,3-sialyltransferase. Yields were above 70% in all cases for semi-preparative scale production (< 100 mg in each case) (Lau *et al*., 2011).

Fucosylated HMO

Classical enzymatic fucosylation of sugars requires GDP-L-fucose as a substrate donor and a fucosyltransferase enzyme. Various fucosyltransferases have been identified in mammals, most of which are located in the Golgi apparatus, with α-1,2, α-1,3/4 and α-1,6 activity, and in different bacteria with α-1,2, α-1,3/4 activity (Han *et al*., 2012).

The enzymatic synthesis of 2'-fucosyllactose was examined using α-1,2-fucosyltransferase from *H. pylori* (FucT2), GDP-L-fucose and lactose as the starting materials to trisaccharide with 65% yield at mg scale (Albermann *et al*., 2001).

Synthesis of the lacto-*N*-difucohexaose I using GDP-β-L-fucose and two fucosyltransferases has been reported with an overall yield of 6% at mg scale (Miyazaki *et al*., 2010). Starting from LNT, produced using Murata and coworkers methodology, human fucosyltransferase I (FUT1) transferred Fuc from GDP-Fuc to the Gal residue of LNT with an α-1,2 linkage. The resulting LNFP-I was subsequently reacted with GDP-β-L-fucose and commercial fucosyltransferase III (FUT3) to generate Lacto-N-difucohexaose I (LNDFH-I) (Figure 1.14).

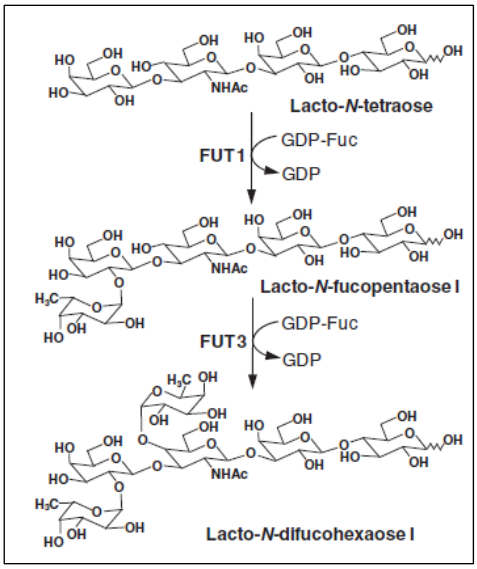


Figure 1.14 Enzymatic synthesis of lacto-N-difucohexaose I.
Human fucosyltransferase I (FUT1) transferred Fuc from GDP-Fuc to the Gal residue of LNT with an α-1,2 linkage. The resulting LNFP-I was subsequently reacted with GDP-β-L-fucose and commercial fucosyltransferase III (FUT3) to generate Lacto-N-difucohexaose I (LNDFH-I). Image from (Han *et al*., 2012).

As is cited in detailed in Chapter 4, one of the most striking advances in glycosynthesis was the development of the glycosynthase technology in 1998 (Mackenzie *et al*., 1998; Malet and Planas, 1998). Glycosynthases are engineered glycosidases devoid of hydrolase activity but are able to efficiently catalyse the formation of glycosidic bonds in high yields when using an activated glycosyl donor. Glycosynthase methodology, then, is postulated as a powerful enzymatic tool for the synthesis of oligo- and polysaccharides (Planas *et al*., 2015; Armstrong and Withers, 2013). According to glycosidases mechanism, glycosynthases can be obtained for retaining and inverting mechanisms.

Concerning HMO synthesis, to date, the unique case of glycosynthase technology strategy has been reported by Wada and coworkers for the production of 2'-fucosyllactose. An inverting 1,2-α-L-fucosidase from *B. bifidum* was converted to a 1,2-α-L-fucosynthase, redesigning its hydrolytic activity to the synthesis of 2'-fucosyllactose from a β-L-fucosyl fluoride donor and lactose acceptor. Even though these approaches opened the way to easily produce α-1,2-fucosyl linkages on galactose residues at the non-reducing ends of sugar chains, the chemical instability of these beta-fucosyl-fluoride donor molecules needs to be solved for an enhanced conversion yield (Wada *et al*., 2008b).

In spite of the reported data, to date, the developed enzymatic methods for producing fucosylated oligosaccharides have obtained a very limited number and quantity of fucooligosaccharides.

1.2.5.4 Prokaryotic biotechnology approach

Recently, promising new strategies for oligosaccharides production have been developed using metabolically engineered bacteria. These methods are based on the utilization of recombinant whole cells that overexpress the glycosyltransferase genes that are naturally involved in the

synthesis of oligosaccharides from sugar-nucleotides and on systems that allow a sugar-nucleotide recycling by these whole cells.

A coupled microbial method was first developed by the Kyowa Hakko Kogyo Company. A *Corynebacterium ammoniagenes* strain was engineered to produce and regenerate UTP from inexpensive substrate by coupling this strain with two recombinant *Escherichia coli*. One of the strains was used to overexpress the genes of sugar nucleotides biosynthesis and the other strain to overexpress a glycosyltransferase gene. The Kyowa researchers were able to produce different oligosaccharides, such as globotriose, lactosamine, and sialyllactose in very high yields (Endo and Koizumi, 2000). The whole cell-based methodology takes advantage of the metabolic machinery of the microorganism used. However, the design of the reactions are limited by the capability of those molecules to get transported into the cell or synthesized *in vivo*. To facilitate passive circulation of the substrates, the method described above requires a permeabilization of the bacteria with xylene (reviewed in Han *et al*., 2012; Priem *et al*., 2002).

Metabolic engineering overcomes coupled microbial method concerns. To date, the oligosaccharides are produced in single growing bacteria, usually *E. coli* that overexpresses the enzyme or enzymes of interest. Up to now, the range of oligosaccharides that could be produced by this *in vivo* method is limited because most of the glycosyltransferases of interest require specific acceptors that are not naturally present in the bacterial cytoplasm.

Small oligosaccharides have been produced by whole-cell systems. However, there are some important bottlenecks in producing HMO at high rates in cells. On the one hand is the massive intracellular enrichment of the produced oligosaccharides and nucleotide by-products, and on the other hand is the extraction of the produced oligosaccharides. Due to intracellular enrichment, the products of the synthesis reaction may gradually develop a product-inhibitory effect on the synthetic enzymes. The purification of the target oligosaccharide from this complex mixture is costly, and thus for most oligosaccharides, not economical. Under these circumstances, biotechnical oligosaccharide production is not efficient enough and hard to control, especially since known biotechnological oligosaccharide production is performed using batch culturing, which, from an economical point of view, is highly unsatisfactory.

Despite these difficulties, several HMO have been obtained and is explained as follows.

LNT-II and LNnT

Priem and coworkers reported the same strategy of using living *E. coli* JM109 cells ($lacY^+$ $lacZ^-$) engineered to overexpress *Neisseria meningitidis* β-1,3-*N*-acetylglucosaminyltransferase (NmLgtA) for large-scale production of LNT-II (6 g/L, 73% yield). In the case of LNnT production (5g/L) *E. coli* cells co-expressing NmLgtA and *Neisseria meningitidis* β-1,4GalT (NmLgtB) were used. In both cases, lactose was transported into the cell by the β-galactoside permease LacY. Intracellular lactose, which could not be degraded because the cell is devoid of β-galactosidase activity, was glycosylated into LNT-II by LgtA using the endogenous pool of UDP-*N*-acetylglucosamine. Glycerol provides the carbon and energy source required for bacterial growth, lactose transport, and LNT-II and LNnT synthesis (Figure 1.15) and LNnT synthesis (Priem *et al*., 2002).

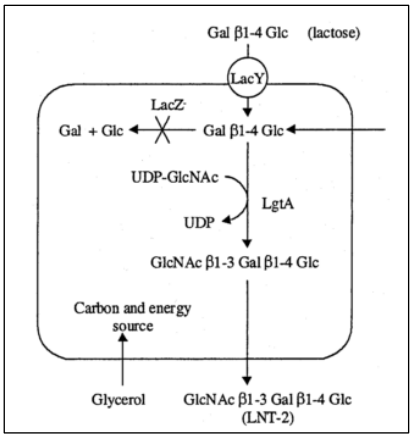


Figure 1.15 Production scheme of the trisaccharide LNT-II by *E. coli*.
Cells expressing the *N. meningitidis* lgtA gene that encodes a β-1,3-*N*-acetylglucosaminyltransferase. Image from Priem *et al*., 2002.

LNT

Large-scale production of LNT in whole-cell was achieved in 2014. *E. coli* strain LJ110 ($lacY^+$ $lacZ^-$) chromosomally integrated with *Neisseria meningitidis* β-1,3-*N*-acetylglucosaminyltransferase (NmlgtA) and *Escherichia coli* O55:H7 β-1,3-galactosyltransferase (wbgO) genes were used. When glucose was used as the carbon source, LNT-II was the major product and only about 5% of the lactose was converted to LNT (See Figure 1.16). Nevertheless, by substituting the glucose with galactose, the yield of LNT production was improved 3.6-fold (Baumgärtner *et al*., 2015). Further studies revealed that fed-batch cultivation with galactose further improved the efficiency and produced 12.72 g/L of LNT (Baumgärtner *et al*., 2014).

Fucosylated HMO

Several fucose-containing lactooligosaccharides have been produced by single cell method. With regard to LNT and LNnT synthesis, the aditional expression of different fucosyltransferases (FutA, FutB, FutC and FutT14) led to different fucosylated lacto-*N*-neotetraoses.

Following LNnT synthesis, the use of *Helicobacter pillory* α-1,3-fucosyltransferase (FutA) produced Lacto-*N*-neodifucohexaose-II at a concentration of 1.7 g/L. The same fermentation driven by another α-1,3-fucosyltransferase (FutB) led to equal amounts of both lacto-*N*-neofucopentaose-V and lacto-*N*-neofucopentaose-II produced at 280 and 260 mg/L, respectively (Dumon *et al*., 2004; Priem *et al*., 2002). Sequential feeding of glucose and lactose in a fed-batch fermentation, using *H. pylori* α-1,2-Fucosyltransferase (FutC) produced 3 g/L of oligosaccharide mixture which was composed of 57%2'-fucosyllactose and 23% lacto-*N*-fucopentaose I (LNFP I). Moreover, the same *E. coli* strain just expressing FutC (without the lgtA and lgtB expression) and with the addition of more lactose, produced 11 g/L of 2'-fucosyllactose (Drouillard *et al*., 2006).

And following LNT synthesis, the additional expression of a recombinant GDP-L-fucose salvage pathway together with a bacterial fucosyltransferase, lacto-*N*-tetraose is further converted into the respective fucosylated compounds (Figure 1.16). The expression of the α-1,2-fucosyltransferase (FutC) in the lacto-*N*-tetraose producing *E. coli* strain led to the formation of

lacto-*N*-fucopentaose I (271.6 mg/L). While the expression of the α-1,4- fucosyltransferase (FucT14) mainly led to the conversion of lacto-*N*-difucohexaose II (133.7 mg/L). The salvage synthetic method was developed using L-fucose for fucosyllactose production. *E. coli* was engineered to overexpress *E. coli* K12 L-fucose transporter (FucP) and *Bacteroides fragilis* fucose kinase/fucose-1-phosphate guanylyltransferase (Fkp) under the control of the T7 promoter. To block the L-fucose metabolism for cell growth, its chromosome was more modified by disruption of the fucA gene encoding fuculose-1-phosphate aldolase. This recombinant *E. coli* strain was able to convert L-fucose to 369 μM GDP-fucose (Baumgärtner *et al*., 2015).

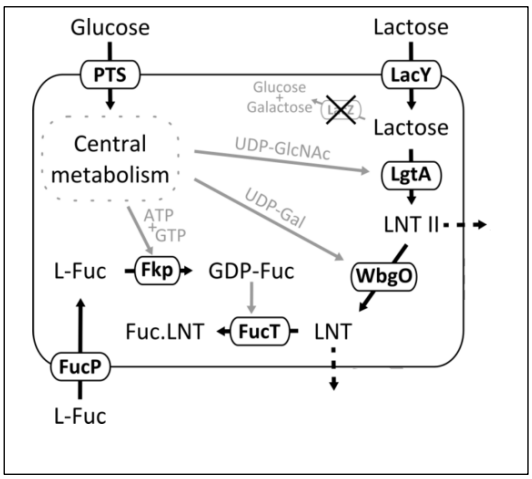


Figure 1.16 *In vivo* synthesis of fucosylated lacto-*N*-tetraoses using Leloir glycosyltransferases and intracellularly generated nucleotide-activated sugars.
Abbreviations are as follows: ATP; adenosine triphosphate, Fuc.LNT; fucosylated lacto-N-tetraose, GDP; guanosine diphosphate, GDP-Fuc; guanosine 5'-diphospho-β-L-fucose, GTP; guanosine triphosphate, L-Fuc; L-fucose, LNT II; lacto-N-triose II, LNT; lacto-N-tetraose, UDP-Gal; uridine 5'-diphosphogalactose, UDP-GlcNAc; uridine 5'-diphospho-N- acetylglucosamine, Fkp; bifunctional L-fucose kinase / L-fucose-1-phosphate guanylyltransferase, FucP; L-fucose permease, FucT; α-1,2- or α-1,4-fucosyltransferase, LacY; lactose permease, LacZ; β- galactosidase, LgtA; β-1,3-*N*-acetylglucosaminyltransferase, PTS; phosphotransferase system, WbgO; β-1,3-galactosyltransferase. Image from Baumgärtner *et al*., 2015.

Sialylated HMO

Sialyllactose or complex sialylated oligosaccharides can be produced in living *E. coli* cells either with synthesis of NeuAc from glucose via the de novo synthetic pathway (Priem *et al*., 2002) or with its uptake from medium via the salvage method (Figure 1.17) (Fierfort and Samain, 2008) .

For the salvage synthesis method, Neu5Ac was transported into the cell by NanT permease and it was protected from degradation because of aldolase (nanA) inactivation (Drouillard *et al*., 2010). And for the *the novo* synthesis, the pathway for CMP-Neu5Ac biosynthesis had to be introduced into *E. coli* strain K12 derivatives, since the *E. coli* strain which is able to produce CMP-Neu5Ac in nature is a pathogenic strain K1, and consequently, unsuitable for biotechnological use. In both systems, lactose was internalized by LacY permease and sialylated by a glycosyltransferase, using CMP-Neu5Ac, which was constantly regenerated by the enzymatic machinery of the living cells (Priem *et al*., 2002). To accomplish the synthesis of 3'-sialyllactose, the α-2,3-sialyltransferase gene from Neisseriameningitidis was coexpressed with *Campylobacter jejuni* genes encoding *N*-acetylglucosamine-6-phosphate-epimerase (neuA), sialic acid synthase (neuB) and CMP-Neu5Ac synthetase (neuC). Simultaneously,

Neu5Ac aldolase, ManNAc kinase, and β-galactosidase activities were eliminated to ensure a higher yield of sialyllactose. Sialyllactose concentration of 25 g/L was obtained with continuous lactose feed in the high cell density culture of the recombinant *E. coli* cells (Fierfort and Samain, 2008)

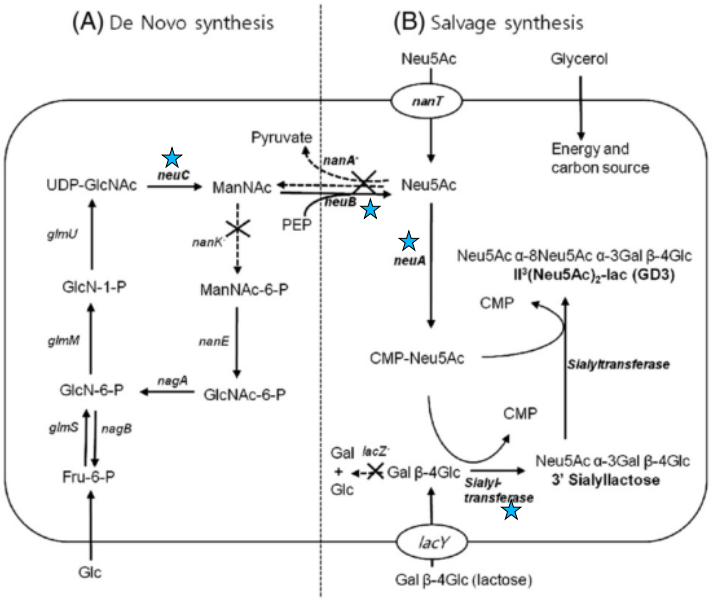


Figure 1.17 Engineered metabolic pathways for the production of sialyllactose.
From an endogenous synthesis of NeuAc from glucose (A) or with its uptake from medium (B). Overexpressed genes are marker with a blue star. Discontinued arrows represent the enzymatic activities that have been eliminated. Glycerol provides the carbon and energy source required for bacterial growth, lactose transport, and sialyllactose synthesis. Image from Han *et al.*, 2012.

This system was further extended with the addition of the appropriate glycosyltransferase genes to produce gangliosides GD3, GM2 and GM1 (reviewed in Han *et al.*, 2012).

To date, a limited number of HMO have been synthesized by whole-cell synthesis and engineered *E. coli* living-cell factory approaches. Despite of some limitations of the technology, the use of genetically modified organisms (GMO) is regulated by the Novel Food Regulation of the European Comission Regulation (EC) No 1829/2003 on genetically modified food and feed. This regulation does not cover the use of GMO derivatives in infant formula (European Commission, 2013), and to date, there is no infant food with the GMO special label in Europe.

1.2.5.5 Eukaryotic cell based assays

Eukaryotic cell based assays are being developed for understanding HMO biosynthesis rather than HMO production.

To date, immortalized or transformed human mammary gland epithelial cells have been studied with limited success. None of the cell lines studied synthesized HMO, not even lactose. Most of them did not even express α-lactalbumin, essential for lactose synthesis (Bode, 2012).

1.2.5.6 Transgenic animal production

Some research has been done with transgenic animals regarding the role of HMO in the health of infants rather than HMO production.

Prieto *et al*., achieved the production of healthy transgenic animals and healthy progeny expressing transgene-encoded glycosyltransferases. The transgenic mice containing cDNA encoding the human alpha 1,2-fucosyltransferase, synthesised 2'-fucosyllactose. Nevertheless, transgenic rabbits generated using the same fusion gene used to produce transgenic mice did not produce enough milk suggesting that the human alpha 1,2-fucosyltransferase arrested the lactation process and apparently modulated the differentiation of the mammary gland (Prieto, 2012).

1.2.6 ALTERNATIVES TO HMO

As previously cited, different methodologies have been developed for the synthesis of HMO which have provided single structures in sufficient quantity to perform clinical studies for formula supplemented with those oligosaccharides. However, there is no available methodology for large scale production of HMO at reasonable prize to introduce them in commercial formula for infants. Furthermore, the human mammary gland produces not only one or two HMO, but more than one hundred different oligosaccharides, so it is needed a mix of structurally more complex oligosaccharides composition to confer the wide spectrum of health benefits given by HMO for the formula-fed infants.

Due to the difficulties in the production of HMO, much simpler structures, such as fructo-oligosaccharides (FOS) and galacto-oligosaccharides (GOS) are introduced as supplements in infant formula (Sela and Mills, 2010).

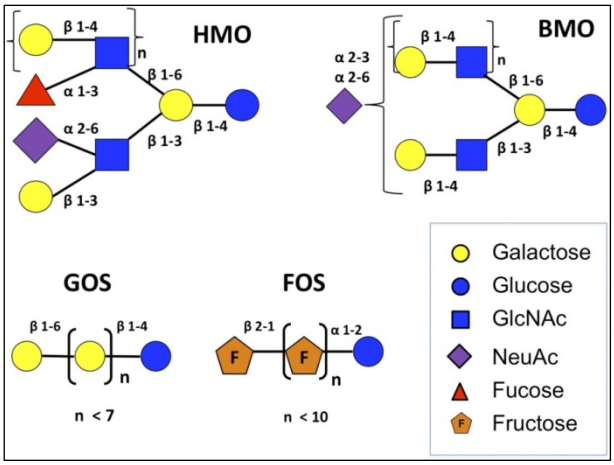


Figure 1.18 Structural comparison of different oligosaccharides.
Galactooligosaccharides (GOS), inulin, human milk oligosaccharides (HMO) and bovine milk oligosaccharides (BMO). Whereas FOS and GOS are linear chains, HMO and BMO are branched structures and contain the structural elements fucose, sialic acid, and N-acetylglucosamine that are completely absent in FOS and GOS. Furthermore, they do not exhibit the inherent structural complexity that is characteristic of HMO and BMO that results from the large number of possible β and α glycosidic linkages present in these oligosaccharides. Glycan structures are depicted according to the recommendations of the Consortium of Functional Glycomics (http://www.functionalglycomics.org). Image adapted from Zivcokic *et al*., 2011.

FOS are composed of glucose at their reducing end bound to repetitive units of fructosyl residues in β 2-1 or β 2-6 linkages. They derived from inulin commonly extracted from chicory or other plant sources through hydrolysis by an endoinulinase to form oligomers of a degree of polymerization of 5 . As an alternative, FOS can be also enzymatically synthesised from sucrose (Boehm *et al*., 2005). GOS, composed of β 1-3/4/6 linked oligo-galactose with lactose at their reducing end, are produced from lactose by β-galactosidase (Intanon *et al*., 2014). It is important to note that FOS and GOS do not naturally occur in human milk. In fact, the fructose monomer itself is not found in human milk. Besides, neither GOS nor FOS are fucosylated nor sialylated (See Figure 1.18).

These simpler oligosaccharides do not contain the structural complexity and diversity of HMO; thus, most of them do not resemble epithelial cell surface glycans, and may not be recognized by these lectins or bind to other lectins that trigger a different cell response. Even prebiotic effects may, to a certain extent, be structure dependent (Ninonuevo and Bode, 2008). However, FOS and GOS are reported to mimic the prebiotic effects of human milk and promote a bacterial microflora that closely resembles that of breast-fed infants (Ninonuevo and Bode, 2008).

To date, GOS, FOS or inulin represent a reasonable and inexpensive way to supplement infant formula with prebiotic oligosaccharides. In the future, better sources of complex oligosaccharides that more closely mimic the structures and functions of HMO are needed to improve upon the existing supplementation strategy (Zivkovic and Barile, 2011).

1.2 LACTO-*N*-BIOSIDASE

Lacto-*N*-biosidase (EC 3.2.1.140) was first found in *Streptomyces sp*. strain 142, a soil actinomycete, in the course of the preparation of an α-fucosidase batch (Sano *et al*., 1992). In 2008, Wada *et al*. identified the second lacto-*N*-biosidase gene from *B. bifidum* JCM1254, and classified it as an exo-β-*N*-acetylhexosaminidase that acts upon the terminal Galβ-1,3GlcNAc residues of HMO to release LNB (Wada *et al*., 2008a). While the role of Lacto-*N*-biosidase in *B. bifidum* is clearly located in the metabolism of the HMO, the role of this enzyme in *Streptomyces sp*. is yet to be fully elucidated (Hattie *et al*., 2012).

1.3.1 Occurrence and physiological role in bifidobacteria

Bifidobacteria are Gram-positive anaerobic bacteria that naturally occur in the human intestinal tract. They are considered to be beneficial commensals for human health because they prevent the growth of pathogenic bacteria by lowering the intestinal pH and stimulate the host's immune system to enhance anti-pathogenic and anti-carcinogenic activities (Miwa *et al*., 2010). After a week of birth, breast-fed infants guts are colonized by bifidobacteria, in particular *B. breve*, *B. longum* subsp. *longum/infantis* and *B. bifidum*, in comparison with those of formula-fed infants, which gut flora is more similar to adults (Harmsen *et al*., 2000).

The lower gastrointestinal tract is an environment poor in mono- and disaccharides because these sugars are consumed by the host. In order to survive, bifidobacteria have different

extracellular glycosidases and a unique sugar metabolic pathway for LNB and GNB (galacto-*N*-biose) called The GNB/LNB pathway, as it is represented in Figure 1.19 (Fushinobu, 2010).

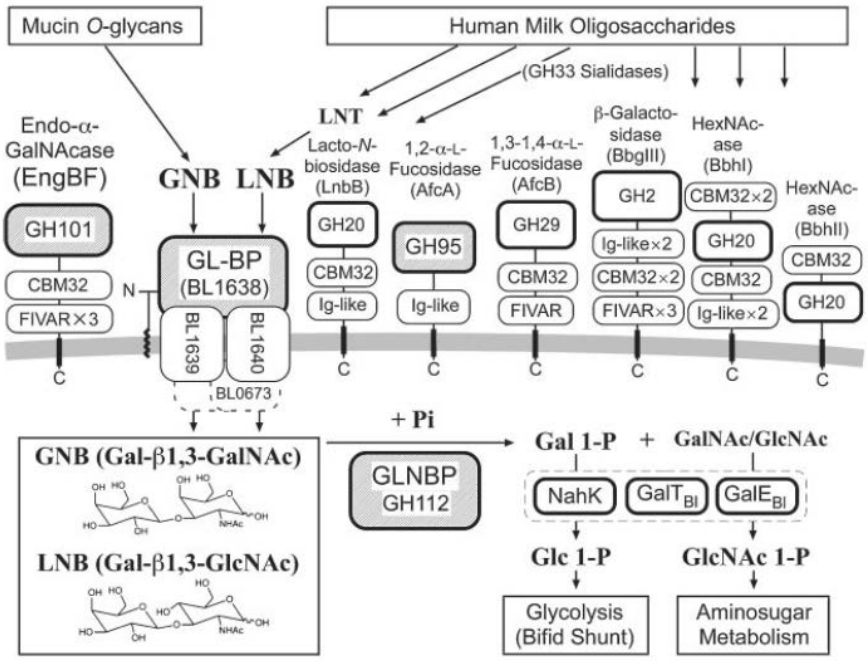


Figure 1.19 GNB/LNB Pathway and Extracellular Glycosidases of Bifidobacteria.
Domain organization of extracellular glycosidases and the GNB/LNB transporter. The crystal structures of enzymes (or proteins) shown as hatched boxes have been determined. EngBF was isolated from B. longum JCM1217, whereas the other extracellular glycosidases were isolated from B. bifidum JCM1254. Image from Fushinobu *et al*., 2010.

LNB is a building unit of type 1 HMO, and GNB is a core 1 disaccharide of O-glycans of mucin glycoproteins that are present in the human intestines and milk (Wada *et al*., 2008a). The GNB/LNB pathway involves the following five proteins required for the uptake and degradation of the disaccharides: GNB/LNB transporter, lacto-N-biose phosphorylase (LnpA), N-acetylhexosamine 1-kinase (NahK), UDP-glucose-hexose 1-phosphate uridylyltransferase (GalT), and UDP-galactose epimerase (GalE). It is likely that LNB and GNB are first imported by GNB/LNB transporter into the cytoplasm through a specific ABC-type transporter and phosphorolyzed by galacto-*N*-biose/lacto-*N*-biose phosphorylase (GNLBP), and finally metabolized in the glycolytic and aminosugar metabolic pathways (Kiyohara *et al*., 2009; Wada *et al*., 2008a).

Based on the dominance of LNB-containing oligosaccharides in HMO and the presence of the GNB/LNB pathway in bifidobacteria, Kitaoka *et al*. 2005, hypothesized that the LNB residues in HMO act as the bifidus factor in breastfed infants (Kitaoka *et al*., 2005). Xiao *et al*. 2010, in a study of LNB as a sole source of carbon for 208 strains of bifidobacteria, confirmed that all the strains that naturally and predominantly colonize the infant gastrointestinal tract grew when LNB was used as sole carbohydrate source, while the strains that are found in adult gastrointestinal tract, human dental caries, and the guts of domestic animals are not able to utilize LNB (Xiao *et al*., 2010).

Since no free form of LNB is present in HMO, extracellular glycosidases responsible for the cleavage and release of LNB from HMO and the release of GNB from mucin glycoprotein have to be present in the gastrointestinal tract.

1.3.2 Characteristics of Lacto-*N*-biosidase from *B.bifidum*

The release of LNB from HMO is the most important step of type 1 HMO degradation. The discovery of GNB/LNB pathway presupposed the presence of lacto-*N*-biosidase in bifidobacteria.

Bifidobacterial lacto-*N*-biosidase (LnbB) was first identified by Wada *et al*., 2008 in *Bifidobacterium bifidum* JCM 1254, and since then, it has also been found in different strains of *B. bifidum* and *B. longum*. The gene (l*nbB*) encoding for LnbB from *B. bifidum* was isolated using degenerated primers based on the amino acid sequence of LnbB from *Streptomyces sp*. strain 142, and the recombinant protein was characterised. LnbB consists of 1112 amino acids and contains an *N*-terminal signal sequence and a C-terminal membrane anchor. LnbB exhibits 38% amino acid identity to *Streptomyces sp* protein. In terms of catalytic classification, LnbB belongs to the GH20 family of glycosyl hydrolases (CAZY database, (Cantarel *et al*., 2009)) since it presents a GH20 catalytic domain (Ito *et al*., 2013; Wada *et al*., 2008a). GH20 enzymes operate by substrate-assisted catalysis meaning the enzyme does not present the nucleophilic amino acid and the *N*-acetamido group of the substrate acts as the nucleophile to generate an oxazolinium intermediate. Most characterized GH20 enzymes are β-*N*-acetylhexosaminidases (EC 3.2.1.52), which release a monosaccharide, GlcNAc or GalNAc, from the non-reducing end of oligosaccharides. However, as it is observed in Figure 1.20, LnbB releases a disaccharide (LNB), implying that this enzyme has an extended -2 subsite, compared to the rest of GH20 enzymes that only present -1 subsite (Ito *et al*., 2013).

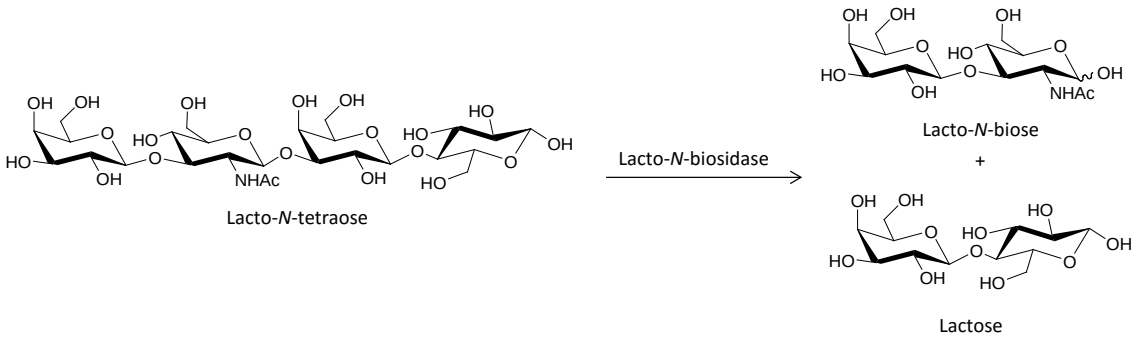


Figure 1.20 Hydrolytic reaction of lacto-*N*-tetraose catalyzed by *B. bifidum* lacto-*N*-biosidase.

Lacto-*N*-biosidase exhibits a very specific activity releasing LNB from lacto-*N*-tetraose (LNT) (Figure 1.20). LnbB shows strict preference for the unmodified β-linked LNB structure, indicating that cleavage of the fucose and sialic acid modifications of HMO is prerequisite for activity (Wada *et al*., 2008a).

Due to the relevant characteristics of LnbB, this thesis envisages LnbB as a promising biocatalyst for HMO synthesis. Accordingly, a deeper characterization of LnbB has been achieved and is revealed in the following chapters.

1.5 GLYCOSIDASES AS TOOLS FOR IMPROVED GLYCOSIDE SYNTHESIS

In nature, glycosidic bond formation is mainly synthesized by two different types of enzymes, which are glycosyltransferases (GTs) and glycosidases (glycoside hydrolases, GHs). GTs catalyse the transfer of a monosaccharide sugar nucleotide donor to an acceptor, producing the new sugar. They are the main catalyst *in vivo*, but the large-scale use *in vitro* is limited. GTs are often membrane bound proteins, hard to isolate and purify, and unstable once they are isolated. In addition, these enzymes require activated glycosyl donor as substrates, which are not readily available and increases the already high cost of the reaction. Therefore, the use of glycosyltransferases is still considered as a general tool for *in vitro* micro-scale carbohydrate synthesis (Planas *et al*., 2015; Wang and Davis, 2013). With regard to HMO synthesis, some simple HMO are produced *in vitro* using the corresponding GTs as previously cited (section 1.2.5.3). However, a deeply understanding of the synthetic pathways of HMO is required for a complete benefit of this methodology.

Alternatively, GHs are degrading enzymes involved in glycan metabolism which hydrolyse glycosidic linkages in a stereospecific manner. Glycoside hydrolases fall into two mechanistic categories, according to the stereochemical outcome of the bond cleavage: inverting enzymes, which hydrolyse the glycosidic bond with inversion of configuration, and retaining enzymes, which do so with net retention of the anomeric configuration. Inverting GHs operate by a single-step mechanism in which a water molecule (with general base catalysis) affects a direct displacement at the anomeric center with protonic assistance by the general acid residue on the departing glycosidic oxygen (Figure 1.21C). While retaining GHs follow a double displacement reaction via formation and hydrolysis of a glycosyl–enzyme intermediate. In the first step (glycosylation), the amino acid residue acts as the general acid and protonates the glycosidic oxygen while the deprotonated carboxylate functions as the nucleophile and attacks the anomeric centre leading to a covalent glycosyl–enzyme intermediate. The second step (deglycosylation) involves the attack by a molecule of water assisted by the conjugate base of the general acid residue, which renders the free sugar (Figure 1.21A). Some retaining glycosidases catalysing glycoside bond hydrolysis in 2-acetamido sugars follow a variation of the retaining mechanism (Figure 1.21B). In this case, the *N*-acetyl group of the substrate acts as an internal nucleophile, forming an oxazolinium intermediate, which is then attacked by a water molecule assisted by the conjugate base of the general acid residue to yield the product. Instead of a nucleophile residue, an auxiliary amino acid hydrogen bonds with the –NH of the acetamido group assisting the formation of the oxazolinium intermediate (Planas *et al*., 2015).

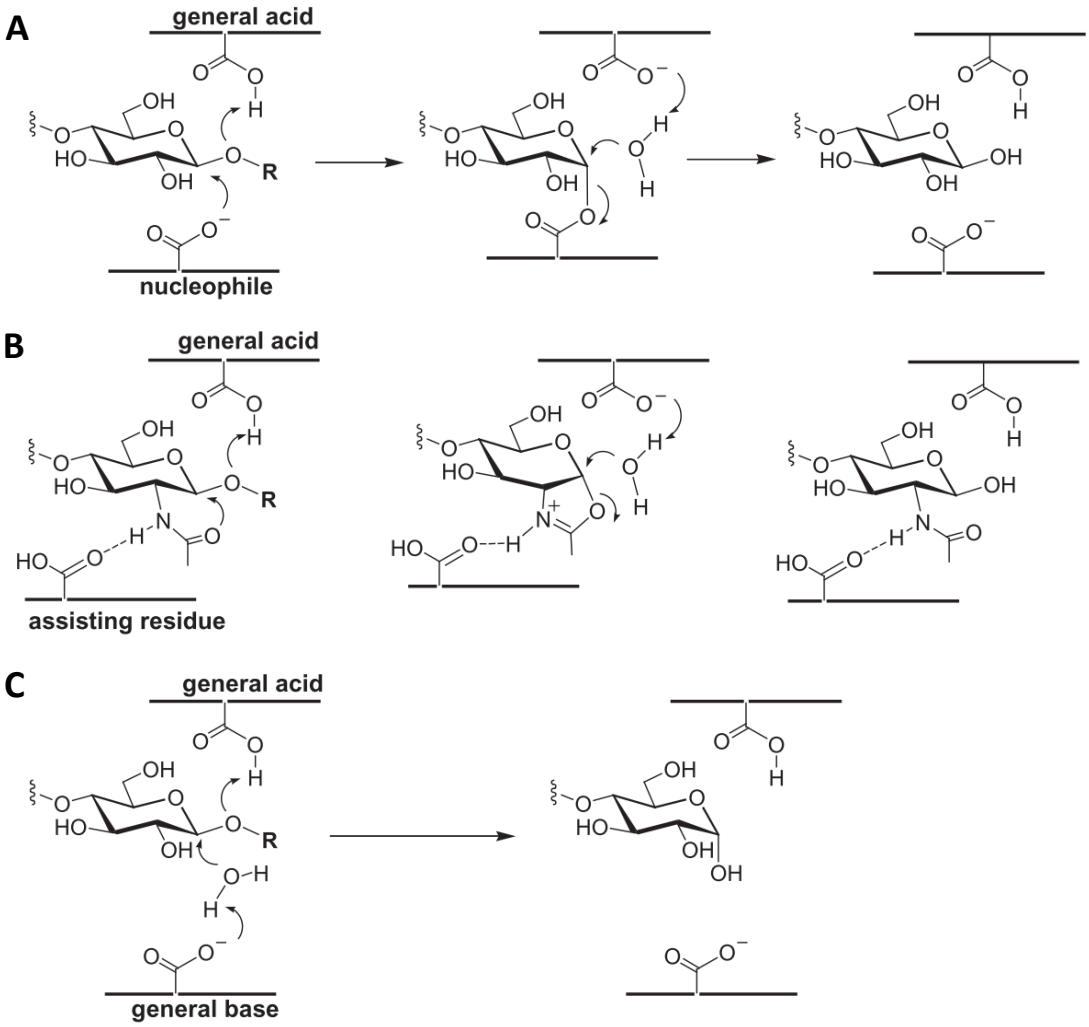


Figure 1.21 Glycosidase mechanism.
(A) Retaining glycosidase with enzyme nucleophile, via a glycosyl–enzyme intermediate in the two-step displacement mechanism. (B) Retaining glycosidase by substrate-assisted catalysis, via an oxazolinium ion intermediate (C) Inverting glycosidase.

GHs hydrolytic activity can be reverted to the synthesis of glycans through transglycosylation reactions under appropriate conditions. GHs are applied as efficient biosynthetic tools due to the easy access and stability of the enzymes, and the use of glycosyl donors that are cheap and relatively simple to produce in large scale. Classical approaches are based on the reversal of their hydrolytic activity. This is achieved either by displacing the equilibrium towards glycoside bond formation (thermodynamically controlled synthesis) or by using activated glycosyl donors (kinetically controlled transglycosylation). However, since these enzymes are still hydrolytically active and reactions are run in water, the product of transglycosylation is always hydrolysed reducing yields. In any case, these methodologies only achieve yields up to 40-50% (Bissaro *et al*., 2015; Planas *et al*., 2015; Wang and Huang, 2009).

Improvement of transglycosylation efficiency of natural GHs can also be achieved by protein engineering. The strategies for engineering glycosidases generally consist of rational design on the basis of available structural information as well as directed evolution with use of powerful screening approaches. In site-directed mutagenesis, the mutations play a role in either

disfavouring the positioning of hydrolytic water molecule or favouring binding of incoming carbohydrate molecule in the aglycon subsite, thereby increasing transglycosylation activity. Directed evolution, on the other hand, can be used for improving transglycosylation activity without prior knowledge of structure-function relationships.

Progress in enzymatic synthesis was achieved in 1998, when the glycosynthase technology was developed (Mackenzie *et al*., 1998; Malet and Planas, 1998). Glycosynthases are nucleophile mutants of retaining glycosidases that efficiently catalyse the synthesis of the glycosidic linkage by condensing an activated glycosyl fluoride donor with a suitable acceptor sugar. They can efficiently catalyse glycoside bond formation and achieve good yields, since they are hydrolytically inactive. As shown in Figure 1.22A, the α-glycosyl fluoride donor mimics the glycosyl-enzyme intermediate in the active site of the nucleophile-less mutant of a β-glycosidase. This enzyme transglycosylates this donor to the acceptor and the product formed is unable to be hydrolyzed by the mutant enzyme, so transglycosylation yields are very high (Planas *et al*., 2015; Armstrong and Withers, 2013).

The glycosynthase technology has thereafter been applied in a wide variety of glycosynthases. The initial concept developed in mutated retaining glycosidases using glycosyl fluoride donors (Figure 1.22B), was then extended to other approaches including the *in situ* generation of an activated glycosyl donor with an external nucleophile (Figure 1.22B), glycosynthases derived from glycosidases acting by substrate-assisted catalysis (Figure 1.22C), and glycosynthases derived from inverting glycosidases (Figure 1.22D) (Armstrong and Withers, 2013; Planas *et al*., 2015).

The applicability of enzymes from thermophilic organisms is limited because of the instability of the glycosyl fluoride donor at high temperatures. The *in situ* generation of the glycosyl donor in the active site of the glycosynthase mutant is used as an alternative strategy. Then, the activity on an aryl glycoside substrate can be rescued by an exogenous nucleophile such as formate or azide and condensation to an acceptor can occur (Trincone *et al*., 2000).

It has also been developed glycosynthases acting *via* substrate-assisted mechanism. In this case, the glycosyl donor for transglycosylation is an activated oxazoline sugar, which binds the enzyme to directly form the reaction intermediate. The first glycosynthase acting *vi*a substrate-assisted catalysis was an endo-β-*N*-acetylglucosaminidase (Endo-M) of GH85 family proteins, and was further extended to other GH85 enzymes (Endo-A and Endo-D) (Wang and Huang, 2009; Yamamoto, 2013) and also to GH18 family proteins (Huang *et al*., 2012; Martinez *et al*., 2012).

The action of inverting enzymes on glycosyl fluorides with an opposite anomer was first mentioned by Honda and coworkers. As inverting glycosidases operate by a single displacement mechanism, enzyme engineering for glycosynthesis in this case is based on a similar but slightly different concept. Inverting glycosynthases transfer the glycosyl fluoride of the opposite anomer of the substrate (e.g. inFigure 1.24, α-glycosyl fluoride by a β-glycoside-hydrolyzing enzyme) to the acceptor molecule. However, although the normal hydrolytic process is substantially slowed, is still significant (Honda and Kitaoka, 2006). As regards HMO synthesis, 1,2-α-L-Fucosidase (AfcA) from *Bifidobacterium bifidum* was engineered to condense a fucose moiety to a lactose to form 2'-fucosyllactose (Wada *et al*., 2008b).

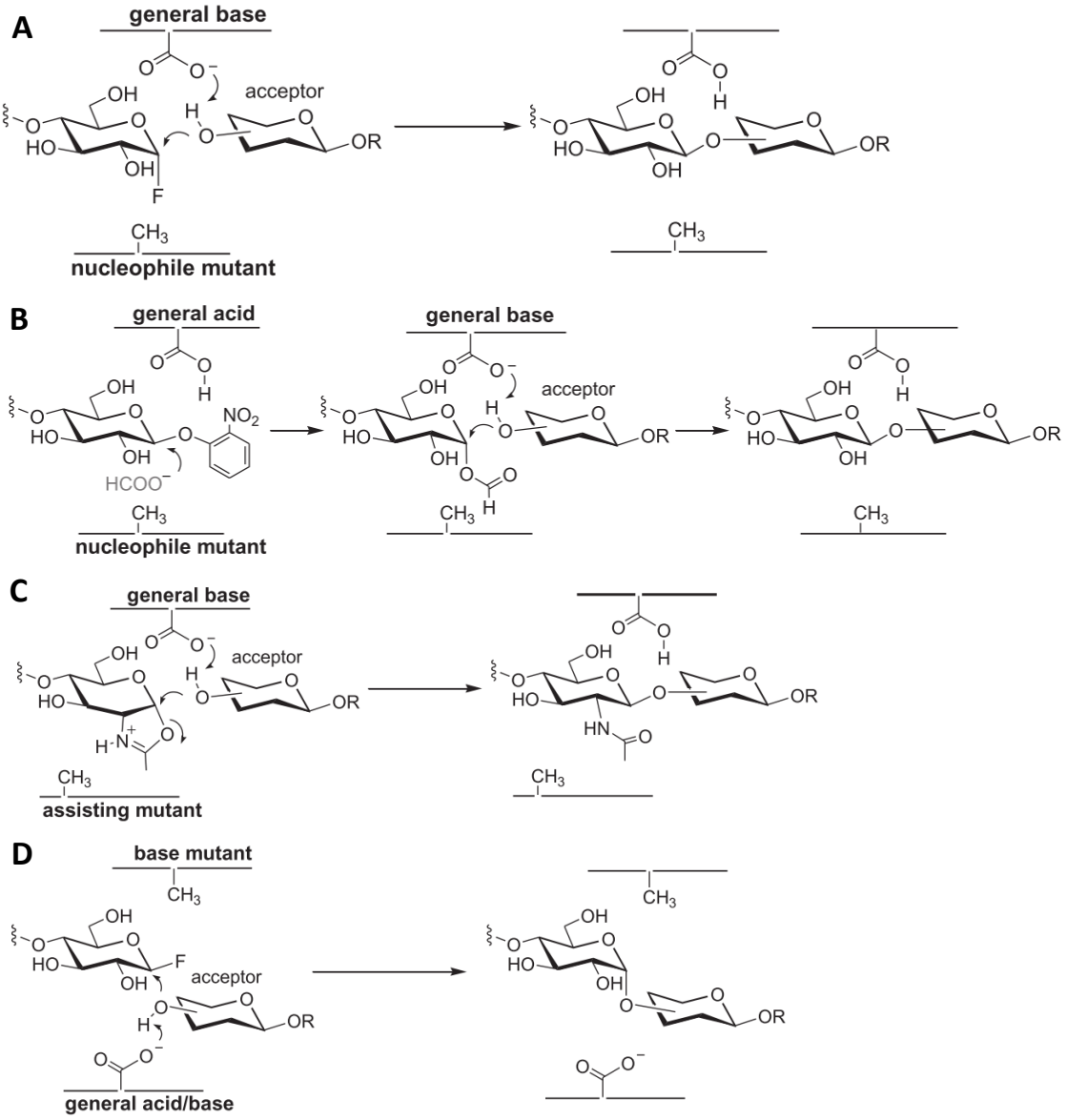


Figure 1.22 Glycosynthase enzymes.
(A) Nucleophile mutant of a retaining GH with glycosyl fluoride donor. (B) Nucleophile mutant of a retaining glycosidase (thermophilic) with in situ generation of the glycosyl donor by and exogenous nucleophile. (C) Mutant at the assisting residue of a retaining glycosidase acting by substrate-assisted catalysis and a sugar oxazoline donor. (D) Glycosynthase-like mutant derived from an inverting glycosidase.

To date, more than 50 glycosynthases have been developed covering a wide repertoire of sugar moieties and glycosidic linkages (Planas *et al*., 2015) but still, no glycosynthase exists to synthesize any core structure of HMO sugars.

1.5 AIM OF THE THESIS

Human milk oligosaccharides (HMO) are of current interest due to the short and long term beneficial effects they provide to the infant. To date, large scale synthesis of HMO has not been possible by any synthetic methodology. Consequently, HMO are not available in infant formulas. Enzymatic synthesis is an alternative strategy that overcomes chemical synthetic limitations since enzymes can form the new glycosidic linkages with high regio- and stereo-selectivity. In particular, glycosidases, which mainly catalyse the hydrolysis of oligo- and polysaccharides *in vivo*, have been shown to be effective biocatalysts when they are converted to glycosynthases.

Using structural-functional knowledge of glycosidase enzymes and rational redesign, this project aims to develop new biocatalysts to achieve the large scale synthesis of type 1 oligosaccharides.

Recently, the Lacto-*N*-biosidase from *Bifidobacterium bifidum* (LnbB) was identified. LnbB specifically hydrolyses the Lacto-*N*-tetraose (type 1 HMO) compound into two disaccharides, lacto-*N*-biose and lactose, *via* substrate-assisted mechanism.

On this basis, the present thesis tackles two main objectives to address the use of LnbB as an efficient biocatalyst: i) the structural-functional study of LnbB and ii) protein engineering of LnbB to generate biocatalysts able to synthesize the target lacto-*N*-tetraose. Modulation of the transglycosylation/hydrolysis activity of LnbB will be evaluated using two different strategies: the glycosynthase-like and the enhanced transglycosylation strategies.

1.6 REFERENCES

Albermann, C., Piepersberg, W., and Wehmeier, U.F. (2001). Synthesis of the milk oligosaccharide 2'-fucosyllactose using recombinant bacterial enzymes. Carbohydr. Res. *334*, 97–103.

Albrecht, S., Lane, J. a, Mariño, K., Al Busadah, K. a, Carrington, S.D., Hickey, R.M., and Rudd, P.M. (2014). A comparative study of free oligosaccharides in the milk of domestic animals. Br. J. Nutr. *111*, 1313–1328.

Andersson, B., Porras, O., Hanson, L.A., Lagergard, T., and Svanborg-eden, C. (1986). Inhibition of Attachment of Streptococcus pneumoniae and Haemophilus influenzae by Human Milk and Receptor Oligosaccharides. J. Infect. Dis. *153*, 232–237.

Andreas, S., Gyula, D., Peter, E., and Schwarz, A. (2013). WO2013185780A1. Enhancing the stability and purity and increasing the bioavailability of human milk oligosaccharides or precursors or blends thereof.

Angeloni, S. (2004). Glycoprofiling with micro-arrays of glycoconjugates and lectins. Glycobiology *15*, 31–41.

Arboe Jennum, C., Hauch Fenger, T., Bruun, L.M., and Madsen, R. (2014). One-pot glycosylations in the synthesis of human milk oligosaccharides. European J. Org. Chem. 3–4.

Armstrong, Z., and Withers, S.G. (2013). Synthesis of glycans and glycopolymers through engineered enzymes. Biopolymers *99*, 666–674.

Asakuma, S., Hatakeyama, E., Urashima, T., Yoshida, E., Katayama, T., Yamamoto, K., Kumagai, H., Ashida, H., Hirose, J., and Kitaoka, M. (2011). Physiology of Consumption of Human Milk Oligosaccharides by Infant Gut-associated Bifidobacteria. J. Biol. Chem. *286*, 34583–34592.

Barboza, M., Sela, D.A., Pirim, C., LoCascio, R.G., Freeman, S.L., German, J.B., Mills, D.A., and Lebrilla, C.B. (2009). Glycoprofiling Bifidobacterial Consumption of Galacto-Oligosaccharides by Mass Spectrometry Reveals Strain-Specific, Preferential Consumption of Glycans. Appl. Environ. Microbiol. *75*, 7319–7325.

Baumgärtner, F., Conrad, J., Sprenger, G.A., and Albermann, C. (2014). Synthesis of the human milk oligosaccharide lacto-N-tetraose in metabolically engineered, plasmid-free E. coli. ChemBioChem 1896–1900.

Baumgärtner, F., Jurzitza, L., Conrad, J., Beifuss, U., Sprenger, G.A., and Albermann, C. (2015). Synthesis of fucosylated lacto-N-tetraose using whole-cell biotransformation. Bioorg. Med. Chem.

Bayón, C., Cortés, Á., Berenguer, J., and Hernáiz, M.J. (2013). Highly efficient enzymatic synthesis of Galβ-(1→3)-GalNAc and Galβ-(1→3)-GlcNAc in ionic liquids. Tetrahedron *69*, 4973–4978.

Bissaro, B., Monsan, P., Fauré, R., and O'Donohue, M.J. (2015). Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. Biochem. J. *467*, 17–35.

Blank, D., Dozt, V., Geyer, R., and Kunz, C. (2012). Human Milk Oligosaccharides and Lewis Blood Group : Individual High-Throughput Sample Pro fi ling to Enhance Conclusions From. Adv. Nutr. *3*, 440–449.

Bode, L. (2006). Recent advances on structure, metabolism, and function of human milk oligosaccharides. J. Nutr. *136*, 2127–2130.

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 1–42.

Bode, L. (2015). The functional biology of human milk oligosaccharides. Early Hum. Dev. *91*, 619–622 4p.

Bode, L., and Jantscher-krenn, E. (2012). Structure-Function Relationships of Human Milk Oligosaccharides. Adv. Nutr. 383–391.

Bode, L., Rudloff, S., Kunz, C., Strobel, S., and Klein, N. (2004). Human milk oligosaccharides reduce platelet-neutrophil complex formation leading to a decrease in neutrophil beta 2 integrin expression. J. Leukoc. Biol. *76*, 820–826.

Boehm, G., and Stahl, B. (2007). Oligosaccharides from milk. J. Nutr. *137*, 847S – 9S.

Boehm, G., Stahl, B., Jelinek, J., Knol, J., Miniello, V., and Moro, G. (2005). Prebiotic carbohydrates in human milk and formulas. Acta Paediatr. *94*, 18–21.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. *37*, D233–D238.

Castagner, B., and Seeberger, P.H. (2007). Automated solid phase oligosaccharide synthesis. Top. Curr. Chem. *278*, 289–309.

Castanys-Muñoz, E., Martin, M.J., and Prieto, P.A. (2013). 2'-Fucosyllactose: an Abundant, Genetically Determined Soluble Glycan Present in Human Milk. Nutr. Rev. *71*, 773–789.

Chen, X. (2015). Human Milk Oligosaccharides (HMOS): Structure, Function, and Enzyme-Catalyzed Synthesis. Adv. Carbohydr. Chem. Biochem. 72, 113–190.

Coppa, G. V, Pierani, P., Zampini, L., Carloni, I., Carlucci, a, and Gabrielli, O. (1999). Oligosaccharides in human milk during different phases of lactation. Acta Paediatr. Suppl. *88*, 89–94.

Coppa, G. V, Zampini, L., Galeazzi, T., Facinelli, B., Ferrante, L., Capretti, R., and Orazio, G. (2006a). Human Milk Oligosaccharides Inhibit the Adhesion to Caco-2 Cells of Diarrheal Pathogens: Escherichia coli, Vibrio cholerae, and Salmonella fyris. Pediatr. Res. *59*, 377–382.

Coppa, G. V., Zampini, L., Galeazzi, T., and Gabrielli, O. (2006b). Prebiotics in human milk: a review. Dig. Liver Dis. *38*, 291–294.

Dallas, D., Sela, D., Underwood, M.A., German, J.B., and Lebrilla, C.B. (2012). Protein-Linked Glycan Degradation in Infants Fed Human Milk. J. Glycomics Lipidomics *s1*.

Drouillard, S., Driguez, H., and Samain, E. (2006). Large-scale synthesis of H-antigen oligosaccharides by expressing Helicobacter pylori alpha1,2-fucosyltransferase in metabolically engineered Escherichia coli cells. Angew. Chem. Int. Ed. Engl. *45*, 1778–1780.

Drouillard, S., Mine, T., Kajiwara, H., Yamamoto, T., and Samain, E. (2010). Efficient synthesis of 6'-sialyllactose, 6,6;-disialyllactose, and 60-KDO-lactose by metabolically engineered E. coli expressing a multifunctional sialyltransferase from the Photobacterium sp. JT-ISH-224. Carbohydr. Res. *345*, 1394–1399.

Dumon, C., Samain, E., and Priem, B. (2004). Assessment of the two Helicobacter pylori ??-1,3-fucosyltransferase ortholog genes for the large-scale synthesis of lewisx human milk oligosaccharides by metabolically engineered Escherichia coli. Biotechnol. Prog. *20*, 412–419.

Egge, H., Dell, A., and Nicolai, H. Von (1983). Fucose-containing oligosaccharides from human milk. I. Separation and Identification of new constituents. Arch. Biochem. Biophys. *224*, 235–253.

Endo, T., and Koizumi, S. (2000). Large-scale production of oligosaccharides using engineered bacteria. Curr. Opin. Struct. Biol. *10*, 536–541.

Engfer, M.B., Stahl, B., Finke, B., Sawatzki, G., and Daniel, H. (2000). Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. Am. J. Clin. Nutr. *71*, 1589–1596.

Escherich, T. (1886). Die darmbakterien des säuglings und ihre beziehungen zur physiologie der Verdauung.

European Commission (2013). Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32003R1829.

Faijes, M., and Planas, A. (2007). In vitro synthesis of artificial polysaccharides by glycosidases and glycosynthases. Carbohydr. Res. *342*, 1581–1594.

Fierfort, N., and Samain, E. (2008). Genetic engineering of Escherichia coli for the economical production of sialylated oligosaccharides. J. Biotechnol. *134*, 261–265.

Fuhrer, A., Sprenger, N., Kurakevich, E., Borsig, L., Chassard, C., and Hennet, T. (2010). Milk sialyllactose influences

colitis in mice through selective intestinal bacterial colonization. J. Exp. Med. *207*, 2843–2854.

Fushinobu, S. (2010). Unique sugar metabolic pathways of bifidobacteria. Biosci. Biotechnol. Biochem. *74*, 2374–2384.

Gabrielli, O., Zampini, L., Galeazzi, T., Padella, L., Santoro, L., Peila, C., Giuliani, F., Bertino, E., Fabris, C., and Coppa, G. V. (2011). Preterm Milk Oligosaccharides During the First Month of Lactation. Pediatrics *128*, e1520–e1531.

Garrido, D., Barile, D., and Mills, D. a (2012). The Glycobiology of Human Milk Oligosaccharides A Molecular Basis for Bifidobacterial Enrichment in the Infant Gastrointestinal Tract 1 – 3. Adv. Nutr. An Int. Rev. J. *3*, 415–421.

Garrido, D., Dallas, D.C., and Mills, D.A. (2013). Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications. Microbiology *159*, 649–664.

Garrido, D., Ruiz-Moyano, S., Lemay, D.G., Sela, D.A., German, J.B., and Mills, D.A. (2015). Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. Sci. Rep. *5*, 13517.

German, J.B., Freeman, S.L., Lebrilla, C.B., and Mills, D. a (2008). Human milk oligosaccharides: evolution, structures and bioselectivity as substrates for intestinal bacteria. Nestle Nutr. *62*, 205–222.

Gibson, G.R., and Roberfroid, M.B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J. Nutr. *125*, 1401–1412.

Goehring, K.C., Kennedy, A.D., Prieto, P.A., and Buck, R.H. (2014). Direct Evidence for the Presence of Human Milk Oligosaccharides in the Circulation of Breastfed Infants. PLoS One *9*, e101692.

Green-Johnson, J.M. (2012). Immunological Responses to Gut Bacteria. J. AOAC Int. *95*, 35–49.

Gyorgy, P., Norris, R.F., and Rose, C.S. (1954). Bifidus factor. I. A variant of Lactobacillus bifidus requiring a special growth factor. Arch. Biochem. Biophys. *48*, 193–201.

Han, N.S., Kim, T.-J., Park, Y.-C., Kim, J., and Seo, J.-H. (2012). Biotechnological production of human milk oligosaccharides. Biotechnol. Adv. *30*, 1268–1278.

Harmsen, H.J.M., Wildeboer–Veloo, A.C.M., Raangs, G.C., Wagendorp, A.A., Klijn, N., Bindels, J.G., Welling, Gjalt W.Harmsen, Hermie J. Wildeboer–Veloo, A.C.M., Raangs, G.C., Wagendorp, A.A., Klijn, N., et al. (2000). Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods. J. Pediatr. Gastroenterol. Nutr. *30*, 61–67.

Hattie, M., Debowski, A.W., and Stubbs, K.A. (2012). Development of tools to study Lacto-N-Biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. ChemBioChem *13*, 1128–1131.

Hennet, T., Weiss, A., and Borsig, L. (2014). Decoding breast milk oligosaccharides. Swiss Med. Wkly. 1–9.

Hoeflinger, J.L., Davis, S.R., Chow, J., and Miller, M.J. (2015). In Vitro Impact of Human Milk Oligosaccharides on Enterobacteriaceae Growth. J. Agric. Food Chem. *63*, 3295–3302.

Honda, Y., and Kitaoka, M. (2006). The first glycosynthase derived from an inverting glycoside hydrolase. J. Biol. Chem. *281*, 1426–1431.

Huang, W., Giddens, J., Toonstra, C., and Wang, L.-X. (2012). Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions. J Am Chem Soc. *134*, 12308–12318.

Intanon, M., Arreola, S.L., Pham, N.H., Kneifel, W., Haltrich, D., and Nguyen, T.-H. (2014). Nature and biosynthesis of galacto-oligosaccharides related to oligosaccharides in human breast milk. FEMS Microbiol. Lett. *353*, 89–97.

Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., Ashida, H., Wakagi, T., Yamamoto, K., Stubbs, K.A., et al. (2013). Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-N-biosidase from Bifidobacterium bifidum. J. Biol. Chem. *288*, 11795–11806.

Jantscher-Krenn, E., Lauwaet, T., Bliss, L.A., Reed, S.L., Gillin, F.D., and Bode, L. (2012a). Human milk

oligosaccharides reduce Entamoeba histolytica attachment and cytotoxicity in vitro. Br. J. Nutr. *108*, 1839–1846.

Jantscher-Krenn, E., Zherebtsov, M., Nissan, C., Goth, K., Guner, Y.S., Naidu, N., Choudhury, B., Grishin, A. V., Ford, H.R., and Bode, L. (2012b). The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats. Gut *61*, 1417–1425.

Jeong, K., Nguyen, V., and Kim, J. (2012). Human milk oligosaccharides: The novel modulator of intestinal microbiota. BMB Rep. *45*, 433–441.

Kitaoka, M., Tian, J., and Nishimoto, M. (2005). Novel Putative Galactose Operon Involving Lacto- N -Biose Phosphorylase in Bifidobacterium longum Novel Putative Galactose Operon Involving Lacto- N -Biose Phosphorylase in Bifidobacterium longum. *71*, 3158–3162.

Kiyohara, M., Tachizawa, A., Nishimoto, M., Kitaoka, M., Ashida, H., and Yamamoto, K. (2009). Prebiotic effect of lacto-N-biose I on bifidobacterial growth. Biosci. Biotechnol. Biochem. *73*, 1175–1179.

Kobata, A. (2003). Possible application of milk oligosaccharides for drug development. Chang Gung Med J *26*, 620–636.

Kobata, A. (2010). Structures and application of oligosaccharides in human milk. Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci. *86*, 731–747.

Kogelberg, H., Piskarev, V.E., Zhang, Y., Lawson, A.M., and Chai, W. (2004). Determination by electrospray mass spectrometry and 1H-NMR spectroscopy of primary structures of variously fucosylated neutral oligosaccharides based on the iso-lacto-N-octaose core. Eur. J. Biochem. *271*, 1172–1186.

Kosloske, and Musemeche (1989). Necrotizing enterocolitis of the neonate. Clin. Perinatol. *16*, 97–111.

Kuntz, S., Rudloff, S., and Kunz, C. (2008). Oligosaccharides from human milk influence growth-related characteristics of intestinally transformed and non-transformed intestinal cells. Br. J. Nutr. *99*, 462–471.

Kuntz, S., Kunz, C., and Rudloff, S. (2009). Oligosaccharides from human milk induce growth arrest via G2/M by influencing growth-related cell cycle genes in intestinal epithelial cells. Br. J. Nutr. *101*, 1306.

Kunz, C. (2012). Historical Aspects of Human Milk Oligosaccharides. Adv. Nutr. An Int. Rev. J. *3*, 430S – 439S.

Kunz, C., Rudloff, S., Schad, W., and Braun, D. (1999). Lactose-derived oligosaccharides in the milk of elephants: comparison with human milk. Br. J. Nutr. *82*, 391–399.

Kunz, C., Rudloff, S., Baie, W., Klein, N., and St, S. (2000). O LIGOSACCHARIDES IN H UMAN M ILK : Structural , Functional , and Metabolic Aspects. Annu. Rev. Nutr. 699–722.

Lane, J.A., and Hickey, R.M. (2014). Food Oligosaccharides (John Wiley & Sons, Ltd).

Lau, K., Yu, H., Thon, V., Khedri, Z., Leon, M., Tran, B., and Chen, X. (2011). Sequential two-step multienzyme synthesis of tumor-associated sialyl T-antigens and derivatives. *79*, 2784–2789.

Li, Y., Yu, H., Cao, H., Lau, K., Muthana, S., Tiwari, V.K., Son, B., and Chen, X. (2008). Pasteurella multocida sialic acid aldolase: a promising biocatalyst. Appl. Microbiol. Biotechnol. *79*, 963–970.

LoCascio, R.G., Ninonuevo, M.R., Freeman, S.L., Sela, D. a., Grimm, R., Lebrilla, C.B., Mills, D. a., and German, J.B. (2007). Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. J. Agric. Food Chem. *55*, 8914–8919.

Lucas, a., and Cole, T.J. (1990). Breast milk and neonatal necrotising enterocolitis. Lancet *336*, 1519–1523.

Mackenzie, L.F., Wang, Q., Warren, R.A.J., and Withers, S.G. (1998). Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. J. Am. Chem. Soc. *120*, 5583–5584.

Malet, C., and Planas, A. (1998). From beta-glucanase to beta-glucansynthase: glycosyl transfer to alpha-glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. FEBS Lett. *440*, 208–212.

Marcobal, A., Barboza, M., Froehlich, J.W., Block, D.E., Bruce, J., Lebrilla, C.B., and Mills, D.A. (2011). Consumption of human milk oligosaccharides by gut-related microbes. J. Agric. Food Chem. *58*, 5334–5340.

Martinez, E.A., Boer, H., Koivula, A., Samain, E., Driguez, H., Armand, S., and Cottaz, S. (2012). Engineering chitinases for the synthesis of chitin oligosaccharides: Catalytic amino acid mutations convert the GH-18 family glycoside hydrolases into transglycosylases. J. Mol. Catal. B Enzym. *74*, 89–96.

Martinez-Ferez, A., Rudloff, S., Guadix, A., Henkel, C. a., Pohlentz, G., Boza, J.J., Guadix, E.M., and Kunz, C. (2006). Goats' milk as a natural source of lactose-derived oligosaccharides: Isolation by membrane technology. Int. Dairy J. *16*, 173–181.

Martín-Sosa, S., Martín, M.-J., and Hueso, P. (2002). The sialylated fraction of milk oligosaccharides is partially responsible for binding to enterotoxigenic and uropathogenic Escherichia coli human strains. J. Nutr. *132*, 3067–3072.

Miller, J.B., and McVeagh, P. (1999). Human milk oligosaccharides: 130 reasons to breast-feed. Br. J. Nutr. *82*, 333–335.

Miwa, M., Horimoto, T., Kiyohara, M., Katayama, T., Kitaoka, M., Ashida, H., and Yamamoto, K. (2010). Cooperation of B-galactosidase and B-N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. Glycobiology *20*, 1402–1409.

Miyazaki, T., Sato, T., Furukawa, K., and Ajisaka, K. (2010). Enzymatic synthesis of lacto-N-difucohexaose I which binds to Helicobacter pylori. Methods Enzymol. *480*, 511–524.

Moracci, M., Trincone, A., and Rossi, M. (2001). Glycosynthases: new enzymes for oligosaccharide synthesis. J. Mol. Catal. B Enzym. *11*, 155–163.

Morrow, A.L., Ruiz-Palacios, G.M., Jiang, X., and Newburg, D.S. (2005). Human-milk glycans that inhibit pathogen binding protect breast-feeding infants against infectious diarrhea. J. Nutr. *135*, 1304–1307.

Murata, T., Inukai, T., Suzuki, M., Yamagishi, M., and Usui, T. (1999). Facile enzymatic conversion of lactose into lacto-N-tetraose and lacto-N-neotetraose. Glycoconj. J. *16*, 189–195.

Newburg, D.S. (2013). Glycobiology of human milk. Biochem. Biokhimiia *78*, 771–785.

Ninonuevo, M.R., and Bode, L. (2008). Infant formula oligosaccharides opening the gates (for speculation). Pediatr. Res. *64*, 8–10.

Ninonuevo, M.R., Park, Y., Yin, H., Zhang, J., Ward, R.E., Clowers, B.H., German, J.B., Freeman, S.L., Killeen, K., Grimm, R., et al. (2006). A strategy for annotating the human milk glycome. J. Agric. Food Chem. *54*, 7471–7480.

Nishimoto, M., and Kitaoka, M. (2007). Practical preparation of lacto-N-biose I, a candidate for the bifidus factor in human milk. Biosci Biotechnol Biochem *71*, 2101–2104.

Nishimoto, M., and Kitaoka, M. (2009). One-pot enzymatic production of β-d-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-d-galactose (galacto-N-biose) from sucrose and 2-acetamido-2-deoxy-d-galactose (N-acetylgalactosamine). Carbohydr. Res. *344*, 2573–2576.

Perez, P.F., Dore, J., Leclerc, M., Levenez, F., Benyacoub, J., Serrant, P., Segura-Roggero, I., Schiffrin, E.J., and Donnet-Hughes, A. (2007). Bacterial Imprinting of the Neonatal Immune System: Lessons From Maternal Cells? Pediatrics *119*, e724–e732.

Planas, A., Faijes, M., and Codera, V. (2015). When Enzymes Do It Better: Enzymatic Glycosylation Methods. In Carbohydrate chemistry:State of the Art and Challenges for Drug Development, p. Chapter 9.

Plante, O.J. (2001). Automated Solid-Phase Synthesis of Oligosaccharides. Science (80-.). *291*, 1523–1527.

Polonovski, M., and Montreuil, J. (1954). Etude chromatographique des polyosides du lait de femme. OMPTES RENDUS Hebd. DES SEANCES L Acad. DES Sci. *238*, 2263–2264.

Priem, B., Gilbert, M., Wakarchuk, W.W., Heyraud, A., and Samain, E. (2002). A new fermentation process allows

large-scale production of human milk oligosaccharides by metabolically engineered bacteria. Glycobiology *12*, 235–240.

Prieto, P.A. (2012). Profiles of Human Milk Oligosaccharides and Production of Some Human Milk Oligosaccharides in Transgenic Animals. Adv. Nutr. *3*, 456S – 464S.

Roberfroid, M. (2007). Prebiotics : The Concept Revisited 1 , 2. J. Nutr. *137*, 830S – 837S.

Rudloff, S. (2006). Incorporation of orally applied 13C-galactose into milk lactose and oligosaccharides. Glycobiology *16*, 477–487.

Ruiz-Palacios, G.M., Cervantes, L.E., Ramos, P., Chavez-Munguia, B., and Newburg, D.S. (2003). Campylobacter jejuni binds intestinal H(O) antigen (Fuc-alpha-1, 2Gal-beta-1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. J. Biol. Chem. *278*, 14112–14120.

Sallomons, E., Wilbrink, M.H., Sanders, P., Kamerling, J.P., Van Vuure, C.A., Hage, J.., and A., J. (2013). Methods for Providing Sialylated Oligosaccharides. WO Pat. 2013085384 A1.

Sano, M., Hayakawa, K., and Kato, I. (1992). An enzyme releasing lacto-N-biose from oligosaccharides. Proc. Natl. Acad. Sci. U. S. A. *89*, 8512–8516.

Schmölzer, K., Czabany, T., Luley-Goedl, C., Pavkov-Keller, T., Ribitsch, D., Schwab, H., Gruber, K., Weber, H., and Nidetzky, B. (2015). Complete switch from α-2,3- to α-2,6-regioselectivity in Pasteurella dagmatis β-d -galactoside sialyltransferase by active-site redesign. Chem. Commun. *51*, 3083–3086.

Schönfeld, H.E.R.B.E.R.T.. (1926). Über die Beziehungen der einzelnen Bestandteile der Frauenmilch zur Bifidusflora. Jahrb. Der Kinderh 16–90.

Sela, D.A., and Mills, D.A. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. Trends Microbiol *18*, 298–307.

Sela, D.A., Chapman, J., Adeuya, A., Kim, J.H., Chen, F., Whitehead, T.R., Lapidus, A., Rokhsar, D.S., Lebrilla, C.B., German, J.B., et al. (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc. Natl. Acad. Sci. U. S. A. *105*, 18964–18969.

Smilowitz, J., Lebrilla, C., Mills, D., German, J., and Freeman, S. (2014). Breast milk oligosaccharides: structure-function relationships in the neonate. Annu Rev Nutr *34*, 143–169.

Thurl, S., Muller-Werner, B., and Sawatzki, G. (1996). Quantification of individual oligosaccharide compounds from human milk using high-pH anion-exchange chromatography. Anal. Biochem. *235*, 202–206.

Tissier, H. (1900). Recherches sur la flore intestinale des nourrissons: état normal et pathologique (Paris, France).

Trincone, A., Perugino, G., Rossi, M., and Moracci, M. (2000). A novel thermophilic glycosynthase that effects branching glycosylation. Bioorganic Med. Chem. Lett. *10*, 365–368.

Tsuchida, A., Okajima, T., Furukawa, K., Ando, T., Ishida, H., Yoshida, A., Nakamura, Y., Kannagi, R., Kiso, M., and Furukawa, K. (2003). Synthesis of disialyl Lewis a (Lea) structure in colon cancer cell lines by a sialyltransferase, ST6GalNAc VI, responsible for the synthesis of alpha-series gangliosides. J. Biol. Chem. *278*, 22787–22794.

Underwood, M.A., Gaerlan, S., De Leoz, M.L.A., Dimapasoc, L., Kalanetra, K.M., Lemay, D.G., German, J.B., Mills, D.A., and Lebrilla, C.B. (2015). Human milk oligosaccharides in premature infants: absorption, excretion, and influence on the intestinal microbiota. Pediatr. Res.

Urashima, T., Saito, T., Nakamura, T., and Messer, M. (2001). Oligosaccharides of milk and colostrum in non-human mammals. Glycoconj. J. *18*, 357–371.

Urashima, T., Fukuda, K., and Messer, M. (2011). Evolution of milk oligosaccharides and lactose: a hypothesis. Animal *6*, 369–374.

Urashima, T., Asakuma, S., Leo, F., Fukuda, K., Messer, M., and Oftedal, O.T. (2012a). The Predominance of Type I Oligosaccharides Is a Feature Specific to Human Breast Milk. Adv. Nutr. *3*, 473S – 482S.

Urashima, T., Asakuma, S., Leo, F., Fukuda, K., Messer, M., and Oftedal, O.T. (2012b). The Glycobiology of Human Milk Oligosaccharides - The Predominance of Type I Oligosaccharides Is a Feature Specific to Human Breast Milk. Adv. Nutr. *3*, 473S – 482S.

Urashima, T., Taufik, E., Fukuda, K., and Asakuma, S. (2013). Recent Advances in Studies on Milk Oligosaccharides of Cows and Other Domestic Farm Animals. Biosci. Biotechnol. Biochem. *77*, 455–466.

Vázquez, E., Barranco, A., Ramírez, M., Gruart, A., Delgado-García, J.M., Martínez-Lara, E., Blanco, S., Martín, M.J., Castanys, E., Buck, R., et al. (2015). Effects of a human milk oligosaccharide, 2'-fucosyllactose, on hippocampal long-term potentiation and learning capabilities in rodents. J. Nutr. Biochem. *26*, 455–465.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T., and Yamamoto, K. (2008a). Bifidobacterium bifidum Lacto-N-Biosidase, a Critical Enzyme for the Degradation of Human Milk Oligosaccharides with a Type 1 Structure. Appl. Environ. Microbiol. *74*, 3996–4004.

Wada, J., Honda, Y., Nagae, M., Kato, R., Wakatsuki, S., Katayama, T., Taniguchi, H., Kumagai, H., Kitaoka, M., and Yamamoto, K. (2008b). 1,2-alpha-l-Fucosynthase: a glycosynthase derived from an inverting alpha-glycosidase with an unusual reaction mechanism. FEBS Lett. *582*, 3739–3743.

Wang, B. (2009). Sialic Acid Is an Essential Nutrient for Brain Development and Cognition. Annu. Rev. Nutr. *29*, 177–222.

Wang, L.-X., and Davis, B.G. (2013). Realizing the promise of chemical glycobiology. Chem. Sci. *4*, 3381–3394.

Wang, L.X., and Huang, W. (2009). Enzymatic transglycosylation for glycoconjugate synthesis. Curr. Opin. Chem. Biol. *13*, 592–600.

Weirich, A., and Hoffmann, G.F. (2005). Ernst Moro (1874–1951)—A great pediatric career started at the rise of university-based pediatric research but was curtailed in the shadows of Nazi laws. Eur. J. Pediatr. *165*, 599–606.

Woo, J.., Sohng, J.., Kim, B.., Kang, S.Y., Kim, D.H., Jang, K.S., Yang, J.Y., Jung, Y.S., Seo, W.M., and Gil, T.G. (2014). N-acetylglucosamine-2-epimerase and method for producing CMP-neuraminic acid using the same. US 8852891 B2.

Wu, S., Grimm, R., German, J.B., and Lebrilla, C.B. (2011). Annotation and structural analysis of sialylated human milk oligosaccharides. J. Proteome Res. *10*, 856–868.

Xiao, J., Takahashi, S., Nishimoto, M., Odamaki, T., Yaeshima, T., Iwatsuki, K., and Kitaoka, M. (2010). Distribution of in vitro fermentation ability of Lacto-N-Biose I, a major building block of human milk oligosaccharides, in Bifidobateria strains. Appl. Environ. Microbiol. *76*, 54–59.

Yamamoto, K. (2013). Recent advances in glycotechnology for glycoconjugate synthesis using microbial endoglycosidases. Biotechnol. Lett. *35*, 1733–1743.

Yu, H., Lau, K., Thon, V., Autran, C.A., Jantscher-Krenn, E., Xue, M., Li, Y., Sugiarto, G., Qu, J., Mu, S., et al. (2014). Synthetic disialyl hexasaccharides protect neonatal rats from necrotizing enterocolitis. Angew. Chemie - Int. Ed. *53*, 6687–6691.

Yu, Z.T., Chen, C., Kling, D.E., Liu, B., McCoy, J.M., Merighi, M., Heidtman, M., and Newburg, D.S. (2013). The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. Glycobiology *23*, 169–177.

Zivkovic, a. M., and Barile, D. (2011). Bovine Milk as a Source of Functional Oligosaccharides for Improving Human Health. Adv. Nutr. An Int. Rev. J. *2*, 284–289.

Zivkovic, A.M., German, J.B., Lebrilla, C.B., and Mills, D.A. (2011). Human milk glycobiome and its impact on the infant gastrointestinal microbiota. Proc. Natl. Acad. Sci. *108*, 4653–4658.

OBJECTIVES

OBJECTIVES

Given the important role of HMO in infant development, and the lack of an efficient synthetic approach for large scale production of these complex molecules, the current thesis envisages the chemo-enzymatic synthesis of lacto-*N*-tetraose (LNT, core type 1), one of the most abundant HMO. With the aim of engineering Lacto-*N*-biosidase form *B. bifidum* (LnbB) as an efficient biocatalyst for LNT synthesis, four main objectives are outlined in this thesis.

1. Synthesis of the key disaccharide Lacto-*N*-biose (LNB) and subsequent substrate donors. The 4-nitrophenyl (*p*-NP) derivative of LNB (*p*-NP-LNB) for the enhanced transglycosylation strategy, and 1,2-oxazoline derivative of LNB (oxa-LNB) for the glycosynthase strategy.

2. Kinetic characterization of Lacto-*N*-biosidase from *B. bifidum*, and identification of the catalytic residues.

3. Structural-functional analysis of GH20 enzymes to evaluate the minimal functional unit of LnbB that retains the enzymatic activity.

4. LnbB engineering, either as a glycosynthase-like or a transglycosidase, for the enzymatic synthesis of LNT.

CHAPTER 2.
SYNTHESIS OF THE SUBSTRATES

2.1 INTRODUCTION

Chemo-enzymatic glycosylation has become a powerful tool for stereo- and regio-selective synthesis of glycosidic bonds in carbohydrate chemistry. The chemo-enzymatic strategy mainly consists of two steps. The first one involves the synthesis of an activated glycosyl donor. And the second one consists of the subsequent enzyme-catalysed transglycosylation of that glycosyl donor to an acceptor (Noguchi *et al*., 2009).

The improvement of transglycosylation efficiency of natural glycosides can be achieved by protein engineering. Reaction engineering will modify reaction to enhance probability of acceptor condensation instead of water and also reduce of water activity. The strategies for protein engineering consist of rational design based on the structural information available of the active site of the enzyme, as well as directed evolution with using screening resources. Both strategies aim to modify the reaction and favour transglycosylation over hydrolysis to achieve efficient carbohydrate synthesis.

In this context, we aim the chemo-enzymatic synthesis of lacto-*N*-tetraose (LNT, Galβ-1,3GlcNAcβ-1,3Galβ-1,4Glc) using engineered Lacto-*N*-biosidase (LnbB) as biocatalyst. Which consists of, firstly, synthesizing the glycosyl donors (activated LNB) and, secondly, producing a generation of LnbB mutants which are believed to condense these donor disaccharides and the acceptor sugar, lactose, to achieve the target lacto-*N*-tetraose.

Subsequently, the generation of LnbB mutants are addressed to two different LNT synthesis approaches. The first one consists of enhancing transglycosylation activity by engineering reaction conditions and using *p*-NP-LNB as the activated glycosyl donor and also, performing mutations in acceptor recognition residues of LnbB. While the second one is based on the glycosynthase methodology, and consists of generating glycosynthase-like LnbB and providing the corresponding activated donor (oxa-LNB) to the new enzyme.

First we present the enzymatic synthesis of the key disaccharide LNB using a β-galactosidase from *Bacillus circulans.* β-Gal-3 is able to synthesise the target disaccharide (LNB) in high yields in a reaction mixture of ionic liquids in water solution.

Secondly, we evaluate the synthesis of donor substrates, *p*-NP-LNB and oxa-LNB, via two different synthetic procedures. The total synthesis 1 scheme (Figure 2.1) starts with the enzymatic production of Galβ-1,3GlcNAc (**3**) from *p*-NP-β-Gal (**1**) (donor) and GlcNAc (**2**) (acceptor) mediated by a β-galactosidase from *Bacillus circulans* (β-Gal-3) (Bayón *et al*., 2013). Subsequent substrate activation involves chemical synthesis. *p*-NP-LNB (**7**) synthesis starts from the peracetylated LNB converted into the corresponding glycosyl chloride, followed by glycosylation with 4-nitrophenol, and further deacetylation. oxa-LNB (**15**) synthesis is performed directly from LNB by using a newly designed dehydrative condensing agent. By contrast, the total synthesis 2 scheme (Figure 2.2) consists of the enzymatic synthesis of the substrates using β-Gal-3 as biocatalyst and the corresponding monosaccharides. In the case of *p*-NP-LNB (**7**) synthesis, *p*-NP-β-Gal (**1**) as donor and *p*-NP-GlcNAc (**8**) as acceptor were used. And with regard to oxa-LNB (**15**) synthesis, *p*-NP-β-Gal (**1**) as donor and oxa-GlcNAc (**14**) as acceptor were used.

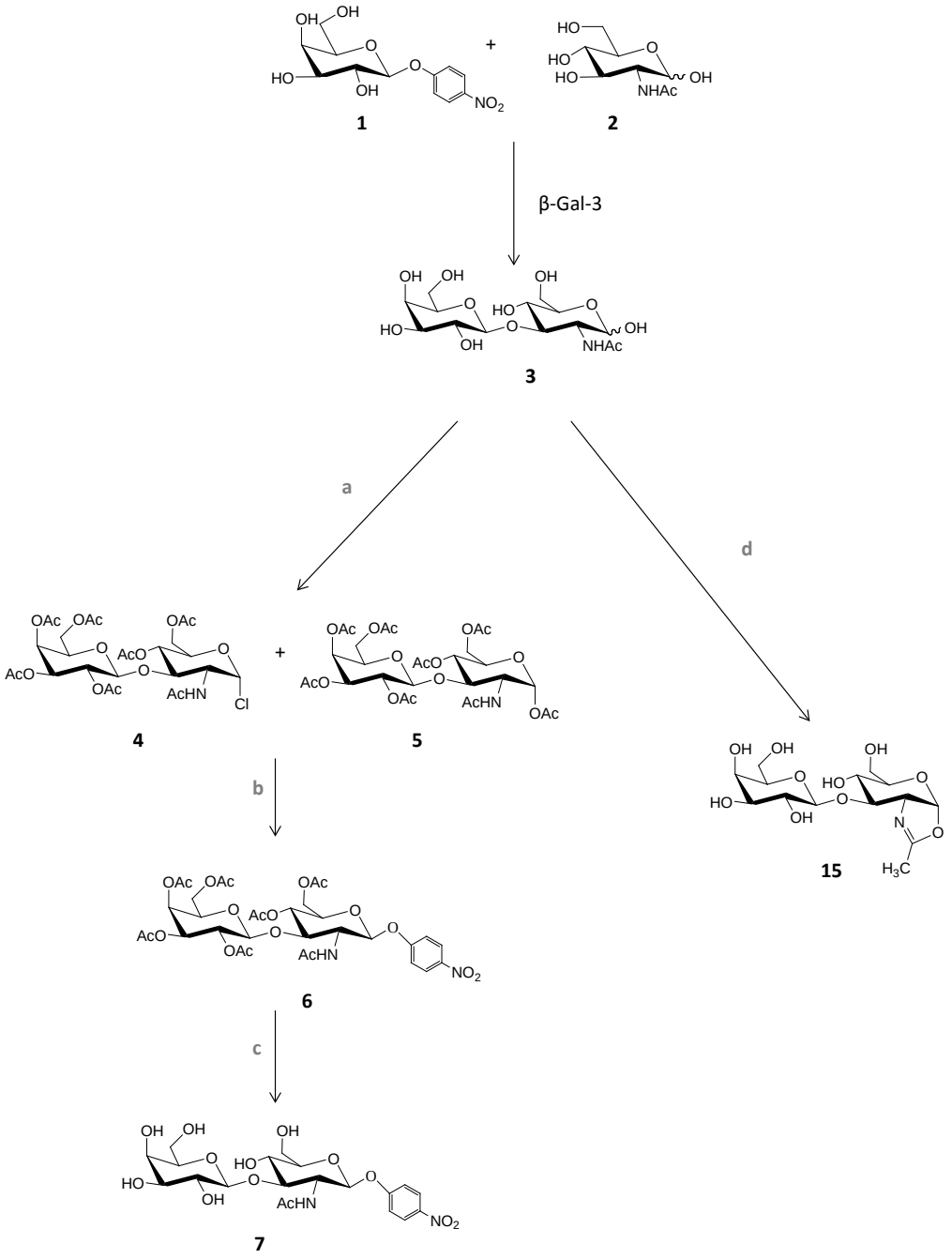


Figure 2.1 Total synthesis 1 scheme.

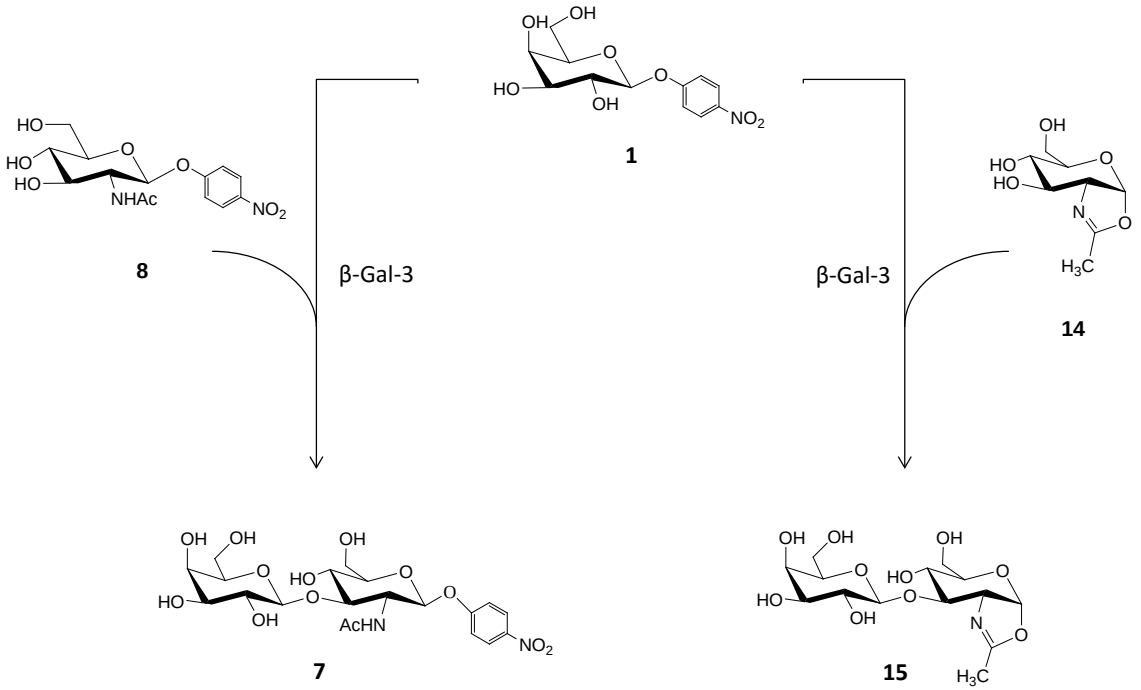


Figure 2.2 Total synthesis 2 scheme.

2.2 MATERIALS AND METHODS

2.2.1 Materials

Solvents

Cyclohexane, chloroform and dichloromethane (DCM) were distilled over $CaCl_2$. In some cases DCM and DCE was distilled a second time over P_2O_5 for a higher desiccation. Ethyl acetate was distilled over K_2CO_3. Pyridine was distilled over KOH. 1,2-dichloroethane (DCE) was first dried with $MgSO_4$, further distilled over P_2O_5 and immediately stored with freshly activated molecular sieves. MeOH was distilled over I_2 (1 g/L of methanol) and Mg (5 g/L of methanol). Molecular sieves were activated in the oven at 450 °C for 3 - 4 hours.

Reactants

Tetrabutylammonium hydrogen sulphate (TBAHS) was recrystallized from acetone. Sodium methoxide in methanol was prepared from Na (2.23 g, 1M) and anhydrous MeOH (100 mL). Amberlite® IR-120 (H+) resin was washed with MeOH until it was colourless.

Substrates

4-Nitrophenyl β-D-galactopyranoside (*p*-NP-Gal, ref N0616) and 4-Nitrophenyl 2-Acetamido-2-deoxy-β-D-glucopyranoside (*p*-NP-GlcNAc, ref N0866) were purchased from TCI America. *N*-acetil-β-D-glucosamina (GlcNAc, Ref. A8625) was purchased from Sigma-Aldrich.

2.2.2 General methods

Monitoring the reactions

Thin-layer chromatography (TLC) was used to follow the reactions. TLC was performed in aluminium sheets coated with regular silica gel (Merck 60 F_{254}) or with silica gel modified with aliphatic hydrocarbons ((Merck 60 RP-18 F_{254}). Once the products are eluted, TLC was analysed under uv light (254 and 360 nm) and then, revealed with the mixture H_2O/MeOH/H_2SO_4 (4.5:4.5:1) at 125 °C.

NMR

^1H and ^{13}C NMR, COSY and HSQC spectra were recorded in a Varian Gemini 400 MHz spectrometer operated at 298 K. For ^1H NMR, chemical shifts were referenced to internal TMS (δ_H=0 ppm) for solutions in $CDCl_3$ (δ_H= 7.26 ppm) and in D_2O (δ_H=4.79 ppm). For ^{13}C NMR, the central peak of the $CDCl_3$ triplet (δ=77.16 ppm) was used as a reference. Chemical shifts (δ, ppm) are indicated followed by the type of signal: s=singlet, d=doublet, t: triplet, dd: double doublet, ddd= double of doublets and m= multiplet. Spin-spin- coupling (J, Hz) and the number of protons according to electronic integration was also indicated.

Lyophilisation

The samples were frozen at -80 °C and then maintained at -50°C and under vacuum conditions to be lyophilised in the Cryodos-45 TELSTAR equipment.

2.2.3 Expression and purification of β-Gal-3 from *Bacillus circulans* ATCC 31382

Recombinant cultures of *E.coli* BL21 were grown in LB medium with kanamycin at final concentration of 30µg/ml for 16 hours, 250 rpm at 37 °C. Protein expression was induced adding IPTG at final concentration of 1mM and after 5 hours of incubation at 250 rpm and 37 °C. Cells were harvested by centrifugation (4,000xg, 15 min at 4 °C), resuspended in 100 mM of phosphate buffer at pH 6 (50 mM Na_2HPO_4, 50 mM NaH_2PO_4), and disrupted by ultrasonic disruption (25 cycles, 10 s on and 35 s off, 50% amplitude). Cell free extracts were centrifuged (14000xg, 4C, 15 min) and the supernatant was applied into a 5mL Ni^{2+}-charged HiTrap chelating column (Bio-Scale™ Mini Profinity™ IMAC Cartridges, Bio-Rad) chromatography. Proteins were purified to homogeneity according to manufacturer's protocol (ref. 10005455, Bio-Rad). Fractions were monitored by absorbance at 280 nm, pooled and concentrated and desalted in an Amicon ultracentrifuge filter (Millipore). The protein concentration was determined by the Bradford method using BSA as standard (Bradford, 1976).

2.2.4 Synthesis of lacto-*N*-biose by transglycosilation using β-Gal-3

Reaction conditions involved 85 mM of *p*-NP-Gal (donor), 425 mM of GlcNAc (acceptor) and 2.5 µM of protein in a final volume of 200 mL (splitted in 4x50 mL volumetric flasks). The reactions took place in 100mM of phosphate buffer at pH 6 (50 mM Na_2HPO_4, 50 mM NaH_2PO_4) with 30 % of ionic liquid ([BMIN][PF6]) at 37 °C. Aliquots were withdrawn at regular time intervals (1, 2, 2.5 and 3 hours) by heating at 80 °C for 10 min, diluted 1:10 in water and analysed by HPLC-ELSD. The reaction was stopped after 3 hours by heating at 80 °C for 30 min. The crude was centrifuged (4 cycles of 15 min at 8000 rpm) to recover the aqueous phase (containing carbohydrate compounds), which was, then, lyophilized to eliminate the water. The yield in the crude reaction crude was 90% (27 g).

Analysis of the reaction was performed by HPLC (Jasco 2000) using an amino column (NH2P50-4E, Asahipak, Japan) and two detectors: ELSD (Evaporative Light Scattering) and UV at 317 nm. Chromatographic conditions were 10 µL injection, flow rate of 0.8 mL/min of an isocratic elution ACN (80%) and H_2O (20%) at 50 °C. Detector conditions regarding ELSD were nebulizer at 60 °C, evaporation at 50 °C and gas pressure of 1.6 SLM. Each experimental assay was determined at least three times with a coefficient of variation of < 5%.

2.2.5 Purification of lacto-*N*-biose from the previous reaction using charcoal-Celite

Chromatography was performed on a mixture of charcoal (ref. 242276, Sigma-Aldrich) and Celite-503 (Merck) (50%, w/w). A mixture of 20:1 ratio (w/w), dry charcoal-Celite mixture: crude fraction was poured into the column and was packed by vacuum. Column was pre-eluted with H_2O at a flow rate of 1mL/L for an overnight. Then, the crude fraction was dissolved in a small amount of H_2O and loaded into the column. Products were eluted with milliQ water in linear gradient of ethanol (from 0% to 20%, v/v). Disaccharide enriched fractions were collected in 10% or 15% ethanol; samples were pooled, solvent was removed and then lyophilized, purity of the solid powder was analysed by HPLC. The yield of pure disaccharide (**3**) after purification through the charcoal-Celite column (compared with initial *p*-NP-LNB) was 90%, although the total chromatographic yield was 60%. Purity of the solid powder was analysed by NMR and HPLC-MS.

Compound **3**.

^{1}H NMR (400 MHz, D_2O) δ 5.19 (d, *J* = 3.5 Hz, 1H, H1^{I}α), 4.76 (d, *J* = 8.2 Hz, 0H, H1^{I}β), 4.13 – 2.95 (H2, H3, H4, H5 and H6), 2.04 (s, 3H, Ac).

^{13}C NMR (100 MHz, D_2O) δ 174.45 (C=O,$COCH_3$), 103.48/103.35 (C1 β/α^{II}), 94.63 (C1β^{I}), 90.95 (C1α^{I}), 82.52/80.08 (C3β/αII), 75.37, 75.20, 75.15, 72.47, 72.44, 71.14, 70.64, 70.60 (C5^{I}, C5^{II}, C3^{II}, C2^{II}), 68.63, 68.59, 68.47, 68.45 (C4^{I},C4^{II}), 60.92, 60.66, 60.49 (C6^{I},C6^{II}), 55.54 (C2β), 52.81 (C2α), 22.17/21.92(CH_3CO).

2.2.6 Purification of lacto-*N*-biose from the previous reaction crude using silica-gel

Flash chromatography was performed on Merck Silica gel (40–63 μm) in a 60:1 dry silica gel:crude ratio (w/w). Column was pre-eluted with the corresponding solvent previously mixed with triethylamine (Et_3N) (1mL/L). The crude fraction was dissolved in a small amount of solvent and loaded into the column. The crude fraction was acetylated, purified using flash chromatography and finally deacetylated.

Acetylation of the crude

Crude fraction (5 g) was reacted with acetic anhydride (Ac_2O, 5 mL, 52.5 mmol), pyridine (2.5 mL, 30 mmol) and 4-dimethylaminopyridine (DMAP, 2.5 mg, 0.02 mmol). The reaction was stirred at room temperature for 18 hours. Then, MeOH (1.5 vol of Ac_2O) was poured into the mixture once cold. After concentration and dilution with chloroform, the organic layer was washed with saturated aq. $NaHCO_3$ and water, dried over $MgSO_4$ and the organic solvent was evaporated under vacuum. In order to remove pyridine residues, the crude was co-evaporated with toluene. Per-*O*-acetated mixture (8.2 g) was obtained.

Silica-gel column chromatography

The per-*O*-acetated crude fraction was dissolved in a small amount of solvent (CH_2Cl_2) and loaded into the column. *p*-NP was first eluted with AcOEt/Cy 4:1. Then, elution was changed to AcOEt/Cy 10:1. And, Ac-GlcNAc (fractions 76-88) (**12**) and Ac-LNB (**16**) (fractions 92-208) were independently eluted). Disaccharide enriched fractions were pooled and solvent was removed by evaporation to dryness *in vacuo.* Ac-LNB (**16**) structure was confirmed by NMR. Chromatographic yield was 98%.

Compound **16**

^{1}H-RMN (400 MHz, $CDCl_3$): δ 6.07 (d, *J* = 3.7 Hz, 0.65H,H1^{I}α), 5.85 (d, *J* = 7.9 Hz, 0.35H, H1^{I}β), 5.76 (d, *J* = 8.8 Hz, 0.35H, NHα), 5.47 (d, *J* = 9.7 Hz, 0.65H, NHβ), 5.37 (dd, *J* = 3.3, 1.2 Hz, 1H, H4^{I}), 5.12-4.89 (m, 3H, H2^{II}, H3^{II},H5^{I}), 4.62 (dd, *J* = 7.6, 3.5 Hz, 0.65H, H2α^{I}), 4.54 (ddd, *J* = 10.4, 9.7, 3.7 Hz, 1H, H1^{II}), 4.31-3.88 (m, 8H,H6^{I}, H4^{II}, H5^{I}, H6^{II}, H3^{I}, H2β^{I}), 2.25-1.90 (m, 24H, CH_3CO).

^{13}C-RMN (100 MHz, $CDCl_3$): δ 171.3-168.2 ($COCH_3$), 100.9/100.8 (C1β/α^{II}) ,91.7 (C1β^{I}) ,91.2 (C1α^{I}), 75.7 (C3β^{I}), 72.7 (C3α^{I}), 70.8, 70.3, 69.8, 69.0, 68.1 (C5α^{I}, C5^{II}, C3^{II}, C5β^{I}, C2^{II}) , 66.6 (C4^{I}), 64.4 (C4^{II}) , 61.8 , 60.8, 60.4 (C6^{I},C6^{II}), 54.4 (C2β), 51.1 (C2α) , 23.45-20.2 (CH_3CO).

2.2.6.1 Zemplen deacetylation of Ac-LNB (16)

Freshly prepared sodium methoxide (1M in methanol, 100 µL) was added to a stirred solution of peracetylated LNB (**16**) (Ac-LNB) in anhydrous methanol. Reaction conditions are detailed in Table 2.2 (section 2.3.2.1). Reactions were neutralized with Amberlite® IR-120 (H+) resin, filtered and analysed by TLC (ACN:H2O, 7:3 as eluent).

2.2.6.2 Methanolic amonia deacteylation of Ac-LNB (16)

Peracetylated LNB (**16**) (Ac-LNB, 100 mg, 0.14 mmol) or peracetylated GlcNAc (**12**) (Ac-GlcNAc, 100 mg, 0.24 mmol) were suspended with methanolic ammonia (NH_3/MeOH, 2mL) and methanol (2 mL). The solutions were stirred at room temperature for 50 hours. The reaction mixtures were then evaporated to dryness *in vacuo* and analysed by TLC (ACN:H_2O, 7:3 as eluent).

2.2.7 Enzymatic synthesis of 4-nitrophenyl 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside (7) using β-Gal-3

Reaction with free enzyme
Reaction conditions involved 85 mM of *p*-NP-Gal, 425 mM or 850 mM of *p*-NP-GlcNAc (donor:acceptor molar ratio, 1:5 or 1:10) and 2 nM of protein in a final volume of 100 µL. The reactions took place in 100 mM of phosphate buffer at pH 6 (50 mM Na_2HPO_4, 50 mM NaH_2PO_4) with different amounts of ionic liquids at 37 °C. At different time intervals, 1 and 3 hours, 20 µL of samples were analysed using HPLC as described in section 2.2.4.

Reaction with immobilised enzyme
Immobilization was carried out with 0.5 mg (7.4 nmol) of enzyme diluted in 5 mL of 100 mM of $NaHCO_3$ buffer at pH 10. In order to avoid conformational changes in the protein due to the high pH, a stabilizing agent was added to the solution (50% of PEG-600 (Polyethyleneglycol). Then, 1 g of 10 % glyoxyl-agarose was added to the solution and was stirred at 25 °C for 30 min. The immobilization was finished adding 5 mg of $NaBH_4$ and the solution was further stirred for 30 min at 25 °C. The final mixture was filtered and washed with 50 mM of phosphate buffer at pH 6.0.

Enzymatic synthesis of *p*-NP-LNB (**7**) using glyoxyl-agarose immobilised enzyme consisted of 85 mM of *p*-NP-Gal, 425 mM of *p*-NP-GlcNAc (1:5 donor:acceptor molar ratio), and 30 mg of immobilised enzyme (0.03 mg of protein) in a final volume of 100 µL (4.4 µM of protein). Enzymatic reactions were performed in 100 mM phosphate buffer pH 6.0 with or without 30% of BMIN [PF6] at 37 °C. At different time intervals, 1 and 3 hours, 20 µL of samples were analysed using HPLC as described in section 2.2.4.

2.2.8 Enzymatic synthesis 1,2-oxazoline derivative of 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (15)

Reactions containing 85 mM of *p*-NP-Gal and 425 mM oxa-GlcNAc (donor:acceptor molar ratio, 1:5) were performed in 100 mM phosphate buffer at different conditions (pH 7.5 and 10 in the presence or absence of PEG and of BMIN [PF6]) in a final volume of 200 µL and at 37 °C. At different time intervals (10 and 30 min, and 1 and 3 hours) 20 µL aliquots were analysed using HPLC-ELSD.

2.2.9 Chemical synthesis of 4-nitrophenyl β-*N*-acetyl-D-glucosamine (8)

Anomeric carbon activation via chlorination

N-acetyl-D-glucosamine (**2**) (5 g, 22.5 mmol) was reacted with acetyl chloride (30mL, 420 mM) in a three-necked volumetric flask with a $CaCl_2$ tube. The reaction was kept at stirring conditions and under nitrogen atmosphere for 24 hours. The reaction was diluted with DCM (50 mL), washed with aq satd $NaHCO_3$ (4x100 mL) and water (3x100mL). The organic layer was dried with $MgSO_4$, filtered and concentrated. Yellow oil was obtained. It corresponded with a mixture of 60:40 fully acetylated α-chloride (**9**) and α-per-*O*-acetated *N*-acetyl-D-glucosamine (**10**) (5.3 g, 55% yield). This mixture was used directly for the glycosylation reaction with 4-nitrophenol.

^{1}H-RMN (400 MHz, $CDCl_3$): 6.19 (d, *J*= 3.7 Hz, 1H, Cl-GlcNAc), 6.17 (d, *J* = 3.7 Hz, 1H, Ac-GlcNAc), 5.81 (d, *J* = 8.7 Hz, 1H, NH-Cl-GlcNAc), 5.56 (d, *J* = 8.9 Hz, 1H, NH-Ac-GlcNAc), 3.89 – 5.34(m, 12H, H2- H6), 1.90-2.25 (m, 27H, CH_3CO).

4-nitrophenol glycosylation

A mixture of fully acetylated α-chloride (**9**) and α-per-*O*-acetated *N*-acetyl-D-gucosamine (**10**) (1.15 g, considering 60% of **9**, 0.68 mg, 1.82 mmol), anhydrous DCM (5 mL), TBAHS (635.6 mg, 1.82 mmol), 4-nitrophenol (520.8 mg, 3.74 mmol) and NaOH 1M (5 mL) was stirred at room temperature for 3 hours. The resulting crude was diluted with anhydrous DCM (50 mL) and washed with NaOH (200mL) and water (1 L) to get rid of unreacted 4-nitrophenol. The organic layer was dried with $MgSO_4$, filtered and concentrated. The crude was recrystallized from MeOH, and *p*-NP per-*O*-acetated *N*-acetyl-D-gucosamine (**11**) was obtained as a white solid in 57% yield (390 mg).

^{1}H-NMR (400 MHz, $CDCl_3$): δ 8.23 (d, *J* = 9.3 Hz, 1H, Ar-H), 7.29 (d, *J* = 9.2 Hz, 1H, Ar-H), 5.77 (d, *J* = 8.4 Hz, 1H, H1), 5.48-5.43 (m, 2H,H1,H3), 5.15 (dd, *J* = 9.9, 9.2 Hz, 1H, H4), 4.29 (dd, *J* = 12.3, 5.6 Hz, 1H, H6a), 4.17-4.14 (m, 1H, H6b), 4.13-4.11 (m, 1H, H2), 3.94 (ddd, *J* = 9.9, 5.6, 2.5 Hz, 1H, H5), 2.08-1.84 (12H, CH_3CO), 1.96 (s, 3H, $NHCOCH_3$).

Zemplen deacetylation

Freshly prepared sodium methoxide (1M in methanol, 100 µL) was added to a stirred solution of *p*-NP per-*O*-acetated *N*-acetyl-D-gucosamine (**11**) (1g, 2 mmol) in anhydrous methanol (25 mL), and the mixture was stirred at RT for 18 hours. A white solid and a yellow solution were obtained. The white solid was filtered, and the yellow solution was neutralized with Amberlite® IR-120 (H+) resin and filtered. White powder was then lyophilized to give 4-Nitrophenyl 2-Acetamido-2-deoxy-β-D-glucopyranoside (**8**) in quantitative yield.

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.22 (d, *J* = 9.3 Hz, 1H, Ar-H), 7.19 (d, *J* = 9.3 Hz, 1H, Ar-H), 5.22 (d, *J* = 8.4 Hz, 1H, H1), 4.04 – 3.88 (m, 2H), 3.76 – 3.22 (m, 9H), 1.99 (s, 3H).

^{13}C-NMR (100 MHz, Methanol-d_4): δ 169.8 ($COCH_3$), 125.7-116.0 (Ar-C), 98.8 (C1), 77.8 (C3), 74.3 (C5), 71.5 (C4), 61.9 (C6), 55.7 (C2), 23.5 ($NHCH_3CO$).

2.2.10 Chemical synthesis of 4-nitrophenyl 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside (7)

Anomeric carbon activation via chlorination

2-Acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (**3**) (LNB, 380 mg, 0.99 mmol) was reacted with acetyl chloride (4mL, 56.7 mM) in a three-necked volumetric flask with a $CaCl_2$ tube. The reaction was kept at stirring conditions and under nitrogen atmosphere for 22 hours. The reaction was diluted DCM (20 mL), washed with aq satd $NaHCO_3$ (4x40 mL) and water (3x40mL). The organic layer was dried with $MgSO_4$, filtered and concentrated. Brown oil was obtained. It corresponded with a mixture of 50:50 fully acetylated α-chloride (**4**) and α-per-*O*-acetated (**5**) LNB (404 mg, 53% yield).

^{1}H-NMR (400 MHz,$CDCl_3$): 6.04 (d, *J*=3.7 Hz, 1H, αH1^{I} Cl-LNB), 6.00 (d, *J*= 3.7 Hz, 1H, αH1^{I} Ac-LNB), 5.80 (d, *J*= 8.9 Hz, 1H, NH-Cl-LNB), 5.76 (d, *J*=8.9 Hz, 1H, NH-Ac-LNB), 4.75-3.83 (m, 10 H, H2-H6), 2.23 (m, 42 H, CH3CO), 1.78 (s, 3H, CH3CONH).

^{13}C-RMN (100 MHz, $CDCl_3$): δ 171.3-168.2 ($COCH_3$), 100.9/100.8 (C1β/α^{II}),91.7 (C1β^{I}),91.2(C1α^{I}), 75.7 (C3β^{I}), 72.7 (C3α^{I}), 70.8, 70.3, 69.8, 69.0, 68.1 (C5α^{I}, C5^{II}, C3^{II}, C5β^{I}, C2^{II}), 66.6 (C4^{I}), 64.4 (C4^{II}), 61.8, 60.8, 60.4 (C6^{I},C6^{II}), 54.4 (C2β), 51.1 (C2α), 23.45-20.2 (CH_3CO).

4-nitrophenol glycosylation

A mixture of fully acetylated α-chloride (**4**) and α-per-*O*-acetated (**5**) 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-D-glucopyranose (404 mg, considering 50% of **4**, 202 mg, 0.55 mmol), anhydrous DCM (1.4 mL, 18.38 mmol), TBAHS (187.42 mg ,0.55 mmol), 4-nitrophenol (153.7 mg, 1.104 mmol) and NaOH 1M (1.4 mL, 49.5 mmol) was stirred at room temperature for 3 hours. The resulting crude was diluted with anhydrous DCM (10 mL) and washed with NaOH (40 mL) and water (200 mL) to get rid of unreacted 4-nitrophenol. The organic layer was dried with $MgSO_4$, filtered and concentrated. The crude was recrystallized from MeOH, and 4-nitrophenyl 2-acetamido-4,6-di-*O*-acetyl-2-deoxy-3-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside (**6**) was obtained as a white solid in 49% yield (197.9 mg, 0.25 mmol).

^{1}H-NMR (400 MHz,$CDCl_3$): δ 8.21-8.16 (m, 2H, Ar-H), 7.10-7.04 (m, 2H, Ar-H), 6.04 (d, *J*=3.7 Hz, 1H, αH1^{I} Cl-LNB), 6.00 (d, *J*= 3.7 Hz, 1H, αH1^{I}, Ac-LNB), 5.80 (d, *J*= 8.9 Hz, 1H, NH-Cl-LNB), 5.76 (d, *J*=8.9 Hz, 1H, NH-Ac-LNB), 4.75-3.83 (m, 10 H, H2-H6), 2.23 (m, 42 H, CH3CO), 1.78 (s, 3H, CH3CONH).

^{13}C-RMN (100 MHz, $CDCl_3$): δ 170.7-169.1 ($COCH_3$), 161.4 (NHAc), 100.9/100.8 (C1β/α^{II}), 125.7-116.5 (Ar-C), 91.7 (C1β^{I}), 91.2 (C1α^{I}), 75.7 (C3β^{I}), 72.7(C3α^{I}), 70.8, 70.3, 69.8, 69.0, 68.1 (C5α^{I}, C5^{II}, C3^{II}, C5β^{I}, C2^{II}), 66.6 (C4^{I}), 64.4 (C4^{II}), 61.8, 60.8, 60.4 (C6^{I},C6^{II}), 54.4 (C2β), 51.1(C2α), 23.45-20.2 (CH_3CO).

Zemplen deacetylation

Freshly prepared sodium methoxide (1M in methanol, 100 μL) was added to a stirred solution of **6** (157 mg, 0.2 mmol) in anhydrous methanol (4 mL), and the mixture was stirred at RT for 18 hours. A white solid and a yellow solution were obtained. The white solid was filtered, and the yellow solution was neutralized with Amberlite® IR-120 (H+) resin and filtered. The white

powder was lyophilized to give 4-Nitrophenyl 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside (**7**) (43,7 mg, 0.18 mmol) in 75% yield.

^{1}H NMR (400 MHz, D_2O) δ 8.16 – 8.11 (m, 1H, Ar-H), 7.11 – 7.05 (m, 1H, Ar-H), 5.24 (d, *J* = 8.5 Hz, 1H,H-1), 4.35 (d, *J* = 7.7 Hz, 1H), 4.06 (dd, *J* = 10.4, 8.5 Hz, 1H), 3.87 – 3.77 (m, 2H), 3.73 – 3.38 (m, 6H), 2.09 (s, 11H), 1.88 (s, 1H, $NHCOCH_3$).

2.2.11 Synthesis of 1,2-oxazoline derivative of *N*-acetyl-D-glucosamine (14)

*2.2.8.3.1 Synthesis of 1,2-oxazoline derivative of N-acetyl-D-glucosamine (**14**) starting from the peracetylated sugar*

To a solution of *N*-acetyl-D-glucosamine (**2**) (2 g, 9 mmol) was added acetic anhydride (Ac_2O, 2 mL, 21 mmol), pyridine (1 mL, 12 mmol) and 4-dimethylaminopyridine (DMAP, 1 mg, 0.008 mmol). The reaction was stirred at room temperature for 18 hours. Then, MeOH (1.5 vol of Ac_2O) was poured into the mixture once cold. After concentration and dilution with chloroform, the organic layer was washed with saturated aq. $NaHCO_3$ and water, dried over $MgSO_4$ and the solvent was evaporated at reduced pressure. In order to remove pyridine residues, the crude was co-evaporated with toluene. Per-*O*-acetated *N*-acetyl-D-glucosamine (**12**) (2.7 g, 4.4 mmol) was obtained as a white powder in 90% yield.

^{1}H-RMN (400 MHz, $CDCl_3$): δ 6.16 (d, *J* = 3.4 Hz, 1H, H1α), 5.68 (d, *J* = 8.8 Hz, 0.09H, NHβ), 5.58 (d, *J* = 8.8 Hz, 0.91H, NHα), 5.27-5.15 (m, 2H, H3, H4), 4.52-4.43 (m, 1H, H2), 4.24 (dd, *J* = 12.5, 4.1 Hz, 1H, H6a), 4.06 (dd, *J* = 12.5, 2.4 Hz, 1H, H6b), 4.02-3.96 (m, 1H, H5), 2.17 (dd, *J* = 10.3, 0.4 Hz, 6H, CH_3CO), 2.12-2.01 (m, 10H), 1.93 (d, *J* = 0.4 Hz, 3H, $NHCOCH_3$).

^{13}C-RMN (100 MHz, $CDCl_3$): δ 171.7-168.6 ($COCH_3$), 90.7 (C1β), 70.7 (C3), 69.7 (C5), 67.4(C4), 61.5 (C6), 51.0 (C2), 23.0-20.5 (CH_3CO).

Using trimethylsilyl bromide and boron trifluoride: Per-*O*-acetated *N*-acetyl-D-glucosamine (**12**) (120 mg, 0.3 mmol) was dissolved in anhydrous DCE (10 mL), and then trimethylsilyl bromide (TMSBr) (197 µL, 1.5 mmol), $BF_3{\cdot}Et_2O$ (163 µL, 1.5 mmol) and 2,4,6-collidine (187 µL, 1.5 mmol) were added sequentially under nitrogen atmosphere. The mixture was stirred at room temperature for 12 h and then diluted with chloroform. The organic layer was washed with saturated aq. $NaHCO_3$ and water, dried over $MgSO_4$ and evaporation of the solvent under reduced pressure. Reaction was followed by TLC (AcOEt).

Using trimethylsilyl trifluoromethanesulfonate: A solution of per-*O*-acetated *N*-acetyl-D-glucosamine (**12**) (1g, 2.6 mmol) in 31 mL of DCE was heated at 50 °C under nitrogen atmosphere, followed by the dropwise addition of trimethylsilyl trifluoromethanesulfonate (TMSOTf, 550 µl, 3.03 mmol). After 20 h, the reaction was neutralised by slowly adding triethylamine (Et_3N, 1 mL) at 0 °C and stirring conditions for 10 min. After concentration and dilution with chloroform, the organic layer was washed with saturated aq. $NaHCO_3$ and water, dried over $MgSO_4$ and evaporation of the solvent under reduced pressure. Per-*O*-acetated 1,2-oxazoline derivative of *N*-acetyl-D-glucosamine (**13**) as an orange-ish oil was obtained (1.8 g).

^{1}H-RMN (400MHz, $CDCl_3$): δ 5.97 (d, *J* = 7.4 Hz, 1H, H1oxa), 5.27 (t, *J* = 2.5 Hz, 1H, H3), 4.94 (ddd, *J* = 9.3, 2.1, 1.2 Hz, 1H, H4), 4.19-4.17 (m, 2H,H2, H6a), 4.14 (ddq, *J* = 6.0, 3.1, 1.6 Hz, 1H,H6b), 3.65 – 3.57 (m, 1H, H5), 2.19 – 2.01 (m, 12H,CH_3CO).

^{13}C-RMN (100 MHz, $CDCl_3$): δ 170.6-166.6 ($COCH_3$), 99.4 (C1oxa), 70.4 (C3), 68.4 (C5), 67.5 (C4), 64.9 (C6), 63.3 (C2), 20.9-13.9 (CH_3CO).

Zemplen deacetylation

Freshly prepared sodium methoxide (1M in methanol, 450 µL) was added to a stirred solution of **13** (1g) in anhydrous methanol (100 mL), and the mixture was stirred at 0 °C for 1 hour (final 5 mM NaOMe concentration). Solvent was removed by evaporation *in vacuum* without previous neutralization, and the oil obtained was dissolved in water and lyophilised. An orange solid corresponding to 1,2-oxazoline derivative of *N*-acetyl-D-glucosamine (**14**) was obtained with an overall yield of 60% (1 g, 4.9 mmol).

^{1}H-RMN (400MHz, D_2O):) δ 6.10 (d, *J* = 7.3 Hz, 1H, H1oxa), 4.14 (dddd, *J* = 7.3, 3.8, 1.8, 1.2 Hz, 1H, H3), 3.99 (dd, *J* = 3.8, 3.3 Hz, 1H, H4), 3.83 (dd, *J* = 12.4, 2.5 Hz, 1H, H6b), 3.68 (dd, *J* = 12.4, 6.6 Hz, 1H, H6a), 3.62 (ddd, *J* = 9.0, 3.3, 1.2 Hz, 1H,H2), 3.40 (dddd, *J* = 9.0, 6.5, 2.5, 0.6 Hz, 1H, H5), 2.07 (d, *J* = 1.7 Hz, 3H, CH_3CO).

^{13}C-RMN (100 MHz, D_2O): 168.0 $COCH_3$, 100.4 (C1oxa), 72.9 (C3), 71.6 (C5), 68.6 (C4), 65.8 (C6), 61.6 (C2), 12.9 (CH_3CO).

*2.2.8.3.2 Synthesis of 1,2-oxazoline derivative of N-acetyl-D-glucosamine (**14**) starting from the unprotected sugar*

Using 2-chloro-1,3-dimethylimidazolinium chloride (DMC): N-acetyl-D-glucosamine (**2**) (100 mg, 495 µmol) with Et_3N (620 µL, 4.4 mmol) were dissolved in 1.98 mL of 10% D_2O in water at 0 °C and stirring conditions. Then DMC (251 mg, 1.5 mmol) was added and let to react for 15 min. NMR analysis showed the formation of 1,2-oxazoline derivative of *N*-acetyl-D-glucosamine (**14**) in 90% yield. DMI product (that comes from DMC after reaction) was also present in the final product since it was not possible to be removed.

Using 2-chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride (CDMBI): a solution of *N*-acetyl-D-glucosamine (**2**) (20 mg, 94.6 µmol) with Na_3PO_4 (128 mg, 709 µmol) in 630 µL of 10% D_2O in water, CBMBI (61.5 mg, 238.6 µmol) was added. The solution was stirred at room temperature for 1 h, then, centrifuged to removed DMBI (product from CDMBI reaction), and finally, lyophilized to give 1,2-oxazoline derivative of *N*-acetyl-D-glucosamine (**14**) in 80% yield (16 mg, 77.12 µmol). There was no DMBI in the final crude, but an excess of salt remained in the final product. Before further transglycosylation reactions, solution of oxa-LNB should be dropwise neutralized with HCl to decrease Na_3PO_4 resultant basic pH.

^{1}H-RMN (400MHz, D_2O):) δ 6.10 (d, *J* = 7.3 Hz, 1H, H1oxa), 4.14 (dddd, *J* = 7.3, 3.8, 1.8, 1.2 Hz, 1H, H3), 3.99 (dd, *J* = 3.8, 3.3 Hz, 1H, H4), 3.83 (dd, *J* = 12.4, 2.5 Hz, 1H, H6b), 3.68 (dd, *J* = 12.4, 6.6 Hz, 1H, H6a), 3.62 (ddd, *J* = 9.0, 3.3, 1.2 Hz, 1H,H2), 3.40 (dddd, *J* = 9.0, 6.5, 2.5, 0.6 Hz, 1H, H5), 2.07 (d, *J* = 1.7 Hz, 3H, CH_3CO).

^{13}C-RMN (100 MHz, D_2O): 168.0 $COCH_3$, 100.4 (C1oxa), 72.9 (C3), 71.6 (C5), 68.6 (C4), 65.8 (C6), 61.6 (C2), 12.9 (CH_3CO).

2.2.12 Synthesis of 1,2-oxazoline derivative of 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (15)

Using 2-chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride (CDMBI): a solution of 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (**3**) (LNB, 16 mg, 41 µmol) with Na_3PO_4 (55.7 mg, 307.5 µmol) in 800 µL of 10% D_2O in water, CBMBI (27 mg, 123 µmol) was added. The solution was stirred at 4 °C for 90 min. After that, reaction was centrifuged to removed DMBI (product from CDMBI reaction), and finally, lyophilized to give 1,2-oxazoline derivative of LNB (**15**) in 80% yield (13 mg, 33.3 µmol). There was no DMBI in the final crude, but an excess of salt remained in the final product. Before further transglycosylation reactions, solution of oxa-LNB should be dropwise neutralized with HCl (5N) to decrease Na_3PO_4 resultant basic pH.

^{1}H NMR (400 MHz, D_2O) δ 6.13 (d, *J* = 7.4 Hz, 0.8H, $H1^{I}$-oxa), 5.28 (d, *J* = 3.6 Hz, 0.2H, $1H^{I}$-α), 4.61 (d, *J* = 7.9 Hz, 1H, $H1^{II}$), 4.19 (ddd, *J* = 7.2, 3.3, 1.7 Hz, 1H), 4.04 (t, *J* = 3.0 Hz, 1H), 3.83 – 3.50 (m, 11H), 3.28 – 3.22 (m, 1H), 1.95 (t, *J* = 2.0 Hz, 2H).

2.2.13 Stability studies of 1,2-oxazoline derivative of *N*-acetyl-β-D-glucosamine

Different solutions of **14** in 50 mM citrate-phosphate buffer with 10% D_2O in a pH range of 3.5-9.0 were studied. At WET_1D NMR conditions, three signals were followed to determine the evolution of the hydrolysis of the oxazoline ring. oxa-GlcNAc is followed with the signal at 5.95 ppm (*J* = 7.4 Hz), and the α-GlcNAc and β-GlcNAc formed due to hydrolysis were followed at 5.05 ppm (*J* = 3.7 Hz) at 4.50ppm (*J* = 8.9 Hz) respectively (see Figure 2.12).

2.3 RESULTS AND DISCUSSION

2.3.1 Enzymatic synthesis of LNB by β-Gal-3

LNB (**3**) was prepared as starting material for the synthesis of the target activated donors *p*-NP-LNB (**7**) and oxa-LNB (**15**) according to the total synthesis scheme 1. The reaction was catalysed by β-galactosidase (β-Gal3) from *Bacillus circulans* ATCC 31382 using *p*-NP-Gal (as donor) and GlcNAc (as acceptor) to yield LNB (Galβ-1,3GlcNAc).

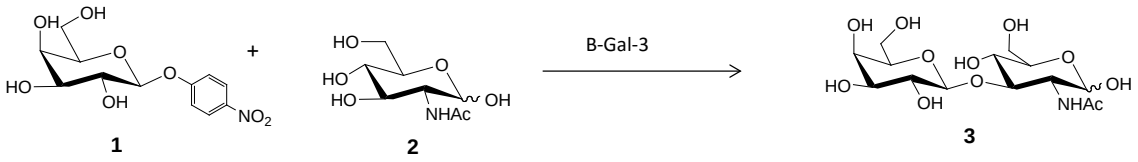


Figure 2.3 Enzymatic synthesis of Lacto-*N*-biose (3) by β-Gal-3 transglycosilation.

β-Gal-3 was first identified by Ito and coworkers. This enzyme, with a molecular size of 67 kDa, belongs to the family GH35 glycosyl hydrolases. β-Gal-3 hydrolyses β-1,3 linkages, but also performs transglycosylation (Fujimoto *et al*., 1998). By using different ionic liquids, Hernáiz and coworkers reported the synthesis of LNB mediated by β-Gal-3 transglycosylation with yields of 99 % (Bayón *et al*., 2013).

With the collaboration of Dr. Hernáiz (Department of Organic and Pharmaceutical Chemistry, Faculty of Pharmacy, Complutense University of Madrid), β-Gal3 was successfully expressed to obtain N-terminally His-tagged recombinant protein in a yield of 10 mg of protein per litre of culture.

The transglycosylation reaction was performed at 85 mM of *p*-NP-Gal, 425 mM of GlcNAc and 2.5 µM of protein in 100 mM phosphate buffer pH 6.0 and 30% of [BMIN][PF6] at 37 °C. Reaction was followed by HPLC using two detectors: ELSD (Evaporative Light Scattering) and UV at 317 nm. Final recovery of 27 grams of dried crude was obtained.

Figure 2.4 shows the chromatograms at different reaction times. According to standards, peaks before 3 min corresponds to [BMIN][PF6] and salts. The peak at 4 min corresponds to *p*-NP-Gal, which decreases while the reaction evolves. The peak at 10 min corresponds to GlcNAc, which is fully saturated during the whole analysis due to the high amount added to the reaction (5-fold excess of acceptor compared to the donor). During the reaction, a new peak appears at 17 min, which corresponds to LNB. As is observed, the LNB peak is maximum at 2.5 hours and, from then, *p*-NP-Gal had completely reacted and the enzyme hydrolysed the disaccharide product. For this reason, reaction yield of this scale-up reaction was 78% instead of the reported 98% (Bayón et al., 2013).

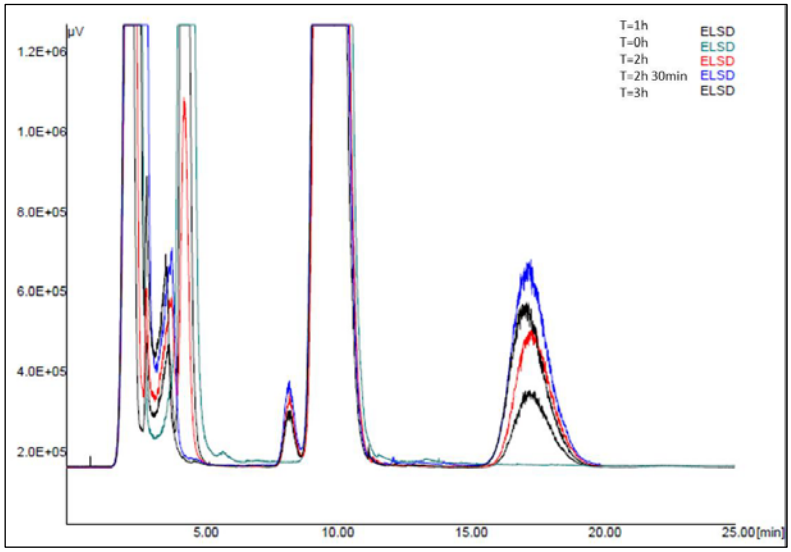


Figure 2.4 Transglycosylation reaction followed using HPLC-ELSD.

Purification of Lacto-*N*-biose from the enzymatic crude

It was expected that the enzymatic crude was composed of LNB, GlcNAc and *p*-NP in a ratio of 1:4:1 according to reaction conditions. In addition, TLC analysis confirmed traces of Gal and impurities.

Two different strategies were used for the purification of the Lacto-*N*-biose. The first one was directly performed with the crude that contained unprotected sugars using a charcoal-celite column. The second one was performed with the acetylated sugars and involved three steps, sugar acetylation, purification using a silica-gel column, and deacetylation.

Purification of Lacto-N-biose using charcoal-celite

The crude was dissolved in water (<2% of the total volume of the column) and loaded into the charcoal-celite column. The elution was performed as Table 2.1 indicates. LNB, eluted with 5% of ethanol, could be purified from GlcNAc and impurities, and it was identified by TLC analysis (ACN:H_2O, 80:20). Enriched carbohydrate fractions were lyophilised and stored at -20 °C.

Table 2.1 Elution gradient of charcoal-celite chromatography.

Eluent	Recovered product	Elution volume (L)
H2O	GlcNAc	3
1% EtOH	-	1
2% EtOH	Impurities	0.5
5% EtOH	LNB	3
20% EtOH	Impurities	1

Column recovery was 60% of the total crude loaded into the column. LNB recovery yield was 70 %, since at least, *p*-NP did not elute from the column due to the hydrophobic properties of the charcoal-celite stationary phase. The structure of LNB was confirmed by the characteristic

signals using NMR analysis (δ 5.19 ppm (d, *J* = 3.5 Hz, 1H, $H1^{I}\alpha$)), and by MS analysis (MS m/z 384 $[M+H]^+$; 406 $[M+Na]^+$) .

Purification of Lacto-N-biose using silica-gel

The enzymatic crude was acetylated in quantitative yield. Flash chromatography was performed on silica-gel using a gradient of AcOEt/Cy as eluent in 95-98% yield. The elution was as it follows: *p*-NP, peracetylated GlcNAc (Ac-GlcNAc) and peracetylated LNB (Ac-LNB). Carbohydrate enriched fractions were evaporated and stored at -20 °C. The corresponding structures were confirmed by NMR analysis (see the experimental section).

Deacetylation with NaOMe/MeOH was performed to deacetylate the peracetylated LNB. Standard reaction conditions involving low NaOMe/Ac-LNB molar ratio (< 1), 24 hours of reaction and at room temperature (Table 2.4, entries 1 and 2), rendered hydrolysis of the disaccharide to yield the corresponding monosaccharides. These results were not expected, since these conditions are extensively used for deacetylation of β-1,3 and β-1,4 sugars. It suggests that oligosaccharides with β-1,3 linkages and an acetamide group in C-2 position are not stable at high basic conditions. Subsequently, different reaction conditions were studied: NaOMe concentrations, Ac-LNB/NaOMe molar ratio, temperatures and reaction time. Reactions were analysed by TLC. Detail conditions are shown in Table 2.2.

Table 2.2 Screening of Zemplen reaction of compound 16.

Reaction number	[Ac-LNB] (mM)	[NaOMe] (mM)	NaOMe/Ac-LNB molar ratio (mM)	Reaction time (h)	Temperature
1	50.4	24	0.5	20	R.T.
2	26.2	10	0.4	20	R.T
3	8.2	0.4	0.05	4	0 °C
4	9.2	1	0.1	20	0 °C
5	4.5	29.1	6.5	0.5	R.T
6	4.5	29.1	6.5	1	R.T
7	4.4	38.5	8.7	1.5	R.T
8	4.4	47.6	10.8	2	R.T
9	4.4	47.6	10.8	2.5	R.T
10	4.5	19.6	4.3	0.5	R.T
11	4.5	19.6	4.3	1	R.T
12	4.5	29.1	6.5	1.5	R.T
13	4.5	29.1	6.5	2	R.T
14	81.7	23.1	0.3	1	R.T
15	81.7	23.1	0.3	1.5	0°C
16	39.7	23.1	0.6	6	0°C

When performing a similar NaOMe/Ac-LNB molar ratio conditions (< 1) and at 0°C (entries 3 and 4), reactions did not evolved. In contrast, at high NaOMe/Ac-LNB molar ratio (entries 4-11) and shorter time of reaction (<2.5 h) (entries 5-12), partial deacetylation of the disaccharide was observed in all cases, and also another compound, most probably *O*-methyl disaccharide

derivative. Best conditions for deacetylation of the Ac-LNB were found at low NaOMe/Ac-LNB molar ratio (< 1) and shorter time of reaction (Table 2.4, entries 14-16). Best conditions at 0°C (entry 15) rendered total deacetylation of the disaccharide with a minor amount of *O*-methyl disaccharide derivative.

Given the difficulties due to the NaOMe basic conditions, a milder strategy using methanolic ammonia (NH_3/MeOH) was used (Neilson and Werstiuk, 1971). Reactions of peracetylated LNB and GlcNAc (as control) were performed and followed by TLC.

According to literature, after 3 h of reaction, deacetylation should be complete. However, in our hands, peracetylated products were present at 3 h of reaction, and after 50 h, deacetylation was complete, but a minor amount of sugar with the OMe group at the C-1 position were also found.

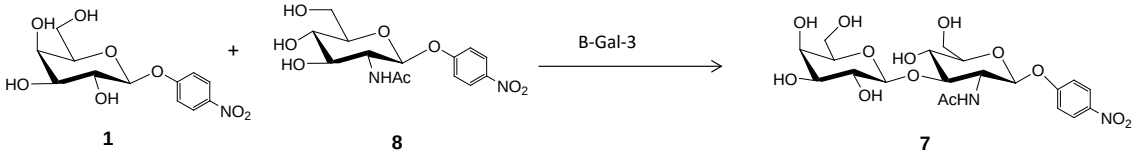
Considering these results, purification of the enzymatic crude with unprotected sugars and using the charcoal-celite column was chosen.

2.3.2 Synthesis of *p*-nitrophenyl β-lacto-*N*-bioside (*p*-NP-LNB)

Since the transglycosylation strategy for the synthesis of LNT involves *p*-NP-LNB (**7**) as substrate, different syntheses of this compound were evaluated. Total synthesis 2 scheme proposes the enzymatic synthesis of *p*-NP-LNB from monosaccharides, and in total synthesis 1 scheme it is developed the chemical synthesis from LNB.

2.3.2.1 Enzymatic synthesis of *p*-nitrophenyl β-lacto-*N*-bioside *via* β-Gal-3 transglycosylation

β-Gal-3 enzyme was evaluated as biocatalyst of the reaction between *p*-NP-Gal (as donor) and *p*-NP-GlcNAc (as acceptor) to yield *p*-NP-LNB (*p*-NP-Galβ-1,3GlcNAc) (**7**).



1 **8** **7**

Figure 2.5 Enzymatic synthesis of compound 7 using β-Gal-3 transglycosylation.

The transglycosylation activity was assayed using 2 nM of protein, 85 mM of *p*-NP-Gal and of *p*-NP-GlcNAc at 1:5 or 1:10 donor:acceptor molar ratio in a final volume of 100 µL. Enzymatic reactions were performed in 100 mM phosphate buffer pH 6.0 at 37 °C and different amounts of ionic liquids. At different times, 1 and 3 hours, 20 µL of samples were taken and analysed using HPLC-ELSD-UV. Specific reaction details are listed in Table 2.3.

Table 2.3 Screening of different ionic liquids for transglycosylation reactions using β-Gal-3 and *p*-NP-GlcNAc as acceptor.
BMIN [PF6]: 1-Butyl-3-methylimidazolium hexafluorophosphate, TROMA [NTF2]: methyltrioctylammoniumbis (trifluoromethylsulfonyl)-imide, BMIN [NTF2]: 1-Butyl-3-methylimidazolium bis(trifluoromethylsulfony1) imide.

Reaction number	Ionic liquid	Final amount used (v/v)	Molar ratio donor : acceptor
1	BMIN [PF6]	30%	1:5
2	BMIN [PF6]	30%	1:10
3	TROMA [NTF2]	30%	1:5
4	BMIN [NTF2]	30%	1:5
5	GC12	43%	1:5
6	GC9	49%	1:5

In Figure 2.6, chromatograms obtained after 1 hour of reaction are shown. Figure 2.6A shows *p*-NP-LNB as standard. Figures 2.6B-G show peaks before 3 min, which corresponds to ionic liquids and buffer, and peak at 4 min to the substrates *p*-NP-Gal and *p*-NP-GlcNAc. While reactions evolved, new peaks at higher retention times appeared. Peak at 5.9 min corresponds to *p*-NP-LNB and peaks at 7.7 and 8.7 correspond to *p*-NP derivatives since they absorbed at UV (317 nm). It is known that β-Gal-3 can also formed β -1,6 linkages when GlcNAc was used as acceptor (Bayón *et al*., 2013). Here, using *p*-NP-GlcNAc as acceptor might change the regiospecifity of the enzyme, and peaks at 7.7 and 8.7 min could correspond to the β-1,4 and β-1,6 isomers of LNB, but they were not further identified.

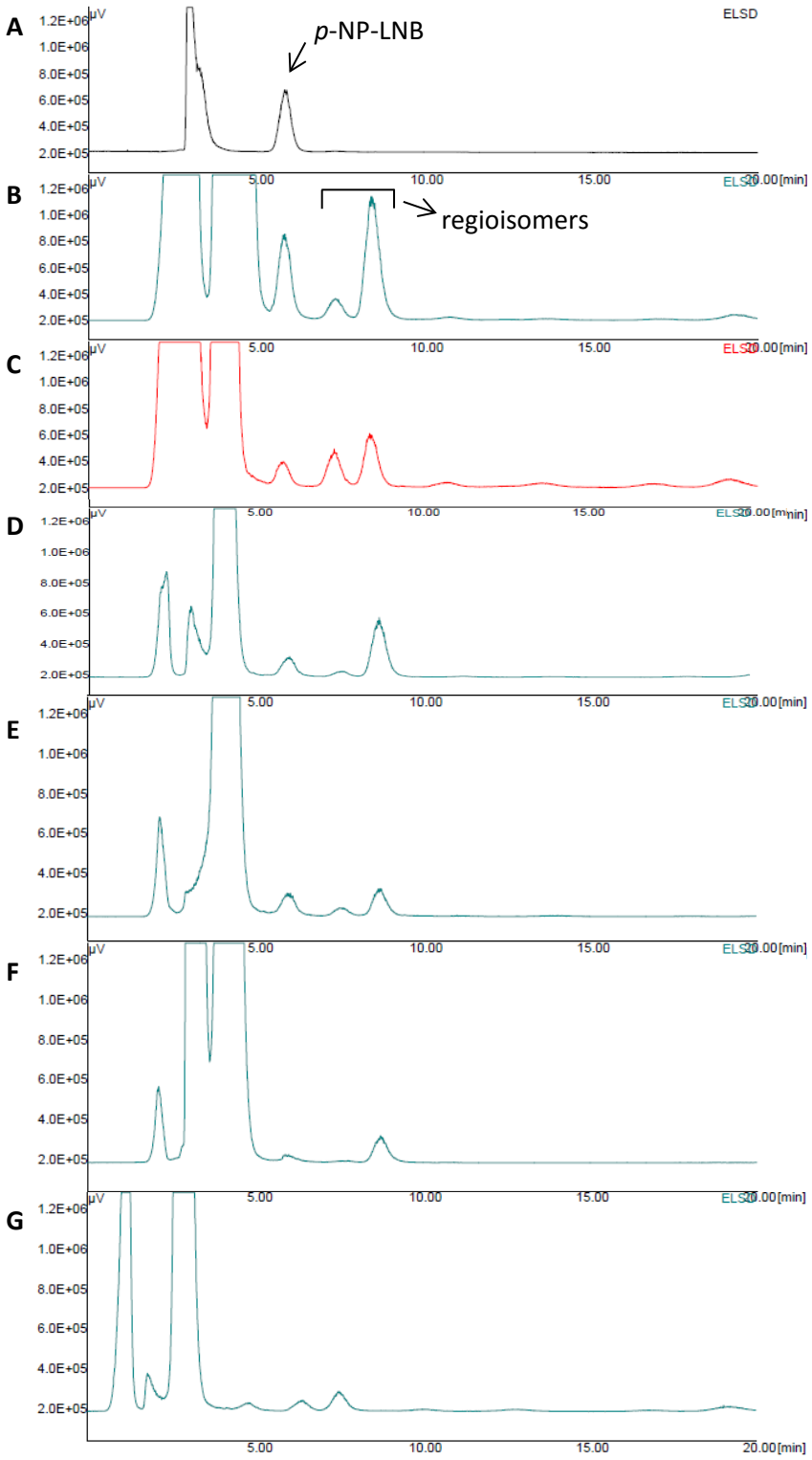


Figure 2.6 HPLC-ELSD chromatograms of the transglycosylation reaction of *p*-NP-LNB at 1 hour.
(A) p-NP-LNB standard in phosphate and reactions in the presence of (B) BMIN [PF6] 1:5 (C) BMIN [PF6] 1:10 (D) BMIN [NTF2] 1:5 (E) GC 9 (F) GC 12 (G) TROMA [NTF2].

The same products were obtained in all six reactions, regardless of the ionic liquid used. However, *p*-NP-LNB was never the major product obtained. Transglycosylation in the presence of BMIN [NTF2], GC12 and GC9 (Figure 2.6E-G) was very low, while the best conditions were obtained in the presence of BMIN [PF6] (Figure 2.6B). In this case, LNB was obtained in 32% of the total products formed. In addition, transglycosylation was not increased when a higher excess of acceptor was used (Figure 2.6C), suggesting that *p*-NP-GlcNAc could act as inhibitor. Moreover, increasing the reaction time to 3 hours, products were hydrolysed. As is shown in Figure 2.7, peaks at 5.9, corresponding to *p*-NP-LNB and peak at 8.7 min decreased from 1 hour of reaction (Figure 2.7B) to 3 hours of reaction (Figure 2.7C).

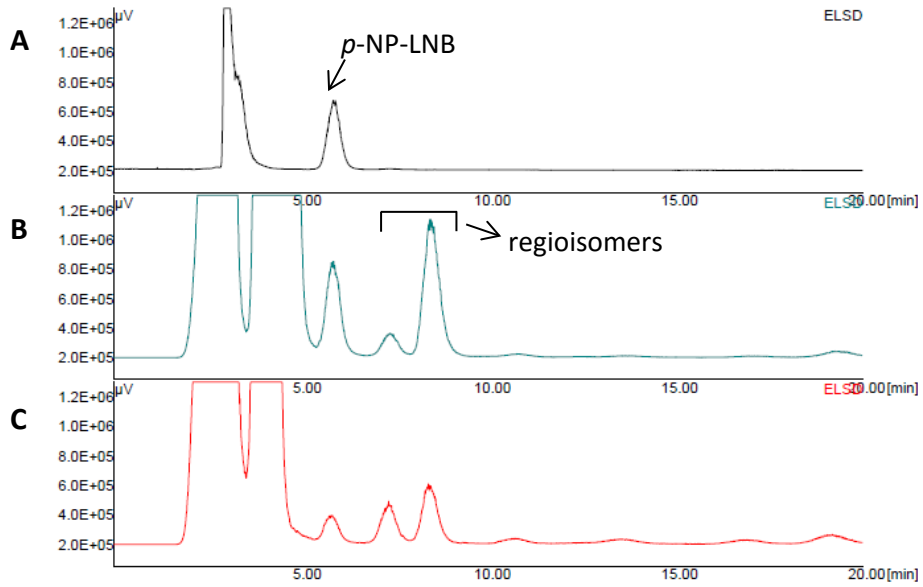


Figure 2.7 HPLC-ELSD chromatograms of the transglycosylation reaction of *p*-NP-LNB using BMIN [PF6].
(A) (A) *p*-NP-LNB standard in phosphate. Reactions LNB using 30% BMIN [PF6] and 1:5 molar ratio (donor:acceptor) at (B)1 hour of reaction (C) 3 hour of reaction.

Moreover, transglycosylation reaction between *p*-NP-Gal and *p*-NP-GlcNAc was also evaluated with immobilised β-Gal-3. Immobilisation of the enzyme can offer several advantages such as repeated recycle of the enzyme, simple separation of reaction products from the biocatalyst and enhance enzyme stability (Hernaiz and Crout, 2000a). In some cases, even changes in regioselectivity have been observed. For example, immobilised β-galactosidase from *B. circulans* (Biolacta N5) showed a specific regioselectivity for β-1,4 *N*-acetyllactosamine (Galβ-1,4GlcNAc), while the free form of the enzyme yielded β-1,4 and β-1,6 isomers (Hernaiz and Crout, 2000b).

In collaboration with Dr. Hernáiz group, β-Gal-3 was immobilised with glyoxyl-agarose in order to assess whether the regioselectivity of the enzyme could be modified. Glyoxyl agarose is constituted by thick agarose fibres containing a large number of very stable aldehyde groups attached to the support by short spacer arms. Under alkaline conditions, the protein is immobilised. After borohydride reduction the enzyme remains attached to the support by means of very stable secondary amino groups (Mateo *et al*., 2006).

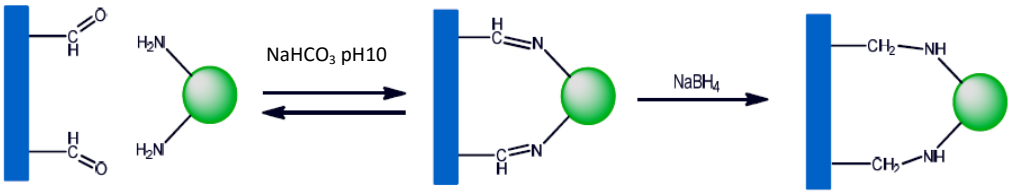


Figure 2.8 Immobilization strategy with glyoxyl-agarose.

The transglycosylation activity was assayed using 30 mg of immobilised enzyme (0.03 mg of protein), 85 mM of *p*-NP-Gal and 425 mM of *p*-NP GlcNAc (1:5 donor:acceptor molar ratio) in a final volume of 100 µL. Enzymatic reactions were performed at 37 °C and 100 mM phosphate buffer pH 6.0 in the absence or the presence of 30% of BMIN [PF6].

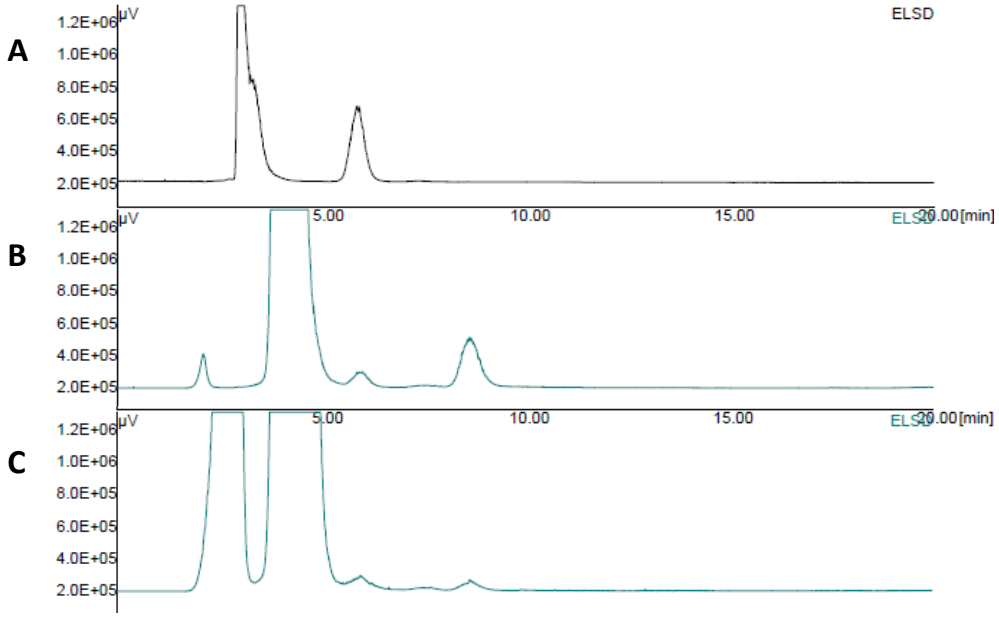


Figure 2.9 HPLC-ELSD Chromatograms of the transglycosylation reaction of *p*-NP-LNB using immobilised β-Gal-3 at 1 hour of reaction.
(A) *p*-NP-LNB standard, (B) 100 mM phosphate buffer, (C) phosphate buffer and BMIN [PF6].

In Figure 2.9, chromatograms after 1 hour of reaction are shown. In addition to the peaks at 3 and 4 min corresponding to BMIN [PF6], salts and substrates (*p*-NP-Gal and *p*-NP-GlcNAc), two new peaks, at 5.9 and 8.7 appeared. None of the reaction conditions assayed were satisfying. In the absence of BMIN [PF6], other regioisomer than β-1,3 was mainly formed, and in the presence of BMIN [PF6] transglycosylation yield was very low. In conclusion, the catalytic efficiency of immobilised β-Gal-3 when using *p*-NP-GlcNAc as acceptor was decreased compared to the free enzyme. These results are in agreement with the literature, since immobilised β-Gal-3 also showed a reduction in catalytic efficiency when GlcNAc was used as acceptor to synthesise LNB. In the presence of phosphate buffer without ionic liquid, transglycosylation yield using immobilised β-Gal-3 was reduced from 51% to 31%, and in the presence of BMIN [PF6], the yield was decreased from 99% to 55% (Carlos Bayón Sánchez, 2013).

2.3.3.2 Chemical synthesis of *p*-nitrophenyl β-lacto-*N*-bioside

Since the chemical synthesis of *p*-NP-LNB was not reported by the time of this study, the approach was done based on the synthesis of *p*-NP-GlcNAc from GlcNAc (Chen and Withers, 2007; Hattie *et al*., 2012; Sauerzapfe *et al*., 2008). As total synthesis 1 scheme shows, this involves the activation of the anomeric carbon *via* chlorination, followed by *p*-NP glycosylation, and further deacetylation. More details are presented in Figure 2.10.

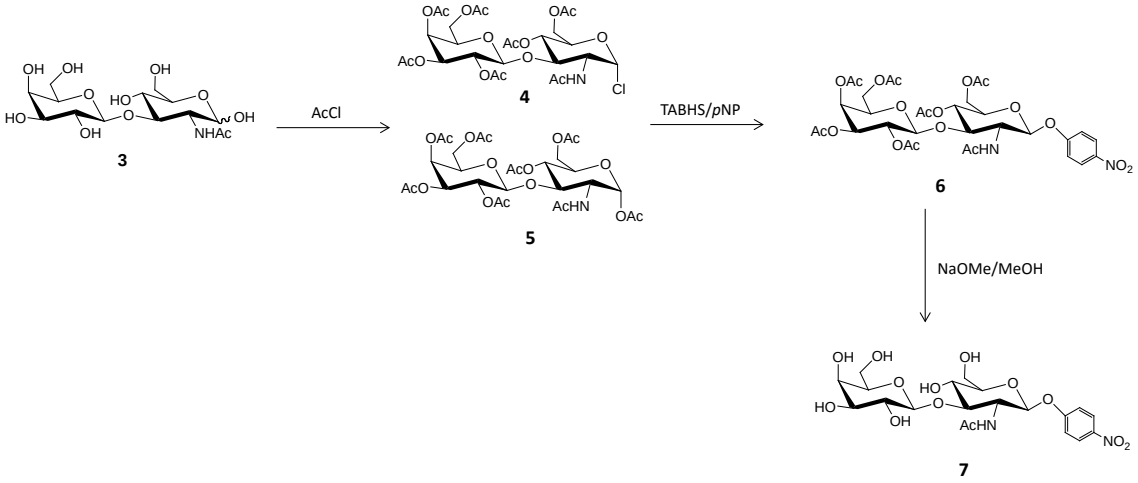


Figure 2.10 Synthesis of the substrate 7.
Anomeric carbon activation by chlorination, further 4-nitrophenol glycosylation and finally sugar deprotection.

LNB (**3**) was reacted with acetyl chloride at room temperature to give a mixture of both sugars, the fully acetylated α-chloride (**4,** 50%) and the α-per-*O*-acetate (**5,** 50%). MS (m/z 676 $[M+Na]^+$ of Cl-LNB; 700 $[M+Na]^+$ of Ac-LNB) and NMR analysis confirmed the presence of both products (δ 6.04 ppm (d, *J*=3.7 Hz, 1H, αH1^{I} Cl-LNB)) and 6.00 ppm (d, *J*= 3.7 Hz, 1H, αH1^{I} Ac-LNB)). This mixture, without further purification, was used for the phase transfer glycosylation reaction with 4-nitrophenol (*p*-NP). After several aqueous ammonium carbonate washing steps, the crude was treated under Zemplen conditions to remove the acetyl protecting groups. The sugar crystallised during deacetylation and was washed after filtration with MeOH to give pure *p*-NP-LNB (**7**, 11% yield). MS (m/z 527.2 $[M+Na]^+$) and NMR analysis confirmed the structure of compound **7** (8.20 ppm (d, *J* = 9.2 Hz, 2H, Ar-H), 7.18 ppm (d, *J* = 9.2 Hz, 2H, Ar-H), 5.24 ppm (d, *J* = 8.5 Hz, 1H^{I}).

The same reaction was performed with GlcNAc (**2**) to obtain *p*-NP-GlcNAc (**8**). Contrary to compound **7,** compound **8** was obtained in 65% of yield. The low yield obtained in *p*-NP-LNB synthesis was mainly due to the workup after glycosylation. Later, Thiem and coworkers reported the synthesis and substituted the workup step for a chromatography step increasing the yield up to 33% (Böttcher and Thiem, 2014).

2.3.3 Synthesis of 1,2-oxazoline derivatives

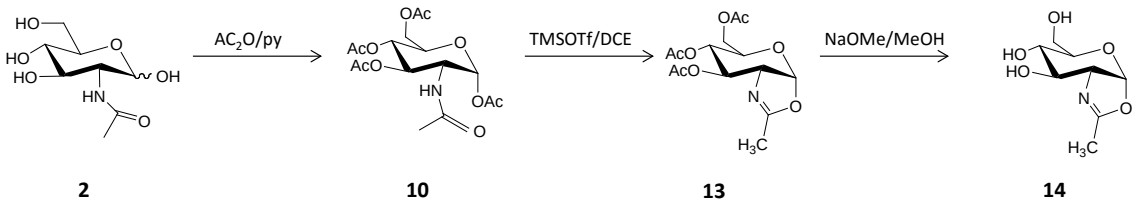
Since the glycosynthase strategy for the synthesis of LNT involves oxa-LNB (15) as substrate, different syntheses of this compound were evaluated. Total synthesis 2 scheme proposes the enzymatic synthesis of oxa-LNB from monosaccharides and total synthesis 1 scheme, the chemical synthesis from LNB.

2.3.3.1 Chemical synthesis of 1,2-oxazoline derivative of GlcNAc

β-Gal-3 enzyme was evaluated as biocatalyst for the synthesis of the oxazoline derivative of Lacto-*N*-biose (oxa-LNB). The reaction involved *p*-NP-Gal (as donor) and the oxazoline derivative of GlcNAc (oxa-GlcNAc, as acceptor) to yield *p*-NP-LNB (*p*-NP-Galβ-1,3GlcNAc).

Oxazoline ring formation of *N*-acetylglucosamine (**14**) was evaluated using two different methodologies. It was first attempted by acetylation followed by treatment of (**10)** with TMSBr and boron trifluoride in the presence of collidine as reported by Huang *et al*., 2009. Other side reaction products were seen and should be separated by a chromatographic step, therefore, and other methodology was applied.

Then, the synthesis of **14** was prepared essentially as originally reported in Shoda *et al*., 2002. Compound **2** was acetylated, followed by oxazoline ring formation with trimethylsilyl trifluoromethanesulfonate (TMSOTf). Finally, the acetyl groups were removed by NaOMe in MeOH. Typical neutralization with Amberlite® IR-120 (H+) resin was not applied, since this acidic condition opens the oxazoline ring. The characteristic NMR signal of the proton of the anomeric carbon (5.95 ppm, *J* = 7.3 Hz) confirmed the structure of **14**. Pure compound **14** was obtained with an overall yield of 60%.



2 **10** **13** **14**

Figure 2.11 Synthesis of compound 14.
Acetylation, followed by oxazoline ring formation with trimethylsilyl trifluoromethanesulfonate (TMSOTf) and further deacetylation.

pH-dependence study of **14**

Stability of the oxazoline ring at different pHs was evaluated in order to perform further transglycosylation reactions. Hydrolysis rates of **14** were studied by NMR using Water Suppression Enhanced analysis (WET_1D) (Figure 2.12). The WET method uses selective pulses, applied on the solvents resonance, followed by pulsed field gradients to dephase residual solvent magnetization. Oxazoline ring is more stable if the amide group is deprotonated. When it is protonated, the oxazoline ring is opened *via* hydrolysis, and GlcNAc (**2**) is obtained. At WET_1D NMR conditions, proton of the anomeric carbon of the oxazoline derivative can be observed at 5.95 ppm (*J* = 7.4 Hz) and the corresponding to α-GlcNAc and β-GlcNAc appeared at 5.05 ppm (*J* = 3.7 Hz) and 4.50ppm (*J* = 8.9 Hz), respectively.

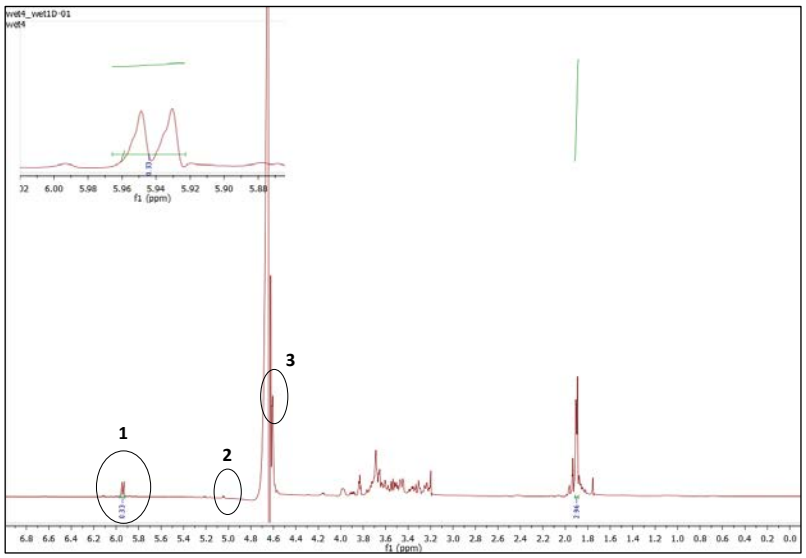


Figure 2.12 NMR analysis of compound 14 after 16 h in 50 mM citrate-phosphate buffer at pH 7.
Characteristic signals are highlighted. 1: proton of the anomeric carbon when the oxazoline ring is formed (d, J = 7.4 Hz), 2: proton of the anomeric carbon when the oxazoline ring is opened in an alpha conformation (d, J = 3.7 Hz) and 3: proton of the anomeric carbon when the oxazoline ring is opened in an beta conformation (d, J = 8.9 Hz).

Below pH 4.5 the oxazoline ring was immediately hydrolysed. In Figure 2.13A the evolution of the stability of the oxazoline ring at different pHs is shown. Oxazoline ring suffered a fast hydrolysis rate at pH 4.5, 5.0 and 6.0. In contrast, at pH 6.5 and 7.0 the hydrolytic rate was slower, and the complete hydrolysis occurred after 24 h. Above pH 7.5 the oxazoline ring remained stable for at least 48 h.

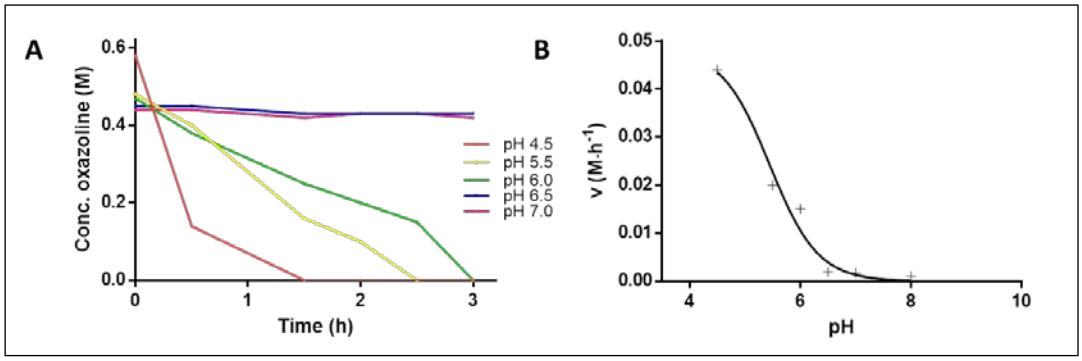
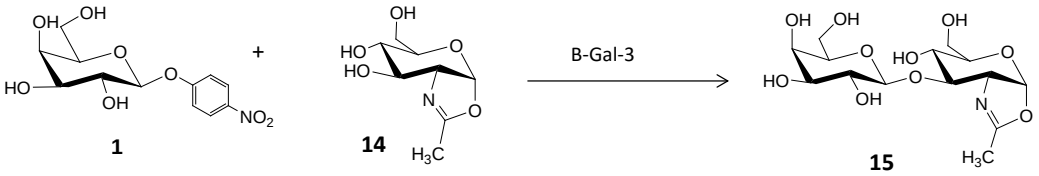


Figure 2.13 pH-dependence study of compound 14.
(A) Hydrolysis of the oxazoline ring of 14 at different pHs. (B) Kinetic profile of the hydrolysis of the oxazoline ring of 14.

As shown in Figure 2.13B, when hydrolytic rates at different pHs are plotted, an estimated p*Ka* value of 5.4 was obtained. When the amide group of the oxazoline is protonated, the nitrogen has a positive charge, favouring oxazoline ring opening. In conclusion, enzymatic reactions with oxazoline derivative substrates are recommended to be performed at pH above 6.5.

2.3.3.2 Enzymatic synthesis of 1,2-oxazoline derivative of Lacto-*N*-biose by β-Gal3

β-Gal-3 was also evaluated as biocatalyst for the synthesis of the oxazoline derivative of Lacto-*N*-biose (oxa-LNB). As total synthesis 2 scheme shows, the reaction using *p*-NP-Gal as donor, and oxa-GlcNAc as acceptor to yield oxa-LNB (**15**) was explored. Oxa-GlcNAc was synthesised following the procedure cited in section 2.3.3.1.



1 **14** **15**

Figure 2.14 Enzymatic synthesis of compound 15 by β-Gal-3 transglycosylation.

β-Gal-3 remains active (>80%) in a pH range of 5-8. At higher pH values, addition of 50% PEG-600 to the reaction media stabilised the protein, and it remained still catalytically active (>60%) at pH 10 for 120 min (Carlos Bayón Sánchez, 2013).

Considering pH dependence stability of the oxazoline ring and the catalytic activity of β-Gal-3 at basic conditions, transglycosylation reactions for oxa-LNT synthesis were set at pH 7.5 and 10.

Final reaction conditions involved 2 nM of protein, 85 mM of *p*-NP-Gal and 425 mM oxa-GlcNAc (donor:acceptor molar ratio, 1:5) in a final volume of 200 µL and at 37 °C. Enzymatic reactions were performed in 100 mM phosphate buffer and in the presence or absence of PEG and BMIN [PF6].

Table 2.4 Different screening conditions for transglycosylation reactions using β-Gal-3 and oxa-GlcNAc as the acceptor.

Reaction number	pH	PEG (v/v)	BMIN [PF6] (v/v)
1	7.5	-	-
2	7.5	50%	-
3	7.5	-	30%
4	10	50%	-

Transglycosylation was not observed in any of the reaction conditions checked. Product formation was not observed by HPLC-ELDS. Since the oxazoline ring was stable at these conditions, it is concluded that β-Gal-3 did not recognize oxa-GlcNAc as acceptor.

2.3.3.3 Chemical synthesis of 1,2-oxazoline derivative of Lacto-*N*-biose

When the same methodologies described for the monosaccharide **12** using TMSOTf or TMSBr and boron trifluoride (reported in section 2.3.3.1) was applied to the peracetylated disaccharide LNB (**16**), reactions rendered the hydrolysis of the disaccharide. Although these reactions were successfully reported in the literature for β-1,4 linked oligosaccharides, here we can confirm that are not applicable to β-1,3 bound disaccharides.

Alternatively, the synthesis of oxa-LNB (**15**) was assayed directly with LNB (**3**) and 2-chloro-1,3-dimethylimidazolinium chloride (DMC) in aqueous media as reported by Noguchi *et al*., 2009 (Figure 2.15).

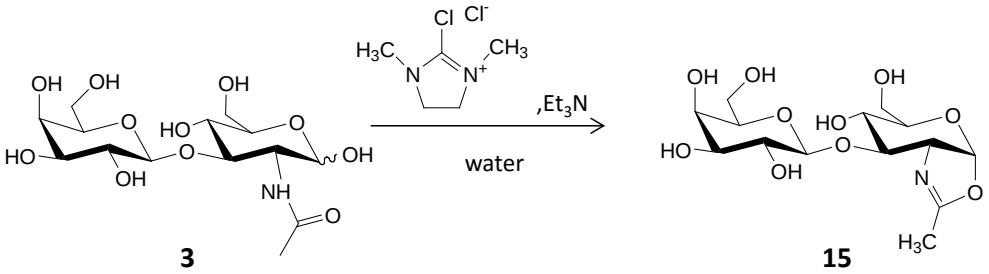


Figure 2.15 Synthesis of compound 15 by using DMC.

The mechanism of oxazoline ring formation in water involves the preferential attack of the hydroxyl group of the beta anomer to the 2-position of DMC to give a reactive intermediate with β-configuration (Figure 2.16). Then, an intra-molecular attack of the carbonyl oxygen of the 2-acetamido group, followed by the abstraction of the amide proton by triethylamine, render a sugar oxazoline and 1,3-dimethyl-2-imidazolidinone. The hydroxyl group of the α-anomer, which is in equilibrium with the β one, may also react with DMC to form an R-adduct. However, the resulting intermediate is immediately hydrolysed by the attack of water, regenerating the starting free sugar, because intra-molecular ring formation from the R-adduct is sterically prohibited. In addition, a considerable amount of DMC was decomposed by the attack of H_2O before being converted to the intermediate (glycosyl dimethylimidazolinium salt) giving rise to the hydrolysed reagent, 1,3-dimethyl-2-imidazolidin-2-one (DMI).

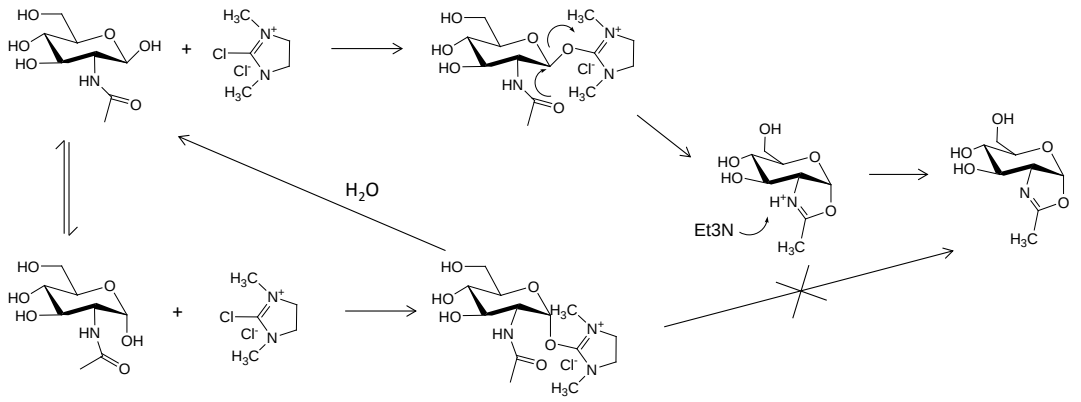


Figure 2.16 Plausible Mechanism for sugar oxazoline formation by DMC.

Synthesis of the disaccharide **15** and monosaccharide **14** (used as control) were successfully achieved in high yields (90%). However, products were not pure since DMI was not easy to remove and appeared in the final product. Other disadvantages concerning the use of DMC involved that it was extremely hygroscopic. Moreover, DMI can strongly inhibit enzymatic reactions according to literature (Noguchi *et al*., 2009).

Thus, an alternative synthesis of the sugar oxazoline derivatives was performed by using a newly designed dehydrative condensing agent, 2-chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride (CDMBI) (Figure 2.17). In this case, the existence of the benzene ring decrease the solubility of the 1,3-dimethyl-1H-benzimidazol-2-one produced (DMBI) as a result of oxazoline formation and can be removed by filtration (Noguchi *et al*., 2012).

The synthesis of compound **15** and **14** (used as control) were prepared directly from the corresponding 2-acetamido-2-deoxy sugars in aqueous media by using CDMBI and Na_3PO_4 to yield 80% in both cases. During the reaction DMBI precipitated. NMR analysis confirmed the structure of oxa-LNB (6.13 ppm (d, *J* = 7.4 Hz, 0.8H, $H1^{I}$-oxa)) with a minor presence of LNB (5.28 ppm (d, *J* = 3.6 Hz, 0.2H, $1H^{I}$-α)). In this case, there was no side product (DMBI) in the final crude, but an excess of salt remained in the final product. Before its use in transglycosylation reactions, the solution of oxa-LNB was neutralised with HCl to decrease the pH around 7.

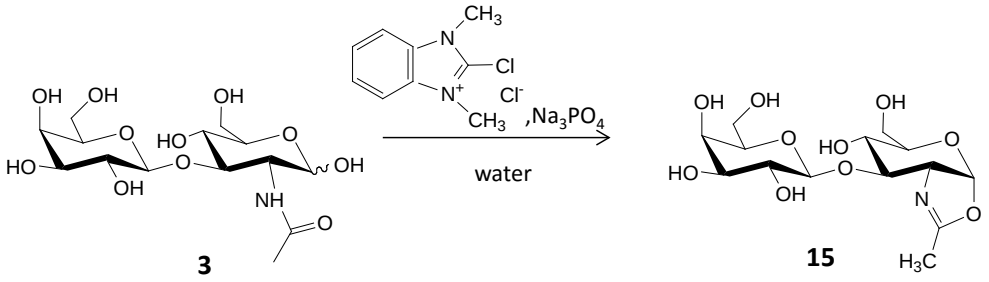


Figure 2.17 Synthesis of compound 15 by using CDMBI.

2.4 CONCLUSIONS

The synthesis of LNB using β-Gal-3 transglycosylation with *p*-NP-Gal and GlcNAc and further purification with charcoal/celite chromatography successfully rendered LNB. By contrast, LNB could not be obatined when silica-gel purification was performed after acetylation and followed by Zemplen reaction. Deacetylation on peracetylated LNB under standard conditions yielded hydrolysis meaning high lability of the β-1,3 linkage when an acetamide group in position 2 is present. Best deacetylation reaction conditions were low sodium metoxide/peracetylated LNB ratio, low temperature and short time reaction.

The synthesis of oxa-LNB, substrate for the glycosynthase approach, was achieved from LNB *via* chemical oxazoline ring formation using a newly designed dehydrative condensing agent, CDMBI, in water solution. By contrast, transglycosylation reaction between *p*-NP-Gal and oxa-GlcNAc using β-Gal-3 as biocatalyst did not occur.

The synthesis of *p*-NP-LNB, substrate for the transglycosylation approach, was afforded *via* chemical synthesis with chlorination step, phase transfer glycosylation and deacetylation with a yield of 11%. Using β-Gal-3 transglycosylation with *p*-NP-Gal and *p*-NP-GlcNAc, the reaction was not regioselective and β-1,3, and others isomers were formed.

p-NP-LNB and oxa-LNB are ready to be used for the kinetic study of the enzyme lacto-*N*-biosidase, and for the subsequent enzyme-catalysed transglycosylation using lacto-*N*-biosidase mutants to achieve the target lacto-*N*-tetraose.

2.5 REFERENCES

Bayón, C., Cortés, Á., Berenguer, J., and Hernáiz, M.J. (2013). Highly efficient enzymatic synthesis of Galβ-(1→3)-GalNAc and Galβ-(1→3)-GlcNAc in ionic liquids. Tetrahedron *69*, 4973–4978.

Böttcher, S., and Thiem, J. (2014). Facile preparation of indoxyl- and nitrophenyl glycosides of lactosamine and isolactosamine. RSC Adv. *4*, 10856.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Carlos Bayón Sánchez (2013). Obtención de biocatalizadores para la síntesis enzimática de glicoestructuras de interés biológico en condiciones sostenibles. Universidad Complutense de Madrid. Tesis Doctoral.

Chen, H.M., and Withers, S.G. (2007). Syntheses of the 3- and 4-thio analogues of 4-nitrophenyl 2-acetamido-2-deoxy-B-D-gluco- and galactopyranoside. Carbohydr. Res. *342*, 2212–2222.

Fujimoto, H., Miyasato, M., Ito, Y., Sasaki, T., and Ajisaka, K. (1998). Purification and properties of recombinant beta-galactosidase from Bacillus circulans. Glycoconj. J. *15*, 155–160.

Hattie, M., Debowski, A.W., and Stubbs, K.A. (2012). Development of tools to study Lacto-N-Biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. ChemBioChem *13*, 1128–1131.

Hernaiz, M.J., and Crout, D.H.G. (2000a). A highly selective synthesis of N-acetyllactosamine catalyzed by immobilised b-galactosidase from Bacillus circulans. J. Mol. Catal. 403–408.

Hernaiz, M.J., and Crout, D.H.G. (2000b). Immobilization/stabilization on Eupergit C of the B-galactosidase from B. circulans and an a-galactosidase from Aspergillus oryzae. Enzyme Microb. Technol. *27*, 26–32.

Huang, W., Lia, C., Lia, B., Umekawab, M., Yamamotob, K., Zhanga, X., and Wang, L.-X. (2009). Glycosynthases Enable a Highly Efficient Chemoenzymatic Synthesis of N-Glycoproteins Carrying Intact Natural N-Glycans. J Am Chem Soc. *131*, 2214–2223.

Mateo, C., Palomo, J.M., Fuentes, M., Betancor, L., Grazu, V., Lopez-Gallego, F., Pessela, B.C.C., Hidalgo, A., Fernandez-Lorente, G., Fernandez-Lafuente, R., et al. (2006). Glyoxyl agarose: A fully inert and hydrophilic support for immobilization and high stabilization of proteins. Enzyme Microb. Technol. *39*, 274–280.

Neilson, T., and Werstiuk, E. (1971). Preparation of 2 ’-0-tetrahydropyranyl Derivatives of Adenosine and Cytidine Necessary for Insertion in Stepwise Synthesis. Can. J. Chem. *49*.

Noguchi, M., Tanaka, T., Gyakushi, H., Kobayashi, A., and Shoda, S. (2009). Efficient Synthesis of Sugar Oxazolines from Unprotected N -Acetyl-2-amino Sugars by Using Chloroformamidinium Reagent in Water Sugar oxazoline derivatives were directly synthesized from the corresponding N -acetyl-2-amino sugars in aqueous media by using. J. Org. Chem. *74*, 2210–2212.

Noguchi, M., Fujieda, T., Huang, W.C., Ishihara, M., Kobayashi, A., and Shoda, S.I. (2012). A practical one-step synthesis of 1,2-oxazoline derivatives from unprotected sugars and its application to chemoenzymatic β-N-acetylglucosaminidation of disialo-oligosaccharide. Helv. Chim. Acta *95*, 1928–1936.

Sauerzapfe, B., Namdjou, D.-J., Schumacher, T., Linden, N., Křenek, K., Křen, V., and Elling, L. (2008). Characterization of recombinant fusion constructs of human β1,4-galactosyltransferase 1 and the lipase pre-propeptide from Staphylococcus hyicus. J. Mol. Catal. B Enzym. *50*, 128–140.

Shoda, S. ichiro, Fujita, M., Lohavisavapanichi, C., Misawa, Y., Ushizaki, K., Tawata, Y., Kuriyama, M., Kohri, M., Kuwata, H., and Watanabe, T. (2002). Efficient method for the elongation of the N-acetylglucosamine unit by combined use of chitinase and ??-galactosidase. Helv. Chim. Acta *85*, 3919–3936.

CHAPTER 3.
LACTO-*N*-BIOSIDASE CHARACTERIZATION

3.1 INTRODUCTION

Bifidibacterium bifidum Lacto-*N*-biosidase (LnbB) is a critical enzyme in the selective promotion of the growth of bifidobacteria in the infant gut ecosystem. It has recently been found that bifidobacteria assimilate HMO as their sole carbon source, since they possess a unique metabolic pathway specific for Lacto-*N*-biose (LNB) called the GNB/LNB pathway. In particular, LnbB is the responsible of the hydrolysis of the lacto-*N*-tetraose (LNT, **19**) into lacto-*N*-biose (LNB, **3**) and lactose (**18**) disaccharides (Figure 3.1) (Wada *et al*., 2008).

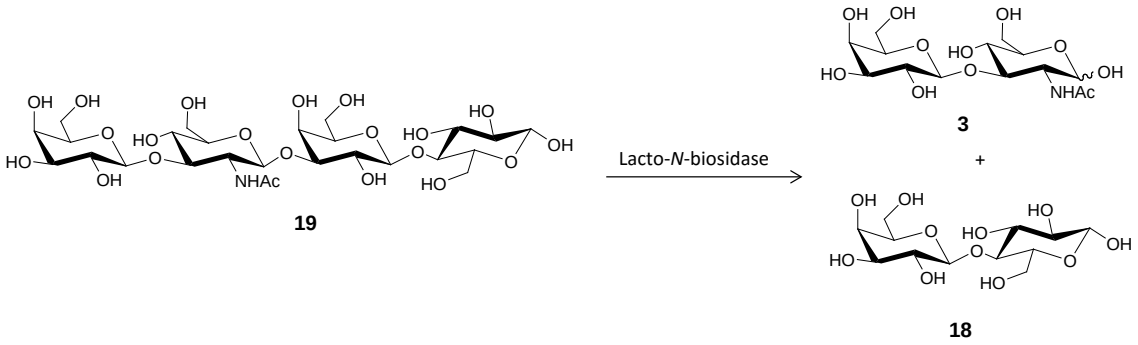


Figure 3.1 Hydrolysis of Lacto-*N*-tetraose catalysed by Lacto-*N*-biosidase to yield lacto-*N*-biose and lactose.

LnbB consists of 1112 amino acids and contains a signal peptide, a catalytic GH20 domain, a carbohydrate-binding module (CBM) from family 32, a bacterial Ig-like domain, and a membrane anchor according to Wada and coworkers (Wada et al., 2008).

LnbB is classified as a member of family 20 glycosyl hydrolases (GH20), and as other members of this family, it has been shown to use a two-step catalytic mechanism involving substrate-assisted catalysis.

In terms of substrate specificity, it is known that LnbB does not hydrolyse the fucosylated forms of lacto-*N*-tetraose (lacto-*N*-fucopentaose I and II) or lacto-*N*-neotetraose (LNnT, type 2 chain), indicating that LnbB specifically releases LNB from the non-reducing end of unmodified type 1 oligosaccharides. In addition, among the *para*-nitrophenyl (*p*-NP) sugars tested, the enzyme is more active toward *p*-NP-LNB and less active toward *p*-NP-GNB (30% compared to *p*-NP-LNB), and did not hydrolyse α-linked disaccharides. The LnbB protein does not act on β-linked *p*-NP monosaccharides, including *p*-NP-GlcNAc and *p*-NP-GalNAc. With respect to pyridyl amino (PA) oligosaccharides, the enzyme acts on lacto-*N*-tetraose, but not on the ganglioside structure with a β-linked GNB (Wada *et al*., 2008).

Due to the relevant characteristics of LnbB, we envisage LnbB as a promising biocatalyst for HMO synthesis. In this chapter, we report the kinetic characterization of the recombinant LnbB and the identification of its catalytic residues.

3.2 MATERIALS AND METHODS

3.2.1 Bioinformatics

Protein sequences were retrieved from UniProt (http://www.uniprot.org). Assignment of protein domains were obtained from different databases (Pfam, Superfamily, PSI-BLAST and CAZY) and also using predicting online tools; Jpred3 (Cole *et al*., 2008) and HHpred (Söding *et al*., 2005)Sequence alignment was performed with PROMALS (Pei and Grishin, 2007) and ClustalW with default options.

3.2.2 General materials and experimental procedures

3.2.2.1 Materials

All reagents are from Sigma-Aldrich unless otherwise designated.

Bacterial strains
Escherichia coli DH5α (Invitrogen) was used as host strains for regular cloning experiments.
Escherichia coli BL21 (DE3) (Invitrogen) was used for protein expression.

Plasmids
pET 24b and pET 28a+ (Novagen) were used for cloning and protein expression.

Oligonucleotides
The oligonucleotides used in this work were purchased from Stab Vida Laboratories.

LB broth (Luria-Betrani)
10 g/L bacto-tryptone, 5 g/L bacto-yeast extract and 10 g/L NaCl, pH 7.2
To prepare solid medium, 15 g/L bacto agar was added to LB and poured on Petri dishes.

Kanamycin stock solutions
Kanamycin (Kan) stock solutions (30 mg/mL) were prepared in distilled sterile water and filtered with 0.2 µm syringe filters.
Final concentration at culture media was 30 µg/mL.

Autoinduction media
LB media supplemented with; 25mM Na_2HPO_4, 25mM K_2HPO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 2mM $MgSO_4$, 0.5% (w/v) glycerol, 0.05% (w/v) glucose and 0.2% (w/v) lactose.

Protease inhibitor cocktail (PMSF) solutions
PMSF stock solutions (100mM) were prepared in isopropanol.

DNA loading buffer (6X)
360 mM Tris-base at pH 8.8, 9% (w/v) SDS, 0.9% (w/v) bromophenol blue, 15% (v/v) β-mercaptoethanol and 30% (w/v) glycerol.

TAE buffer (50X)
2M Tris-base, 50mM EDTA and 17.5% (v/v) glacial acetic acid.

Agarose gel for DNA analysis
0 1% (w/v) of agarose in TAE buffer (1X). Heat the solution on hot plate until rolling boil, let cool for 10 minutes, and add ethidium bromide to a final concentration of 0.5 µg/mL before pouring gel in the container.

Running gel solution for SDS-PAGE analysis
125 mM Tris-base at pH 6.8, 0.1% (w/v) SDS, 10-20% (w/v) ammonium persulfate, 0.1% (v/v) TEMED and 8-16% (v/v) acrylamide.

Stacking gel solution for SDS-PAGE analysis
375 mM Tris-base at pH 8.8, 0.1% (w/v) SDS, 10-20% (w/v) ammonium persulfate, 0.1% (v/v) TEMED and 4% (v/v) acrylamide.

Running buffer for SDS-PAGE analysis
25 mM Tris-base, 190 mM glycine and 0.1% (w/v) SDS, at pH 8.3.

Protein loading buffer (5X)
63 mM Tris-base at pH 6.8, 10% (w/v) glycerol, 2% (w/v) SDS, 5% (v/v) ß-mercaptoethanol and 0.0025% (w/v) bromophenol blue.

Buffer A for enzyme purification
20 mM Na_2HPO_4, 150 mM NaCl at pH 7.5.

Buffer B for enzyme purification (Affinity purification)
20 mM Na_2HPO_4, 150 mM NaCl, 500 mM imidazole at pH 7.5.

Buffer C for enzyme kinetics
25 mM Na_2HPO_4, 25 mM citric acid at pH 4.5

Glycine buffer for enzyme kinetics
500 mM glycine at pH 11

Buffer D for enzyme purification (HPSEC)
20 mM Na_2HPO_4, 300 mM NaCl at pH 7.5.

3.2.2.2 General methods

Bacterial growth in LB medium
The cells were grown overnight at 37 °C, 250 rpm in LB medium and diluted 200 fold to obtain starting main culture. If proceed, LB medium was complemented with the appropriate antibiotic for cell selection, 30 µg/mL of kanamycin.

Preparation of heat shock competent cells

A single colony of *E. coli* was inoculated into 2 ml LB medium and incubated at 37 °C overnight 250 rpm. The overnight culture was then used to inoculate 500 ml LB. Cells were grown to OD_{600} of around 0.4, placed on ice for 20 min and then harvested by centrifugation at 3000 g for 10 min at 4 °C. Cell pellets were gently resuspended in 30 ml of ice-cold 100mM $CaCl_2$ and incubated on ice for 30 min. Cells were harvested by centrifugation as described above and gently resuspended in 8 ml 100 mM $CaCl_2$,containing 15% glycerol. The aliquots (100 µL) were stored at -80 °C.

Plasmid DNA isolation

Plasmidic DNA was isolated by following the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich).
A single colony of *E. coli* DH5α, previously transformed with the target plasmid, was inoculated into 5 ml LB medium and incubated at 37 °C overnight, 250 rpm. The o/n culture was then used following the manufacturer's instructions.

Oligonucleotides

Primers used for PCR amplification of fragments were designed following standard rules: the length of most primers was ranged from 18-25 nucleotides with a >50% of guanine/cytosine (GC) content greater than 50% in order to result in a melting temperature (Tm) from 50-65 ˚C.

Primers for site-directed mutagenesis were designed following instructions of Liu *et al*., 2008. Each primer pair contains non-overlapping sequences at their 3' end and primer-primer complementary (overlapping) sequences at the 5' end. The mismatching nucleotides for the mutation are located at the centre of the overlapping region. It was considered to introduce as less mismatched nucleotides as possible for a desired mutated amino acid, taking into account the codon usage for E. coli.

Oligonucleotide stock solutions (100 µM) were prepared in nuclease-free sterile water and diluted to working solution (1 µM).
Final concentration at PCR reaction was 0.05-0.02 µM.

Enzymatic digestion

The corresponding DNA (1 µg) was digested with appropriate endonucleases, 10 U/µg of DNA, following the manufacturer's instructions (New England Biolab, NEB). To inhibit restriction enzymes, the digested mixtures were heated at 65 °C for 10 min in the case of heat labile enzymes. For heat resistant enzymes, the DNA fragments resulting from the digestion were purified from an agarose gel after electrophoresis. PCR products were mixed with 6x DNA loading buffer and electrophoresed on a 1% agarose gel.

PCR products purification

The corresponding DNA was purified using a commercially available purification kit: Gen Elute gel extraction kit (Sigma-Aldrich) after electrophoretic separation in 1% agarose gel.

Ligation

Ligation was carried out in a 20 µL final volume reaction mixture containing nuclease-free water, corresponding vector and insert DNA (50 ng vector), 2 µL of 10x ligation buffer (NEB),

and 2.5 U/µL T4 DNA polymerase (NEB). The ligation reaction was incubated o/n at 16 °C or 30 min at room temperature, depending on laboratory timing. A control ligation was set up without insert. DNA ligase inactivation was performed by heating the reaction mixture for 10 min at 65 °C.
The correct vector-insert molar ratio might has to be experimentally determined to adjust the molar ratios from anywhere between 1:1 to 1:5 vector to insert, respectively.

Heat shock transformation
Competent cells (100 µL) were incubated on ice for 30 min with 20 µL of ligation reaction. Then, they were heat shocked at 42 °C for 2 min. Tubes were placed back on ice for a further 5 min. 500 µL of LB media was added and cells were incubated for 1 h at 37 °C, in stirring conditions (250 rpm). Transformation mixtures were spread onto LB agar plates containing 30 µg/mL of Kan for transformant selection. Plates were incubated o/n at 37 °C.

Colony screening by PCR
To screen for recombinant plasmids, colony PCR was performed. Amplification was carried out in a 20 µL final volume reaction mixture containing 0.05-0.02 µM T7 forward and reverse primers, nuclease-free water, REDTaq PCR Reaction Mix (ReadyMix™, Sigma-Aldrich). With a 20 µL pipette, freshly transformed colony was picked and mixed thoroughly in the reaction mixture. PCR conditions were a 3 min initial denaturation at 95 °C followed by 34 cycles: 95 °C 15 s, 50 °C 15 s (corresponding annealing temperature), 72 °C 50 s or 4 min depending on the amplicon, and a 10 min final extension at 72 °C. PCR were carried out using a thermocycler PTC-200 (MJ research).

DNA sequencing
Plasmidic DNA of the new clones was sent to Technical and Scientific Services of University of Barcelona or to StabVida Laboratories to be automatically sequenced by the fluorescent Sanger method. T7 forward and reverse primers were used.

SDS-PAGE analysis
6 mL of running gel solution and 1.5 mL of stacking gel solution were placed in the electrophoresis apparatus according to the manufacturer's instructions (Mini-protean Tetra Vertical Electrophoresis Cell, Bio-Rad). Protein samples were mixed with an appropriate amount of protein loading buffer and heated for 10 min at 95 °C before loading into the gel. The gels were run in running buffer at 100 V for 90 min. After electrophoresis, gels were washed three times with water, stained with the Bio-safe Coomassie from Bio-Rad for 30 min, and finally, rinsed several times with water before visualizing protein bands in them.

Protein concentration quantification Protein concentration was determined by Bradford (Bradford, 1976) at Abs 595 nm using BSA as standard.

3.2.2.3 Design and cloning LnbB

A synthetic gene of Lacto-*N*-biosidase from *Bifidobacterium bifidum* (LnbB), codon-optimized for the expression in *Escherichia coli* were produced by GenScript (GenScript, NJ, USA). The

LnbB-FL gene (37-1064) was cloned into the pET24b plasmid (Novagen, Madison, WI, USA) using the restriction sites NdeI and XhoI to obtain the plasmid pET 24b-LnbB-FL-Cter.

LnbB-FL gene with N-terminal His-tag was also obtained by PCR reaction and cloned into a pET28a+ plasmid (Novagen, Madison, WI, USA) using NdeI and Bpu1102 as flanking regions to obtained the plasmid pET 28a-LnbB-FL-Nter.

Amplification was carried out with iProof High-Fidelity DNA Polymerase from Bio-Rad IProof polymerase (0.04 U/µL, Bio-Rad) in a 20 µl final volume reaction mixture containing plasmidic DNA (50-100 ng), sense primer (0.05 µM), antisense primer (0.05 µM), nuclease-free water and I-proof PCR master mix (10 µL). A 3 min initial denaturation at 95°C was followed by 30 cycles: 95°C 30 s, 30 s of the corresponding annealing temperature for the different primers, 72°C 2.5 min and a 10 min final extension at 72°C. The following forward and reverse primers were used (restriction site are underlined):

Full-length N-his sense	5'-GGAATTCCATATGGCTGATGACTCCGCAGCAGGC-3'
Full-length N-his antisense	5'-TAATAGCTCAGCTTCGGTGCCCGG-3'

PCR were carried out using a thermocycler PTC-200 (MJ research).

PCR products were purified with Gen Elute gel extraction kit (Sigma-Aldrich). Vector pET28a+ (Novagen) and purified PCR products were digested with the specific endonucleases. After ligation, LnbB-FL gene was cloned into the pET28a+ vector with NdeI and Bpu1102 as flanking restrictions sites. Both proteins mentioned above were design to be purified using affinity chromatography (IMAC) with Ni^{2+}-charged columns.

3.2.2.4 Site-directed mutagenesis of active site mutants of LnbB

Mutagenesis of specific catalytic residues of LnbB was carried out using modified QuikChange site-directed mutagenesis protocol (Liu and Naismith, 2008). For each single mutant, a sense/antisense primer pair was designed. Each primer pair contains non-overlapping sequences at their 3' end and primer-primer complementary (overlapping) sequences at the 5' end. The mismatching nucleotides for the mutation are located at the centre of the overlapping region.
It has been tried to introduce as less mismatched nucleotides as possible for a desired mutated amino acid, taking into account the codon usage for *E. coli*.
Primers used are list below (mutated nucleotides are underlined, and codons affected are lowercase letter):

D320A sense	5'-TGGGCGCGgcgGAATACATGATTGGTACC-3'
D320A antisense	5'-CATGTATTCcgcCGCGCCCATGTGCCAATATTTCG-3'
E321A sense	5'-GCGCGGATgcgTACATGATTGGTACCTCG-3'
E321A antisense	5'-AATCATGTAcgcATCCGCGCCCATGTGCCAATATTTCG-3'

Plasmid pET 24b-LnbB-FL-Cter, which encodes for full length C-terminal LnbB was used as template for mutagenesis. PCR were carried out using IProof polymerase (0.04 U/µL, Bio-Rad)

and a thermocycler PTC-200 (MJ research). Amplification was carried out in a 50-20 µl final volume reaction mixture containing plasmidic DNA (50-100 ng), sense primer (0.05-0.02 µM), antisense primer (0.05-0.02 µM), nuclease-free water and I-proof PCR master mix (10-25 µL) (Ref. 1725310, Bio-Rad). A 3 min initial denaturation at 95°C was followed by 30 cycles: 95°C 30 sec, 30 sec of the corresponding annealing temperature for the different primers, 72°C 4 min, and a 10 min final extension at 72°C. After PCR reaction, the samples were digested with DpnI for 1 hour at 37 °C, and then, transformed into *E. coli* DH5α. Positive transformants were confirmed by DNA sequencing and the final constructs were used to transform *E. coli* DH5α and BL21 (DE3) cells.

3.2.2.5 Protein expression trials

The protein expression levels were optimized by adjusting temperature and time of expression, either with IPTG or with autoinduction medium.

The cells were grown overnight at 37° C, 250 rpm in LB medium and diluted 200 fold to obtain starting main culture. In the case of IPTG induction, the starting cultures were induced at OD_{600} 0.5-0.6 and OD_{600} 1 at 25 °C, 30 °C and 37 °C, for 3, 6 and 14 hours of induction. In the case of autoinduction media, cells were allowed to grow for 12, 24, 48 and 60 hours at 22 °C, 30 °C and 37 °C. 1 mL of culture was taken from each experiment. Cells were harvested by centrifugation (4000g, 15 min, 4°C), resuspended in buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) and disrupted by sonication (2 µm probe, 50% amplitude, 15 sec on, 45 sec off for 3 cycles) (Brandson). After centrifugation of the lysate (16000rpm, 1 h at 4 °C), the supernatant and the pellet were analysed by SDS-PAGE.

3.2.2.6 Expression and purification of recombinant Lacto-*N*-biosidase enzymes

Expression of both LnbB was performed either with IPTG or with autoinduction medium.

IPTG induction: recombinant cultures of *E.coli* BL21 were grown in LB medium with kanamycin at final concentration of 30µg/ml for 16 hours, 250 rpm at 37 °C. Protein expression was induced adding IPTG at final concentration of 1mM and after 3hours of incubation at 250rpm and 37 °C, cells were harvested by centrifugation (4000xg, 15 minutes).

Autoinduction media: recombinant cultures of *E. coli* BL21 (DE3) star cells were grown following an autoinduction protocol (Studier, 2005) in LB medium with 30 µg/mL kanamycin, 25mM Na_2HPO_4, 25mM K_2HPO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 2mM $MgSO_4$, 0.5% glycerol, 0.05% glucose and 0.2% lactose at 250 rpm and 30 °C. After 24 h cells were harvested by centrifugation (4000xg, 15 minutes).

Purification was carried out as follows: harvested cells were resuspended in buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) and disrupted by French press (1200 psi). Cell free extracts were centrifuged (20000xg, 4 °C, 1 hour) and the supernatant was applied into a 1mL Ni^{2+}-charged HiTrap chelating column chromatography. Proteins were purified to homogeneity by a linear gradient of Buffer B (Na_2HPO_4 20mM, NaCl 150mM, imidazole 500mM, pH 7.5) at flow rate of

1 mL/min, and appropriate fractions were collected. The protein concentration was determined by Bradford measuring Abs 595 nm using BSA as standard

3.2.2.7 Purification of Lacto-*N*-biosidase enzymes using FPLC

Both proteins were purified further using an XK16/60 Superdex 200 column (GE Healthcare), preequilibrated with buffer A. The calibration curve was obtained with 1-2 mg/mL of standards (Dextran Blue (>2000kDa), Alcohol dehydrogenase (150kDa), BSA monomer (66 kDa), Carbonic anhydrase (29kDa), Cytochrome C (12.4 kDa), Aprotinin (6.5 kDa) Sigma-Aldrich). All purifications were done in an ÄKTA FPLC (GE Healthcare).

1-2 mL of concentrated proteins (standards or samples) from affinity purification was loaded into the Superdex 200 column by loop injection. Elution was done with buffer A at a flow rate of 1 mL/min. SDS-PAGE analysis was performed for all fractions collected, and fractions containing the protein of interest were pooled, concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore) and stored at 4 °C. In general, the elution yielded a single peak of the target protein with >95% of purity.

3.2.2.8 Protein Identification

Since LnbB has a higher molecular mass which is out of the range of detection of the Bruker Daltonics spectrometer, it was analysed by peptide mass fingerprint. Pure proteins obtained from acrylamide gel bands were digested with trypsin (Trypsin Gold Mass Spectrometry Grade, Promega) and peptides were identified using Mascot (Matrix Science, London UK) by search on NCBInr database.

Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK) with a laser light wavelength of 632.8 nm and a scattering angle of 173 degrees. Temperature was set at 25°C. Protein solutions in buffer A were measured without previous dilution.

3.2.2.9 Kinetics of the hydrolase activity

Lacto-*N*-biosidase activity assays with *p*-nitrophenyl β-lacto-*N*-bioside (*p*-NP-LNB, synthesised in this thesis, Chapter 2) were performed at 30 °C in a volume of 0.1 mL using 96-wells microtiter plates using a Bravo liquid handling Robot (Agilent). Enzymatic reactions were performed in a reaction mixture containing the substrate and Buffer C (25mM of citrate-25mM phosphate buffer, pH 4.5). After pre-incubation for 5 min at 30 °C, reactions were initiated by addition of the enzyme (20 nM), specific concentration is determined in each experiment). At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed.

Substrate concentration to determine Michaelis-Menten kinetic parameters was between 0-1 mM. Substrate concentration for specific activity was at 0.25 mM (according to Km and kcat values of 68 µM and 89 s^{-1} of Wada *et al*., 2008, it was performed at saturated conditions.

Initial rates (v_{ow}) were obtained from the linear progress curve of release of *p*-NP, and they were expressed as v_{ow}/[E] in s^{-1}. Kinetic parameters *kcat*, K_M, and $kcat/K_M$ were calculated by nonlinear regression to the Michaelis–Menten equation.

3.2.2.10 pH-activity profile dependence with *p*-NP-LNB

The pH dependence of enzyme activity was studied using *p*-NP-LNB (synthesised in this thesis, Chapter 2). The assays were performed at standard conditions: at 30 °C, pH values ranging from 3 to 8.5 and final volume of 0.1 mL. After pre-incubation of 20nM of the full length protein for 15 min in the appropriate buffer, addition of 0.25 mM substrate initiated the reaction. At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed. The molar extinction coefficients of *p*-nitrophenol at 405 nm were calculated at each pH and substrate stability was also measured at each pH. **I**nitial rates values were plotted against pH according to Equation 3.1.

$$\mathrm{Vo} = \frac{Kcat^{lim}}{1 + (10^{pKa1-\mathrm{pH}}) + (10^{\mathrm{pH}-pKa2})}$$

Equation 3.1

3.2.2.11 Temperature-activity dependence with *p*-NP-LNB

The temperature-activity dependence of enzyme activity was studied using *p*-NP-LNB (synthesised in this thesis, Chapter 2). The assays were performed at standard conditions in a volume of 0.1 mL at 50mM of citrate-phosphate buffer pH 4.5 (Buffer C). The temperature range assayed was from 25-50 °C. After pre-incubation of 20nM of the full length protein for 15 min at the appropriate temperature, addition of 0.25 mM substrate initiated the reaction. At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed. The molar extinction coefficients of *p*-nitrophenol at 400 nm were calculated at each temperature and substrate stability was also measured at each temperature.

3.3 RESULTS AND DISCUSSION

3.3.1 Analysis of domain structure and cloning of Lacto-*N*-biosidase enzyme

Lacto-*N*-biosidase form *B. bifidum* consists of 1112 amino acids and, according to Wada and coworkers, contains a signal peptide, a catalytic GH20 domain, a carbohydrate-binding module (CBM) from family 32, a bacterial Ig-like domain, and a membrane anchor (Wada *et al*., 2008). However, this domain assignment only covers around 65% of the total amino acids contained in LnbB. Therefore, it suggested that some domains are missing in this description.

In order to predict secondary structure of LnbB and further assignment of the domains, prediction and sequence alignment online tools were used, since the crystal structure of the protein was not solved by the time of this study.

Jpred3 (Cole *et al*., 2008) was used to predict LnbB secondary structure. Sequence alignment of all known structure GH20 proteins was performed with PROMALS with default options (Pei and Grishin, 2007). The amino acidic sequence of LnbB showed less than 25% of similarity with the rest of the GH20 proteins. However, in terms of secondary structure, region 1-528 residues of LnbB presented a very similar structure to the others GH20 proteins.

Pfam and Supfam databases assigned region 784-932 and 962-1064 residues of LnbB as CBM32 and Iglike domains, respectively. Nevertheless, region 528-776 residues of LnbB were not assigned and HHpred was used, in this work, to predict the structure. The software is based on identifying homologous sequences with known structure to be used as a template (Söding *et al*., 2005). Results showed that this region was structurally very similar to a lectin domain (ricin-like).

Using data integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY), and the results from the cited online tools (Jpred3, Promals and HHpred), a model of the domain architecture of LnbB was built (Figure 3.2).

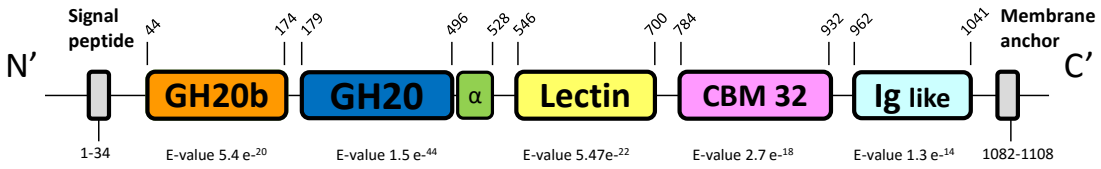


Figure 3.2 Schematic representation of the modular structure of LnbB.
Each box represents the different domains of LnbB. Numbers indicate the amino acids involved in the domain and E-value determines the significance of the assignment.

LnbB is formed by five domains (Figure 3.2). Domains I and II are the GH20b and GH20 ones. Domain III (residues 546-700) is similar to a lectin domain according to an HHpred search. Domain IV (784-932) has sequence homology to CBM32 (Supfam), and domain V (962-1041) is described as a Immunoglobulin-like domain in PFam. In addition, LnbB presents an N-terminal secretion signal peptide and a C-terminal cell wall anchoring sorting signal motif. This indicates that the protein is secreted to the outside of the cell.

In particular, Domain I and II are characteristic domains of glycoside hydrolase family 20. GH20b domain is a non-catalytic domain, which is conserved in most GH20 enzymes although with unknown function. GH20 domain is actually the catalytic domain composed of a $(\beta/\alpha)_8$-barrel (TIM-barrel) architecture, with the active site located at the C-terminal face of the β-barrel (Henn-Sax *et al*., 2001). Lectin and CBM32 are two different carbohydrate binding domains. And the Ig-like domain, which is frequently found in the surface-located enzymes of bacteria, has no assigned function in LnbB.

The design of the gene encoding for LnbB protein was thought to avoid trasnmembrane regions. Accordingly, recombinant LnbB (34-1064 amino acids) lacks the signal peptide and the membrane anchor compared to the *B. bifidum* protein.

A synthetic gene encoding the LnbB from *B. bifidum*, codon-optimized for the expression in *E. coli*, was produced. It was supplied in a pUC57 and in pET24b vector between NdeI and XhoI restriction sites to generate a C-terminal His-tagged protein. By PCR reaction, the gene of the N-terminal His-tagged LnbB protein was also obtained to evaluate the effect of the terminal tag. *E.coli* DH5α and BL21 (DE3) were transformed with the different plasmids.

3.3.2 Expression and purification of Lacto-*N*-biosidase enzymes

Two different protein induction methods were performed, IPTG induction and autoinduction in a supplemented LB media. Expression trials were performed at different temperatures and different times of induction. The trials carried out with autoinduction system showed higher levels of protein expression in the soluble fraction (23 mg/L), compared to IPTG induction (7 mg/L).

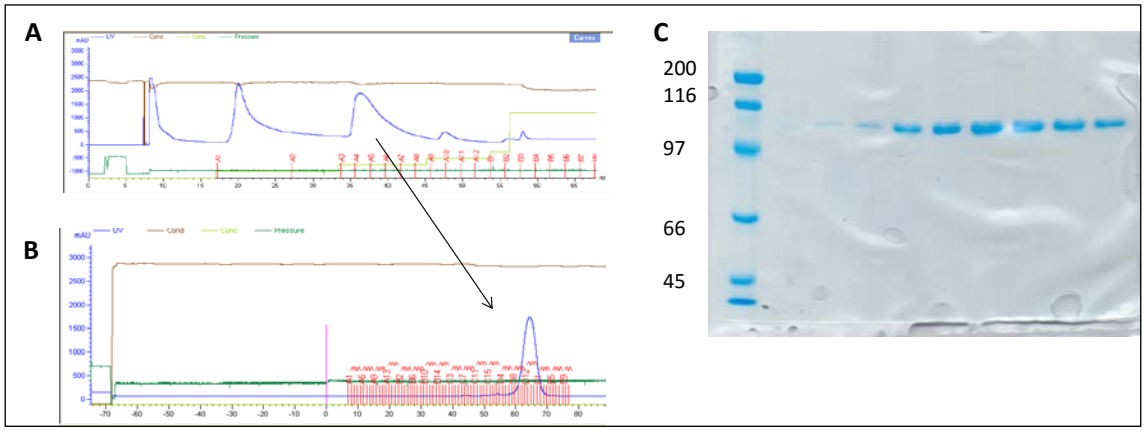


Figure 3.3 C-terminal His-tagged LnbB protein purification.
(A) Chromatogram of affinity chromatography (IMAC). Single peak of full-length was eluted with 50 mM Imidazole in buffer A. (B) Chromatogram of size exclusion chromatography on a Superdex 200 column. Single peak of full-length eluted with buffer A. (C) SDS-PAGE of fractions from size exclusion chromatography. Pure bands of full-length (112kDa) are observed.

Final conditions for the expression of the LnbB proteins (either N- or C-terminal His-tagged) involved growing recombinant cells at 30 °C for 24 h with autoinduction medium. After cell lysis, the soluble fraction obtained was purified by affinity chromatography (Figure 3.3A). SDS-PAGE analysis revealed that full-length migrated with an approximate molecular mass of 112

kDa, which agrees with its respective theoretical molecular mass (Figure 3.3C). Fractions containing the protein were pooled and concentrated before being purified on a Superdex 200 column (GE Healthcare) (Figure 3.3B). After the gel filtration column, fractions were pooled and concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore). Final protein concentration was determined spectrophotometrically by Bradford method measuring Abs 595 nm using BSA as standard (Smith *et al.*, 1985). Average yields of 23 mg of full-length protein per litre of culture were obtained. Both His-tagged LnbB proteins were expressed and purified by following the same protocol and were obtained as soluble proteins with similar yields.

C-terminal His-tagged LnbB protein solutions were also analysed by mass spectrometry (MALDI-TOFF MS) and analytical size-exclusion chromatography (HPSEC).The identification of peptides after enzymatic digestion of wild type LnbB was done, and the molecular mass estimation of the protein in solution was consistent with a monomeric 112 kDa protein. Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK). The full-length protein showed a single peak for monodisperse particles of 10 nm and PDI 0.30, which agrees with monomeric species (Figure 3.4).

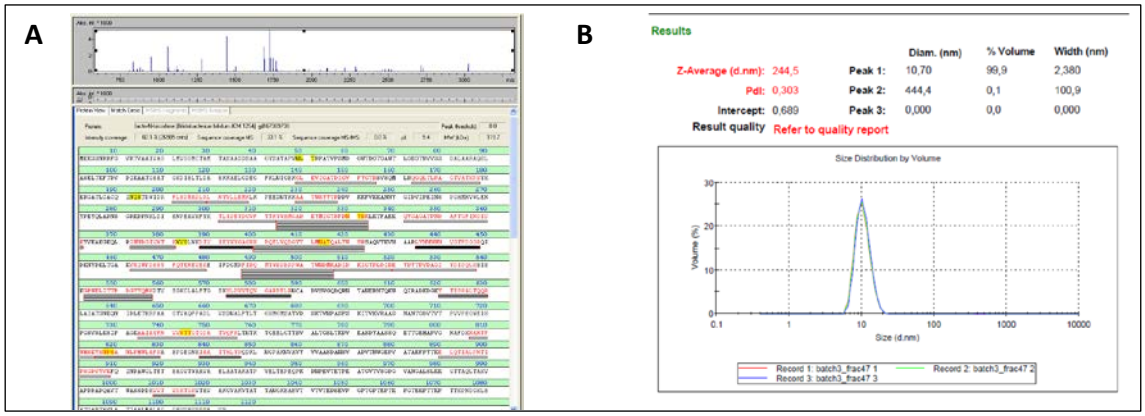


Figure 3.4 Molecular mass and oligomerization state analysis of the C-terminal His-tagged LnbB protein.
(A) MALDI-TOF spectrum of enzymatic digestion of LnbB. (B) Size distribution plot by Dynamic Light Scattering. 99.9 % of the particles of the assayed sample presented a unique size particle of LnbB.

3.3.3 Biochemical characterization of Lacto-*N*-biosidase from *Bifidobacterium bifidum*

3.3.3.1 Kinetic analysis of LnbB using *p*-NP-Galβ1,3GlcNAc as substrate

Lacto-*N*-biosidase hydrolytic activity was characterised using *p*-NP-LNB as substrate (**7**) and the C-terminal His-tagged enzyme.

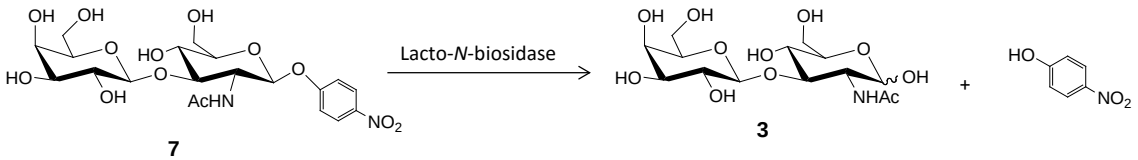


Figure 3.5 Hydrolytic reaction of *p*-NP-LNB mediated by LnbB.

Conditions: 20 mM of LnbB, different *p*-NP-LNB substrate concentrations (0-1 mM), 50 mM citrate-phosphate bufferat pH 4.5 and 30 °C.

Hydrolytic rates at constant LnbB concentration (20 nM) and different substrate concentrations 0-1 mM), showed that the reaction followed saturation kinetics (Figure 3.6) with an apparent k_{cat} value of 25±3 s^{-1} and K_m of 162±34 µM, then k_{cat} / K_m of $154 \cdot 10^3$ $M^{-1} \cdot s^{-1}$.

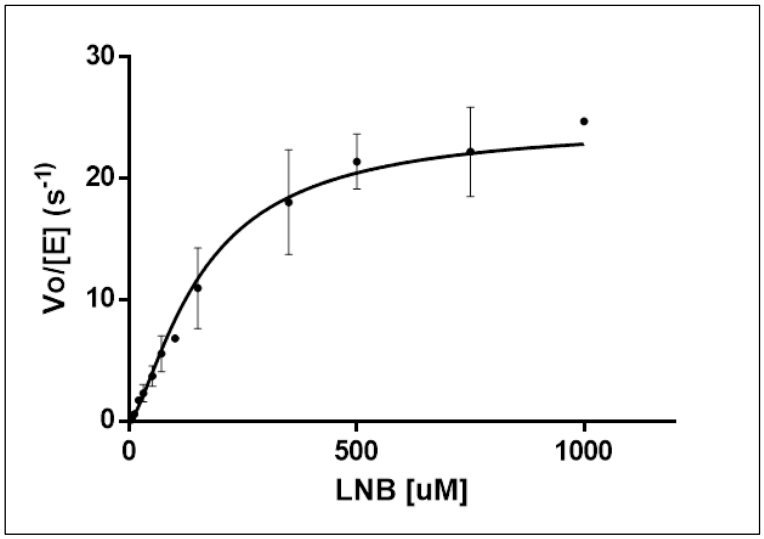


Figure 3.6 Michaelis-Menten kinetics of LnbB.
Reaction conditions involved 20 nM of protein, *p*-NP-LNB substrate (0-1 mM), 50 mM citrate-phosphate buffer at pH 4.5 and 30 °C.

Compared to literature (Table 3.1), these results are in the range of Ito's *et al*. Results from Wada *et al*., and Hattie *et al*., which were determined at pH 4.5 and 37 °C, and showed higher catalytic efficiency.

Table 3.1 Kinetic parameters of the hydrolytic reaction of Lacto-*N*-biosidase from *B. bifidum*.
(Hattie *et al*., 2012; Ito *et al*., 2013; Wada *et al*., 2008)

	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($M^{-1} \cdot s^{-1}$)
Present study	25 ± 3	162 ± 34	$154 \cdot 10^3$
Wada *et al*. 2008	89	68	$1300 \cdot 10^3$
Hattie *et al*. 2012	78± 6	71 ± 11	$1000 \cdot 10^3$
Ito *et al*. 2013	15 ± 1	99 ± 11	$150 \cdot 10^3$

The hydrolytic activities of the C-terminal and the N-terminal LnbB proteins (20 nM) were determined in order to evaluate the effect of the tag. The reaction mixture contained 0.25 mM of *p*-NP-LNB and 50mM of citrate-phosphate buffer at pH 4.5 (Buffer C) and at 30 °C. Both LnbB proteins, N- and C-terminal His-tagged, showed a specific activity of 27.2 ± 3.5 s^{-1}, indicating that the His-tag has no effect on the catalysis rates.

3.3.3.2 pH-activity profile and temperature dependence with *p*-NP-Galβ1,3GlcNAc as substrate

pH dependence of LnbB was analysed using 20 nM of enzyme and 0.25 mM of *p*-NP-LNB in a pH range of 3-8.5 of 50 mM citrate-phosphate buffer and at 30 °C. At different time intervals, the reactions were stopped by addition of 0.5 M of glycine at pH 11. The release of *p*-NP was measured at 405 nm, and the initial rates were determined. These initial rates values were plotted against pH and an approximately bell-shaped profile was obtained (Equation 3.1).

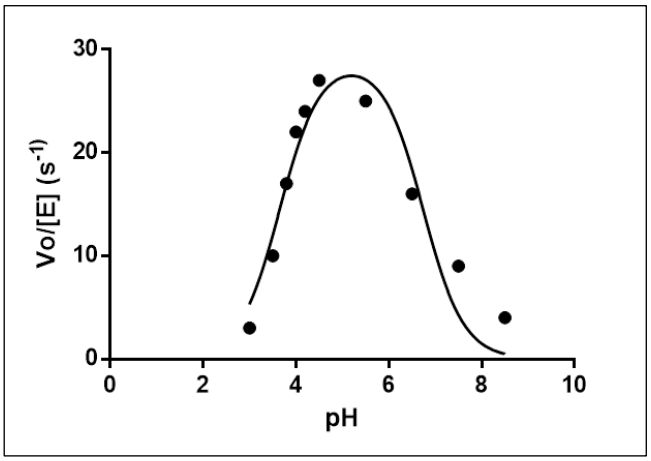


Figure 3.7 pH dependence of LnbB.
Initial hydrolytic rates *vs* pH using *p*-NP-LNB as substrate and 50 mM citrate-phosphate buffer at 30 °C. Profile is fitted to Equation 3.1.

Although the enzyme remained active in a wide pH range, an environment more acidic than pH 3 and more basic than pH 7, results in a loss of activity. Maximum activity was registered at pH 4.5 which is in agreement with other authors (Sano *et al*., 1992; Wada *et al*., 2008).

Temperature dependence of LnbB was analysed using 20 nM of enzyme and 0.25 mM of *p*-NP-LNB, in 50 mM citrate-phosphate buffer at pH 4.5 and in a temperature range of 25-50 °C. The Arrhenius plot is obtained by plotting the logarithm of the rate constant, k, versus the inverse temperature, 1/T (Figure 3.8). As before, reactions were stopped at different times and released *p*-NP was measured.

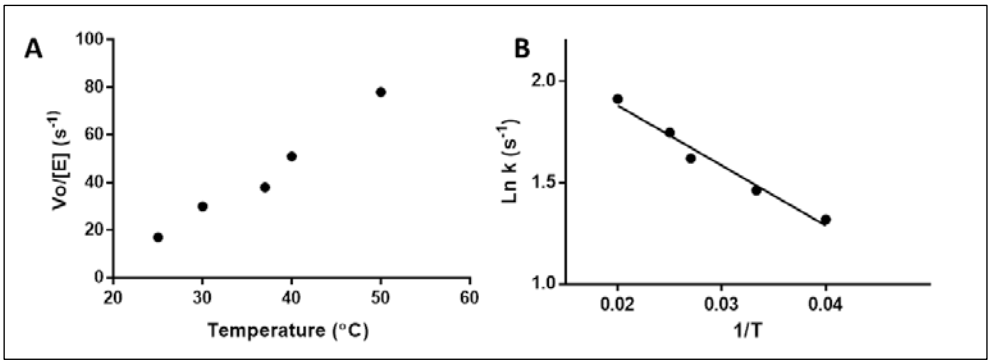


Figure 3.8 Temperature dependence of LnbB.
(A) Temperature dependence of LnbB and (B) Arrhenius plot. Hydrolytic reactions were done at temperature range of 25-50 °C.

Hydrolytic activity of LnbB increased linearly till 50°C. And it showed a linear Arrhenius plot for the reaction with *p*-NP-LNB substrate in the temperature range of 25-50 °C. Hydrolytic reactions were always done at 30 °C, although as shown in Figure 3.8, it is not the optimal temperature for catalysis.

3.3.4 Lacto-*N*-biosidase active site mutants

The catalytic mechanism of GH20 family proteins is a modification of the classical double-displacement process used by retaining glycosidases, and has been described very thoroughly (Mark and James, 2002). GH20 family β-*N*-acetylhexosaminidases operate *via* substrate-assisted catalysis, sometimes also called anchimeric assistance (neighbouring group participation), wherein the 2-acetamido substrate moiety acts as a nucleophile to form an oxazolinium ion intermediate (Figure 3.9).

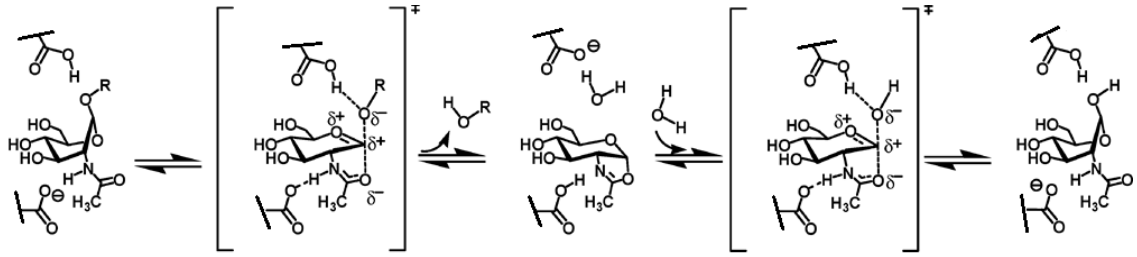


Figure 3.9 Catalytic mechanisms for β-*N*-acetylhexosaminidases.
Double-displacement retaining mechanism of GH20 family proteins: the nucleophile is not provided by the enzyme but by the substrate 2-acetamido group, leading to the formation of an oxazoline intermediate. Image from Slámová *et al*., 2010.

The active site of GH20 enzymes contains a highly conserved pair of catalytic residues Asp-Glu. It has been proposed that the conserved Asp residue plays a role in orientating the 2-acetamido group, and ultimately promoting the nucleophilic attack of this 2-acetamido group at the anomeric position to form the oxazolinium ion intermediate, which is in turn attacked by water to release the hydrolysis product with net retention of the anomeric stereochemistry.

Experiments with mutated hexosaminidases have provided evidence that this aspartate stabilizes the oxazoline transition state and also assists in correctly orienting the 2-acetamido moiety for catalysis. For example, Glu314 and Asp313 are the corresponding essential residues in β-*N*-acetylhexosaminidase from *Streptomyces plicatus* (SpHex). E314Q mutant reduced its V_{max} 296-fold, and its K_M 7-fold, implying that Glu314 is the proton donor in the catalytic reaction of SpHex (Mark *et al*., 1998). The catalytic role of Asp313 was determined by Williams *et al*., obtaining D313A and D313N variants. D313A showed a decrease in the catalytic efficiency of 70-fold, where D313N presented a decrease of $5 \cdot 10^6$-fold. These results were consistent with the important role for Asp313 in carefully positioning the 2-acetamido group in a favourable geometry to act as a nucleophile (Williams *et al*., 2002).

With regard to LnbB protein, Hattie and coworkers confirmed that LnbB uses a mechanism involving substrate-assisted catalysis of the 2-acetamido group of the substrate using 2-acetamido analogues bearing differing degrees of fluorine substitutions (Hattie *et al*., 2012).

The highly electronegative fluorine substituent decreases the basicity of the carbonyl group, and the effect of such substitution is that there is a dramatic decrease in rate for enzymes using anchimeric assistance.

The identification of Asp and Glu catalytic residues of LnbB was based on sequence alignment between LnbB and different members of the GH20 family (Figure 3.10).

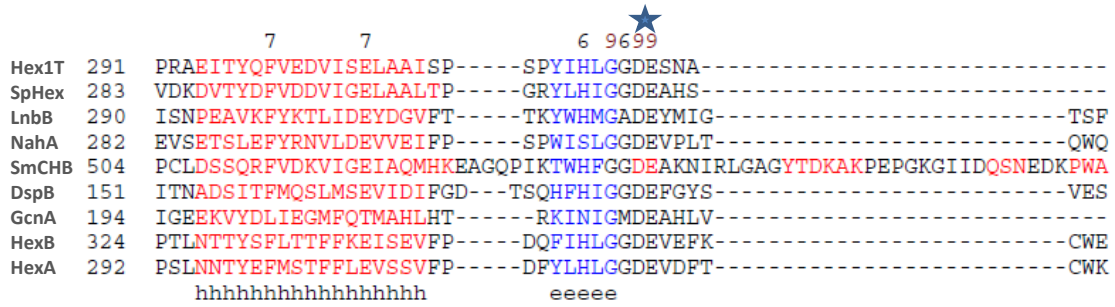


Figure 3.10 Promals multiple alignment of GH20 domain of LnbB and its homologues.
The numbering starts at the initiation of codon of each protein. Conserved catalytic residues are marked by a blue star. The organisms are as follows: Hex1T, *N*-acetylhexosaminidase from Paenibacillus sp; SpHex, *N*-acetylhexosaminidase from *Streptomyces plicatus*; LnbB, Lacto-*N*-biosidase from *B-bifidum*; NahA, *N*-acetylhexosaminidase from *Arthrobacter aurescens*; SmCHB, chitobiose from *Serratia marcescens*; DspB, *N*-acetylhexosaminidase from *Aggregatibacter actinomycetemcomitans*; GcnA, *N*-acetylhexosaminidase from *Streptococcus gordonii*; HexB and HexA, *N*-acetylhexosaminidase from *Homo sapiens*. Full image at Appendix section.

As observed in Figure 3.10 the two catalytic residues (Asp and Glu) are completely conserved among all predicted members of GH20 family. Accordingly, the two corresponding catalytic residues of Lacto-*N*-biosidase are Asp320 and Glu321. Asp320 corresponds to the assisting residue that orientates the 2-acetamido group, and Glu321 acts as the acid/base residue.

Based on this identification, the explanation of the possible catalytic mechanism of LnbB is represented in Figure 3.11. The first step of the reaction (cyclization) proceeds *via* attack of the 2-acetamido carbonyl oxygen on the anomeric carbon to form a covalent bicyclic oxazoline intermediate. This step is facilitated by polarization of the 2-acetamido moiety by Asp320. Departure of the aglycone is facilitated by general acid catalysis provided by Glu321 in the enzyme active site. In the second step (ring opening) the oxazoline intermediate is opened by the general base-catalysed attack of a water molecule on the anomeric carbon. Both steps occur with inversion of stereochemistry at the anomeric centre such that the overall reaction proceeds with net retention of stereochemistry.

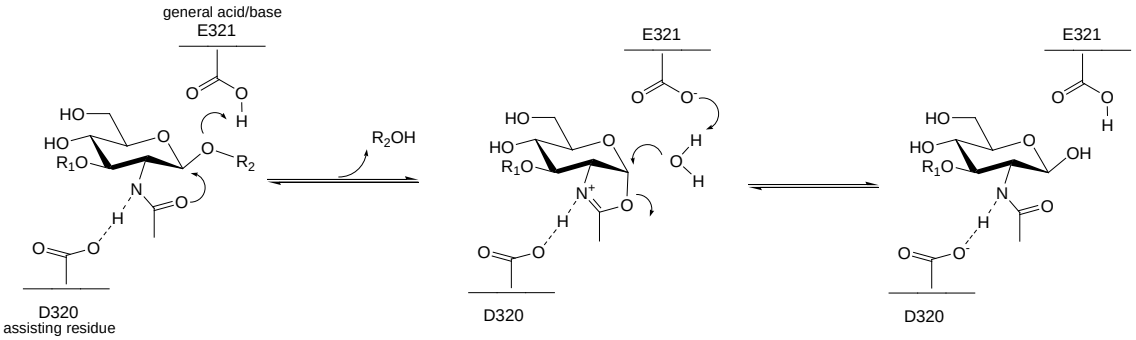


Figure 3.11 Catalytic mechanism of LnbB involving substrate-assisted catalysis.
R_1: Galactose, R_2: Lactose.

Mutants of the catalytic residues of LnbB were designed. Both residues (Asp and Glu) were substituted for alanine (Ala) to generate D320A and E321A mutant proteins. The change to Ala removed the side-chain and the charge from the wild type residue in order to abolish its catalytic capacities.

3.3.4.1 Cloning, expression and purification of Lacto-*N*-biosidase active site mutants

Mutated proteins were successfully obtained by site-directed mutagenesis using wild type Full-length pET24b construct as template (C-terminal His-tagged). Mutations were confirmed by DNA sequencing. Finally, *E. coli* BL21 (DE3) cells were transformed with the final vectors for the expression and purification of mutant recombinant proteins.

Purification of the mutated proteins was performed as described for the wild type LnbB. The protein concentration was determined by Bradford (Bradford, 1976) using BSA as standard. Average yields of 3-7 mg of proteins per litre of culture were obtained. All proteins were considered to be >90% pure by Coomassie staining after SDS-PAGE (Figure 3.12).

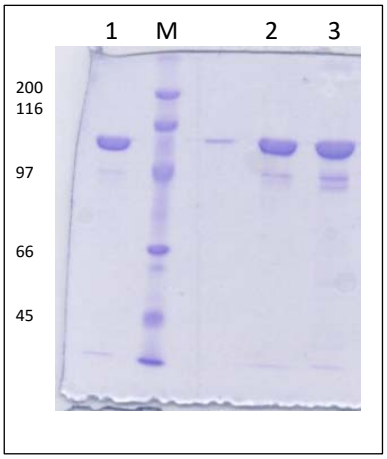


Figure 3.12 SDS-PAGE analysis of the wild type, D320A and E321A mutated proteins.
Lane 1: full-length protein, M: marker, Lane 2: D320A protein, Lane 3: E321A protein. Molecular masses in kDa.

3.3.4.2 Kinetics of Lacto-*N*-biosidase active site mutants

The hydrolytic activities of the mutated proteins and a new batch of wild type LnbB (as control) were also evaluated as described above. Briefly, the hydrolytic activity of the soluble proteins (20 nM) was performed in a reaction mixture containing *p*-NP-LNB (0.25 mM) as substrate in 50 mM citrate-phosphate buffer at pH 4.5 and at 30 °C. The reaction was stopped at different interval times and the release of *p*-NP was measured.

Table 3.2 Kinetic parameters of the hydrolytic reaction of wild type and mutated proteins.
The reactions were performed at 20 nM of proteins, 0.25 mM of substrate, 30 °C in buffer C at pH 4.5.

Protein	Specific activity (s^{-1})
Wild type	31
D320A	0.01
E321A	0.8

Both mutated proteins showed decreased hydrolytic activity compared to the wild type LnbB (Table 3.2). In the case of the D320A mutant, a 3500-fold decrease was observed. This result is consistent with the corresponding D313A mutant of SpHex (Williams *et al*., 2002), indicating that D320 is the candidate to be the assisting residue. Mutation of the proposed acid/base catalytic residue Glu321 to Ala (E321A mutant) yielded a 38-fold less active enzyme. This mutation was expected to abolish the hydrolytic activity of the enzyme but in this case, and in general in GH20 enzymes, the mutation to Ala slightly decreases the activity (E314A of SpHex showed 70-fold decrease). For a retaining β-glycosidase, replacement of the acid/base catalytic residue with a non-ionisable residue should substantially reduce the catalytic efficiency of the mutant enzyme. However, it has a less pronounced effect for substrates bearing good leaving groups that do not benefit significantly from general acid catalysis (Vallmitjana *et al*., 2001). The substitution of Glu321 to Ala321 had little effect and maybe *p*-NP-LNB does not require general acid assistance since *p*-NP is quite good leaving group.

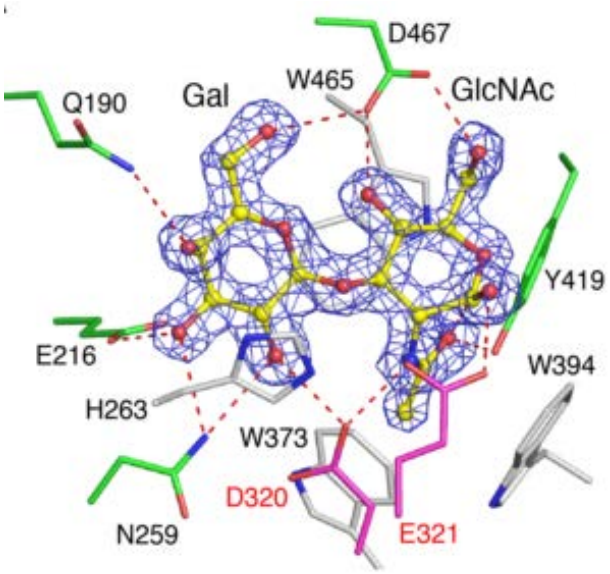


Figure 3.13 Stereoviews of LNB bound to the active site of LnbB.
Catalytic residues, residues forming hydrogen bonds, and residues forming hydrophobic interactions are shown in magenta, green, and white, respectively. Hydrogen bonds are shown as red dashed lines. (PBD entry 4H04). Image from Ito et al., 2013.

These results are consistent with the crystal structure of LnbB in complex with the product LNB (Ito *et al*., 2013), in which, the role of these catalytic positions was confirmed. A clear electron density for ligands at the centre of the barrel domain is observed (Figure 3.13). The two catalytic residues, Asp320 (assisting residue) and Glu321 (acid/base catalytic residue), form hydrogen bonds with the amide nitrogen of the 2-acetamido group and the O1-hydroxyl, respectively (Ito *et al*., 2013).

3.4 CONCLUSIONS

Using data integration from different databases, and the results from structure prediction online tools, a model of the domain architecture of LnbB was built. LnbB is formed by five domains, GH20b and GH20 catalytic domains, a lectin domain, a CBM32 and an Immunoglobulin-like. In addition, LnbB presents an N-terminal secretion signal peptide and a C-terminal cell wall anchoring sorting signal motif.

LnbB has been successfully cloned to obtain N-terminal and C-terminal His-tagged recombinant proteins. The study of the catalytic capacity of both proteins indicated that the terminal tag has no effect on catalysis.

The behaviour of the hydrolytic reaction followed saturation kinetics, with an apparent k_{cat} value of 25±3 s^{-1} and K_m of 162±34 µM, the a catalytic efficiency of k_{cat} / K_m of $154 \cdot 10^3$ $M^{-1} \cdot s^{-1}$.

The pH profile for the hydrolytic activity of the wild type enzyme using *p*-NP-LNB resembled a bell-shaped curve with the optimum catalytic efficiency at a pH of 4.5. In addition, LnbB remained catalytically active in a wide range of temperatures (from 25°C to 50 °C).

The catalytic residues, D320A and E321A, were identified using sequence alignment between LnbB protein and crystallised GH20 proteins. D320A and E321A mutants showed decreased hydrolytic activity compared to the wild type enzyme, 3500-fold and 39-fold decrease, respectively.

3.5 REFERENCES

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Hattie, M., Debowski, A.W., and Stubbs, K.A. (2012). Development of tools to study Lacto-*N*-Biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. ChemBioChem *13*, 1128–1131.

Henn-Sax, M., Höcker, B., Wilmanns, M., and Sterner, R. (2001). Divergent evolution of (betaalpha)8-barrel enzymes. Biol. Chem. *382*, 1315–1320.

Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., Ashida, H., Wakagi, T., Yamamoto, K., Stubbs, K.A., et al. (2013). Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-*N*-biosidase from Bifidobacterium bifidum. J. Biol. Chem. *288*, 11795–11806.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. *8*, 91.

Mark, B.L., and James, M.N.G. (2002). Anchimeric assistance in hexosaminidases. Can. J. Chem. *80*, 1064–1074.

Mark, B.L., Wasney, G. a, Salo, T.J., Khan, a R., Cao, Z., Robbins, P.W., James, M.N., and Triggs-Raine, B.L. (1998). Structural and functional characterization of Streptomyces plicatus beta-*N*-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. J. Biol. Chem. *273*, 19618–19624.

Pei, J., and Grishin, N. V. (2007). PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. Bioinformatics *23*, 802–808.

Sano, M., Hayakawa, K., and Kato, I. (1992). An enzyme releasing lacto-*N*-biose from oligosaccharides. Proc. Natl. Acad. Sci. U. S. A. *89*, 8512–8516.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. Anal. Biochem. *150*, 76–85.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244–W248.

Studier, W. (2005). Protein production by autoinduction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Vallmitjana, M., Ferrer-navarro, M., Planell, R., Abel, M., Ausın, C., Querol, E., Planas, A., and Pérez-Pons, J. (2001). Mechanism of the Family 1 β- glucosidase from *Streptomyces* sp : catalytic residues and kinetic Studies. Biochemistry *40*, 5975–5982.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T., and Yamamoto, K. (2008). Bifidobacterium bifidum Lacto-*N*-Biosidase, a Critical Enzyme for the Degradation of Human Milk Oligosaccharides with a Type 1 Structure. Appl. Environ. Microbiol. *74*, 3996–4004.

Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G., and Withers, S.G. (2002). Aspartate 313 in the Streptomyces plicatusHexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. J. Biol. Chem. *277*, 40055–40065.

CHAPTER 4.
STRUCTURAL-FUNCTIONAL ANALYSIS OF FAMILY GH20 HEXOSAMINIDASES

Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases. Val-Cid, C., Biarnés, X., Faijes, M., and Planas, A. (2015). PLoS One 10, e0128075.

4.1 INTRODUCTION

N-acetyl-β-hexosaminidases, classified in CAZY families GH3, GH20 and GH84 (Carbohydrate Active Enzymes database, www.cazy.org (Cantarel *et al*., 2009)) catalyse the hydrolysis of β-linked GlcNAc or GalNAc from the non-reducing end of glycoproteins, glycosaminoglycans, glycolipids, chitin and other oligosaccharides. They are widely distributed in nature, and they have recently gained a lot of attention, not only for their implication in human physiology and diseases, but also for their potential as biosynthetic tools for carbohydrate and glycomimetic synthesis (Liu *et al*., 2012; Slámová *et al*., 2010). Enzymes from these three families share some basic features in terms of protein structure and catalytic mechanism, but show no sequence homology. GH20 hexosaminidases are the most numerous of them, they are well distributed in different microorganisms, animals, and plants, and they are also involved in many important physiological and pathological processes, such as cell structural integrity, energy storage, pathogen defence, cellular signalling, fertilization, inflammatory events, and lysosomal storage diseases (Intra *et al*., 2008). However, these enzymes show a broad range of different functions depending on the organism but also on the localization of the enzyme within the organism. From the biotechnological point of view, hexosaminidases, together with chitinases, are important enzymes for industrial chitin degradation, the most abundant polysaccharide after cellulose (Slámová *et al*., 2010). In insects, because of the important role of chitin, these enzymes are considered potential targets for insect control, acting as biopesticides (Dowd *et al*., 2007). In humans, the two lysosomal hexosaminidases HexA and HexB are responsible for the hydrolysis of GlcNAc and GalNAc residues of glycosphingolipids such as ganglioside GM2 and their dysfunction is linked to neurodegenerative disorders known as Tay-Sachs and Sandhoff diseases (Gloster and Vocadlo, 2012; Shane Rountree *et al*., 2009). Recent data suggests that lysosomal exoglycosidases, together with many other factors may participate in the progression of development of glial tumours (Wielgat *et al*., 2006). In plants, β-*N*-hexosaminidases have been detected in germinating seeds, suggesting their role in the storage of glycoproteins (Intra *et al*., 2008). In addition, some bacterial hexosaminidases contribute to the virulence of some infections in humans, as in the case of *Actinobacillus actinomycetemcomitans*, an organism that colonizes the human oral cavity and causes periodontitis and *Streptococcus pneumoniae* infections (Slámová *et al*., 2010).

Family GH20 has been mechanistically well characterized; it is known that they catalyse hydrolysis with retention of anomeric configuration with substrate participation mechanism. The 2-acetamido group of the substrate is used for nucleophilic attack at the anomeric carbon to create an oxazoline intermediate (Vocadlo and Davies, 2008).

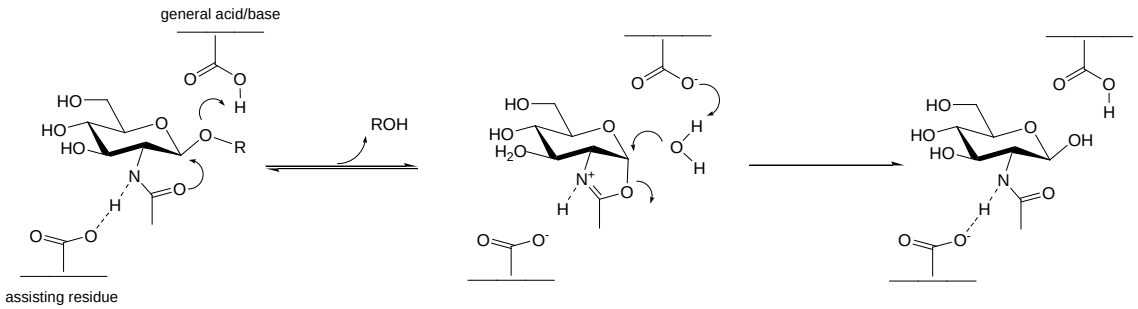


Figure 4.1 Retaining glycosidase by substrate-assisted catalysis, *via* an oxazolonium ion intermediate.

Moreover, protein folding of the catalytic domain is highly conserved among GH20 enzymes; the common feature of all known structures is the $(\beta/\alpha)_8$-barrel (TIM-barrel) architecture, which consists of eight parallel β-sheets forming a cylindrical form, surrounded by the eight α-helices. The active site is located at the C-terminal face of the β-barrel and within the loops that connect the β-strands with the subsequent α-helices (Henn-Sax *et al*., 2001).

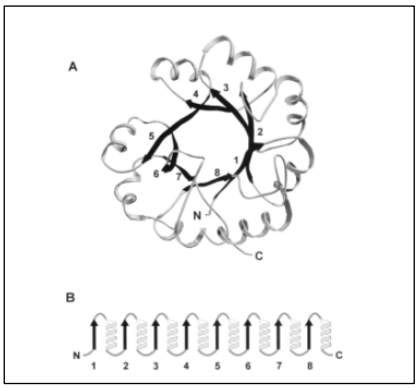


Figure 4.2 The $(\beta\alpha)_8$-Barrel Fold.
(A) View onto the C-terminal ends of the eight β-strands, which form a cylindrical parallel β-sheet. This β-barrel is surrounded by the eight α-helices. β) Topologic diagram of the eight (βα) modules. The active site residues of all known (βα)8-barrel enzymes are located at the C-terminal ends of the β-strands and in the loops that connect the β-strands with the subsequent α-helices. Image from Henn-sax *et al*., 2001.

GH20 enzymes comprise not only β-*N*-acetylhexosaminidases (EC.3.2.1.52), catalysing the removal of terminal non-reducing *N*-acetylhexosamine residues, but also lacto-*N*-biosidases (EC.3.2.1.140), hydrolysing the terminal lacto-*N*-biosyl residues from the non-reducing end of oligosaccharides such as human milk oligosaccharides (HMO).

To date, the structures of thirteen GH20 enzymes are known. Their domain organization is diverse as they present several accompanying domains apart from the catalytic GH20 domain. In addition, some of them exist as dimeric proteins.

Due to their great potential as biocatalysts for biotechnology and biomedical applications, our aim has been to define the minimal functional domain architecture for enzymatic activity in these GH20 family enzymes.

The large *B. bifidum* lacto-*N*-biosidase (LnbB) protein has been chosen due to its biological importance in the metabolism of human milk oligosaccharides and also, because of its complex domain organization. The LnbB gene encodes for a membrane-anchored extracellular enzyme of 1112 amino acid residues. Recently, it was cloned and the recombinant full-length protein was characterized confirming that this enzyme catalyses the hydrolysis of the tetrasaccharide Galβ1,3GlcNAcβ1,3Galβ1,4Glc to lacto-*N*-biose (Galβ1,3GlcNAc) and lactose (Hattie *et al*., 2012; Wada *et al*., 2008).

We here report the structural requirements for functionality of GH20 enzymes to obtain a minimal functional unit that retains the enzymatic activity.

4.2 MATERIALS AND METHODS

4.2.1 Bioinformatics: GH20 structures and sequences retrieval and analysis

The list of currently available 3D structures of GH20 enzymes was retrieved from CAZY database (http://www.cazy.org). Protein structures were downloaded from the Protein Data Bank (http://www.pdb.org). Protein sequences were retrieved from UniProt (http://www.uniprot.org). Assignment of protein domains were obtained from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY) and also using predicting online tools; Jpred3 (Cole *et al*., 2008) and HHpred (Söding *et al*., 2005). Protein structures were analysed and visualized with VMD (Humphrey *et al*., 1996). Structure alignment was performed with MultiSeq (Roberts *et al*., 2006) restricting the alignment region to GH20b and GH20 domains only. Sequence alignment was performed with PROMALS (Pei and Grishin, 2007) and ClustalW with default options.

Table 4.1 PDB entries of the proteins used in the analysis of the GH20 structures.

Enzyme	Organism	PDB
β-*N*-acetylhexosaminidase (SpHex)	*Streptomyces plicatus*	1HP5
β-*N*-acetylhexosaminidase (ScHex)	*Streptomyces coelicolor* A3(2)	4C7G
β-*N*-acetylhexosaminidase (Hex1T)	*Paenibacillus* sp.	3GH5
Chitobiase (SmCHB)	*Serratia marcescens*	1QBB
Lacto-*N*-biosidase (LnbB)	*Bifidobacterium bifidum*	4JAW
β-*N*-acetylhexosaminidase (NahA)	*Arthrobacter aurescens*	3RCN
β-*N*-acetylhexosaminidase βf3009)	*Bacteroides fragilis*	4PYS
β-*N*-acetylhexosaminidase (HexA)	*Homo sapiens*	2GK1
β-*N*-acetylhexosaminidase (HexB)	*Homo sapiens*	1NP0
β-*N*-acetylhexosaminidase (OfHex1)	*Ostrinia furnacalis*	3OZO
β-*N*-acetylhexosaminidase (GcnA)	*Streptococcus gordonii*	2EPN
β-1,6-*N*-acetylglucosaminidase Dispersin B (DspB)	*Aggregatibacter actinomycetemcomitans*	1YHT
β-*N*-acetylhexosaminidase (StrH)	*Streptococcus pneumoniae* R6	3RPM
	Streptococcus pneumoniae TIGR4	2YL8

4.2.2 General materials and experimental procedures

All reagents are from Sigma-Aldrich unless otherwise designated.

General material and methods are detailed in Chapter 3.

Specific substrates for both enzymatic approaches (*p*-NP-LNB and oxa-LNB) were synthesised according to Chapter 2 (see experimental section).

4.2.3 Design and cloning of the different genes of LnbB

The Construct A gene (40-528) of Lacto-*N*-biosidase from *Bifidobacterium bifidum* (LnbB), codon-optimized for the expression in *Escherichia coli* were produced by GenScript (GenScript, NJ, USA). **Construct A gene (40-528)** was cloned into the pET28a+ plasmid (Novagen, Madison, WI, USA) using the restriction sites NdeI and Bpu1102. **Constructs B (176-497) and Bα (176-528) genes** were obtained by PCR reaction using construct A as template, and **constructs C (40-667), D (40-776), E (546-667), and F (546-776) genes** were carried out using the wild type LnbB C-terminal His-tagged gene. The following forward and reverse primers were used (restriction site are underlined):

Construct B and Bα sense	5'-GGAATTCCATATGAAGCCGAAATATAAAGAACGT-3'
Construct B antisense	5'-TAATAGCTCAGCGGCGTCAAC-3'
Construct Bα antisense	5'-TAATAGCTCAGCACGGGAGTCACTCCA-3'
Construct C and D sense	5'-GCGTATCCATGGCTGATGACTCCGCAGCAGGCTA-3'
Construct C and E antisense	5'-GCGTATCTCGAGCAGCGCGTTGTCGC-3'
Construct D and F antisense	5'-GCGTATCTCGAGCAGGCTGCCGGTCAGC-3'
Construct E and F sense	5'-GCGTATCCATGGTGGATGCGGGTATC-3'

Amplification was carried out with iProof High-Fidelity DNA Polymerase (2U/µL) from Bio-Rad in a 50-20 µl final volume reaction mixture containing plasmidic DNA (~150 ng), 0.02- 0.05 µM of sense primer, 0.02-0.05 µM of antisense primer, nuclease-free water and I-proof PCR master mix (Ref. 172-5310, Bio-Rad). A 3 min initial denaturation at 95°C was followed by 30 cycles: 95°C 30 sec, 30 sec of the corresponding annealing temperature for the different primers, 72°C 50 sec or 2.5 min depending on the amplicon, and a 10 min final extension at 72°C. Cloning primers incorporated restriction sites for directed cloning.
PCR were carried out using a thermocycler PTC-200 (MJ research).

PCR products were purified with Gen Elute gel extraction kit (Sigma-Aldrich), and vector pET28a+ (Novagen) and purified PCR products were digested with the specific endonucleases. After ligation, all constructs were cloned into the pET28a+ vector with NdeI and Bpu1102 as flanking restrictions sites, except for construct D and F that were flanked by NcoI and XhoI sites.

All proteins mentioned above were design to be purified using affinity chromatography (IMAC) with Ni^{2+}-charged columns.

Construct F with Strep-tag sequence was obtained by PCR reaction using LnbB gene as template using cloning primers that incorporated streptavidin sequence and restriction sites for directed cloning.

The gen was inserted in pET28a+ among NcoI and Bpu1102 restriction sites. The following primers were used (restriction site are underlined and Strep-tag sequence is bold):

Construct F Strep-tag sense	5'-GCGTATCCATGGTGGATGCGGGTATC-3'
Construct F Strep-tag antisense	5'TAATAGCTCAGCCTA**TTTTTCAAATTGAGGATGTGACCAC GCGCC**CAGGCTGCCGGTCAGCGCAACCG-3'

E. coli DH5α cells were transformed with the different plasmids and positive transformants were verified by DNA sequencing. Afterwards*, E.coli* BL21 (DE3) (Stratagene, La Jolla, CA, USA) were also transformed for protein production. Both strains were cultivated in LB-medium under selective conditions (final concentration of kanamycin 30µg/ml).

4.2.4 Site-directed mutagenesis of loop mutants of LnbB

Mutagenesis of specific residues of loop 573-575 of LnbB was carried out using modified QuikChange site-directed mutagenesis protocol (Liu and Naismith, 2008). For each single mutant, a sense/antisense primer pair was designed. Each primer pair contains non-overlapping sequences at their 3' end and primer-primer complementary (overlapping) sequences at the 5' end. The mismatching nucleotides for the mutation are located at the centre of the overlapping region.
It has been tried to introduce as less mismatched nucleotides as possible for a desired mutated amino acid, taking into account the codon usage for *E. coli*.
Primers used are list below (mutated nucleotides are underlined, and codons affected are lowercase letter):

H573A sense	5'-GGTTCAAAAgcgCTGGATGTCGTTACCCAAG-3'
H573A antisense	5'-CGACATCCAGcgcTTTTGAACCGGTAAACAGTGCC-3'
L574A sense	5'- CAAAACAT gcgGATGTCGTTACCCAAGTGG-3'
L574A antisense	5'-GTAACGACATCcgcATGTTTTGAACCGGTAAACAGTGCC-3'
L574R sense	5'-CAAAACATcgcGATGTCGTTACCCAAGTGG-3'
L574R antisense	5'-GTAACGACATcgcGATGTTTTGAACCGGTAAACAGTGCC-3'
D575A sense	5'-AAACATCTGgcgGTCGTTACCCAAGTGGGTGC-3'
D575A antisense	5'-GGTAACGACcgcCAGATGTTTTGAACCGGTAA-3'
H573A-L574A-D575A sense	5'-GGTTCAAAAgcggcggcgGTCGTTACCCAAGTGGGTGC -3'
H573A-L574A-D575A antisense	5'- GGTAACGACcgccgccgcTTTTGAACCGGTAAACAGT-3'

Plasmid pET 24b-LnbB-FL, which encodes for wild type C-terminal His-tagged LnbB was used as template for mutagenesis. PCR were carried out using IProof polymerase (Bio-Rad) and a thermocycler PTC-200 (MJ research). Amplification was carried out in a 20-50 µl final volume reaction mixture containing plasmidic DNA (~150 ng), 0.02- 0.05 µM of sense primer, 0.02-0.05 µM of antisense primer, nuclease-free water and I-proof PCR master mix (Ref. 172-5310, Bio-Rad). A 3 min initial denaturation at 95 °C was followed by 30 cycles: 95 °C 30 s, 30 s of the corresponding annealing temperature for the different primers, 72 °C 4 min, and a 10 min final extension at 72°C. After PCR reaction, the samples were digested with DpnI for 1 hour at 37 °C, and then, transformed into *E. coli* DH5α. Positive transformants were confirmed by DNA sequencing and the final constructs were used to transform *E. coli* DH5α and BL21 (DE3) cells.

4.2.5 Protein expression trials

Construct A, B and Bα expression levels were optimized by adjusting the temperature and the time allowed for expression, either with IPTG or with autoinduction medium.

The cells were grown in overnight at 37° C, 250 rpm in LB medium and diluted 200 fold to obtain starting main culture. In the case of IPTG induction, the starting cultures were induced at OD_{600} 0.5-0.6 and OD_{600} 1 at 25 °C, 30 °C and 37 °C, for 3, 6 and 14 hours of induction.

In the case of autoinduction media, cells were allowed to grow for 12, 24, 48 and 60 hours at 22 °C, 30 °C and 37 °C in LB media supplemented with25mM Na_2HPO_4, 25mM K_2HPO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 2mM $MgSO_4$, 0.5% glycerol, 0.05% glucose and 0.2% lactose.

1 mL of cultures was taken from each experiment. Cells were harvested by centrifugation (4000g, 15 min, 4°C), resuspended in buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) and disrupted by sonication (2 µm probe, 50% amplitude, 15 sec on, 45 sec off for 3 cycles) (Brandson). After centrifugation (16000rpm, 1 h at 4 °C), supernatant and pellet were analysed by SDS-PAGE.

4.2.6 Expression and purification of wild type and mutants forms of Lacto-*N*-biosidase enzymes with His-tag in *Escherichia coli*

Recombinant cultures of *E. coli* BL21 (DE3) star cells were grown following an autoinduction protocol (Studier, 2005) in LB medium with 30 µg/mL kanamycin, 25mM Na_2HPO_4, 25mM K_2HPO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 2mM $MgSO_4$, 0.5% glycerol, 0.05% glucose and 0.2% lactose at 250 rpm and 30 °C for 24 h. Cells were harvested by centrifugation (4000g, 15 min, 4 °C), resuspended in buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) and disrupted by passing once through a cell disrupter (20000 psi) (Constant Systems, UK). After centrifugation (20000g, 4 °C, 2 h), the supernatant was loaded into a 1mL Ni^{2+}-charged HiTrap chelating column chromatography (GE Healthcare). Proteins were purified to homogeneity by a linear gradient of 50 mL of Buffer B (20 mM Na_2HPO_4, 150 mM NaCl, 500 mM imidazole at pH 7.5) at a flow rate of 1 mL/min (Na_2HPO_4 20 mM, NaCl 150 mM, imidazole 500mM, pH 7.5). Collected fractions were analysed by SDS-PAGE (8%-12% acrylamide) and fractions containing the proteins were combined and dialyzed against buffer A with a 10 kDa cut-off membrane centricon (Merck-Millipore).

4.2.7 Expression and purification of truncated Lacto-*N*-biosidase with Strep-tag in *Escherichia coli*

Recombinant cultures of *E. coli* BL21 (DE3) star cells were grown following an autoinduction protocol, cited above. Once the cells were disrupted, the lysate was centrifuged and the pellet was discarded. The clarified lysate was loaded onto 1 mL Strep-Taq column (GE Healthcare). The protein was eluted with 2.5 mM d-desthiobiotin in Buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5). Collected fractions were analysed by SDS-PAGE (12-16% acrylamide) and fractions containing the proteins were combined and dialyzed against buffer A with a 10 kDa cut-off membrane centricon (Merck-Millipore).

4.2.8 Purification of Lacto-*N*-biosidase enzymes using FPLC

All proteins, except loop mutants and the Strep-tagged construct F, were purified further using an XK16/60 Superdex 200 column (GE Healthcare), preequilibrated with buffer A. The calibration curve was obtained with 1-2 mg/mL of standards (Dextran Blue (>2000kDa), Alcohol dehydrogenase (150kDa), BSA monomer (66 kDa), Carbonic anhydrase (29kDa), Cytochrome C (12.4 kDa), Aprotinin (6.5 kDa) Sigma-Aldrich). All purifications were done in an ÄKTA FPLC (GE Healthcare).
1-2 mL of concentrated proteins from affinity purification was loaded into the Superdex 200 column by loop injection. Elution was done with buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) at a flow rate of 1 mL/min. SDS-PAGE analysis was performed for all fractions collected, and fractions containing the protein of interest were pooled, concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore) and stored at 4°C. In general, the elution yielded a single peak of the target protein with >95% of purity.

4.2.9 Protein Identification

Construct A solution was analyzed by mass spectrometry (MALDI-TOFF MS) for molecular mass determination in a Bruker Daltonics spectrometer with 2,5-dihydroxybenzoic acid as matrix. Reflectron, positive ion mode, 19 kV acceleration voltage, and 20 kV reflector voltage were used.

Other soluble proteins were analysed by peptide mass fingerprint. In those cases, pure proteins obtained from acrylamide gel bands were digested with trypsin (Trypsin Gold Mass Spectrometry Grade, Promega) and peptides were identified using Mascot (Matrix Science, London UK) by search on NCBInr database.

Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK) with a laser light wavelength of 632.8 nm and a scattering angle of 173 degrees. Temperature was set at 25°C. Protein solutions in buffer A were measured without previous dilution.

4.2.10 Size exclusion chromatography (HPSEC)

HPSEC analyses to determine molecular mass estimation of the protein in solution were performed on an Agilent 1100 HPLC system (Agilent) equipped with a UV detector. 10 µL of each sample or standard was injected into a Shodex-803KW column (8.0 x 300 mm) thermostated at 25°C, and Buffer D (50mM Na_2HPO4, 300mM NaCl, pH 7.0) as eluent at a flow rate of 0.8 mL/min. The calibration curve was obtained with 1-2 mg/mL of standards (Dextran Blue (>2000kDa), Alcohol dehydrogenase (150kDa), BSA monomer (66 kDa), Carbonic anhydrase (29kDa), Cytochrome C (12.4 kDa), Aprotinin (6.5 kDa) Sigma-Aldrich).
All samples were filtered with PVDF 0.45 µm filters (Teknokroma) prior to be injected into the column.

4.2.11 Kinetics of the hydrolase activity

Lacto-*N*-biosidase activity assays with *p*-nitrophenyl β-lacto-*N*-bioside (Toronto Research Chemicals Inc.) were performed at 30 °C in a volume of 0.1 mL using 96-wells microtiter plates measuring *p*-nitrophenol concentration at 405 nm. Enzymatic reactions were performed in a reaction mixture containing 0.25 mM substrate and Buffer C (50mM of citrate-phosphate buffer, pH 4.5). After pre-incubation for 5 min at 30 °C, reactions were initiated by addition of the enzyme (20 nM-4 µM, specific concentration is determined in each experiment). At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed. The molar absorption coefficient of *p*-nitrophenol at 405 nm was 5117 $M^{-1} \cdot cm^{-1}$ at pH 4.5 at 30 °C.
Initial rates (v_o) were obtained from the linear progress curve of release of *p*NP, and they were expressed as $v_o/[E]$ in s^{-1}.

4.2.12 Kinetic complementation assays

Enzymatic reactions with constructs A and F were performed using standard conditions mentioned in C2.2.11 in a Bravo liquid handling Robot (Agilent). After pre-incubation of construct A (50 nM) and F (25 nM-500 nM) for 5 min at 30 °C, addition of the substrate (0.25 mM of *p*-nitrophenyl β-lacto-*N*-bioside) initiated the reaction. Triplicate measurements at each concentration were performed.
Enzymatic activities of construct A (50 nM) in the presence of BSA at 1:1, 1:4 and 1:10 ratio were also analysed using standard conditions.

4.2.13 Differential scanning fluorimetry (ThermoFluor assay)

ThermoFluor assay to assess the stability of proteins was performed. 5 µM of protein solution in Buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) was added to 200X SYPRO Orange dye at 50 µl of final volume. The scan of the fluorescence was done at 492nm/516nm, excitation/emission, by increasing 1 °C each min, from 25 °C to 75 °C. Temperature and fluorescence were followed using qPCR equipment, Rotor- Gene 3000 (Corbett Research).

4.3 RESULTS AND DISCUSSION

According to CAZY classification (Cantarel *et al*., 2009), the GH20 family is mainly composed of both prokaryotic and eukaryotic β-*N*-acetylhexosaminidases with different substrate specificities, EC.3.2.1.52 and EC.3.2.1.140. This family is very wide, with more than 2000 proteins identified. However, only thirteen of them have the 3D structure solved (Figure 4.3). Their domain organization is quite diverse, being the catalytic module assigned to a GH20 domain (PFAM code PF00728).

Among known structure GH20 proteins, the catalytic GH20 domain appears to be a single domain, as it is shown in Dispersin B from *A. actinomycetemcomitans*. Or it can be accompanied by other domains with quite diverse functionalities, such as; a non-catalytic domain (commonly called GH20b) which is conserved in most GH20 enzymes although with unknown function, several lectin domains, different carbohydrate binding domains (CBM), and other domains of unknown function. In addition, the oligomerization state of the different hexosaminidases is also varied, from monomeric to dimeric proteins.

Several studies on the structure-function relationships at the active site have been performed (Mark *et al*., 2001; Pluvinage *et al*., 2011; Prag *et al*., 2000; Sumida *et al*., 2012; Williams *et al*., 2002) but only a few of them have addressed the structural organization of these enzymes (Ito *et al*., 2013; Mark *et al*., 2001; Sumida *et al*., 2009).

In this thesis a detailed analysis of the structure-function relationships among GH20 enzymes is reported.

4.3.1 Structural domains organization of GH20 enzymes

Using data integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases), a consensus organization for the known structure GH20 enzymes has been created (Table 4.2). Regardless of the number of accompanying domains, in this study two different models of organization are n proposed (Table 4.2): Model A, comprising of proteins with at least both GH20b and GH20 domains, which is characterized by a GH20b-GH20-α architecture, where GH20 is always accompanied by GH20b at the N-terminus and followed by an extra α-helix after the GH20 (β/α)8-barrel, and Model B with proteins that only present the GH20 domain without GH20b, or the α-helix .

To date, eleven of the thirteen GH20 enzymes structures fit into Model A (Figure 4.3). As summarized in Table 4.2, five of them are monomeric proteins in solution (SpHex, ScHex, Hex1T, SmCHB, and LnbB), whereas other four are dimeric proteins (HexA, Hex B, OfHex1, and GcnA). The oligomerization state of the *Arthrobacter aurescens* (NahA) and *Bacteroides fragilis* (Bf3009 a) enzymes remains unknown.

Table 4.2 Domain organization of GH20 β-*N*-acetylhexosaminidases.
Data obtained using integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases). PDB structure accession code and chain identifier used for structural analysis are indicated and domains actually present in the PDB structure are underlined.[1].*: Oligomerization state in solution is unknown.

Model A

Enzyme	Organism	Length (aa)	PDB	Domains[1]	GH20b-GH20-α location (aa)
Monomeric					
β-*N*-acetylhexosaminidase (SpHex)	*Streptomyces plicatus*	506	1HP5[A]	GH20b-GH20-α	19-494
β-*N*-acetylhexosaminidase (ScHex)	*Streptomyces coelicolor* A3(2)	535	4C7G[A]	GH20b-GH20-α	11-511
β-*N*-acetylhexosaminidase (Hex1T)	*Paenibacillus* sp.	978	3GH5[A]	GH20b-GH20-α-Lectin-CBD	15-493
Chitobiase (SmCHB)	*Serratia marcescens*	885	1QBB[A]	CHB_HEX-GH20b-GH20-α-CHB_HEX_C	215-814
Lacto-*N*-biosidase (LnbB)	*Bifidobacterium bifidum*	1112	4JAW[A,B]	GH20b-GH20-α-Lectin-CBM32-Iglike	34-515
β-*N*-acetylhexosaminidase (NahA) *	*Arthrobacter aurescens*	540	3RCN[A]	GH20b-GH20-	5-506
β-*N*-acetylhexosaminidase βf3009) *	*Bacteroides fragilis*	518	4PYS[A,B]	GH20b-GH20-α	21-498
Heterodimeric					
β-*N*-acetylhexosaminidase (HexA)	*Homo sapiens*	529/556	2GK1 [A,B,C,D,E,F,G]	GH20b-GH20-α	α:23-510 β:56-540
Homodimeric					
β-*N*-acetylhexosaminidase (HexB)	*Homo sapiens*	556/556	1NP0[A,B]	GH20b-GH20-α	56-540
β-N-acetylhexosaminidase (OfHex1)	*Ostrinia furnacalis*	593/593	3OZO[A]	GH20b-GH20-α	65-574
β-N-acetylhexosaminidase (GcnA)	*Streptococcus gordonii*	627/627	2EPN[A,B]	GH20b-GH20-α-Domain III	2-414

Model B

Enzyme	Organism	Length (aa)	PDB	Domains[1]	GH20 domain location
Monomeric					
β-1,6-*N*-acetylglucosaminidase Dispersin B (DspB)	*Aggregatibacter actinomycetemcomitans*	361	1YHT[A]	GH20	20-342
β-*N*-acetylhexosaminidase (StrH)	*Streptococcus pneumoniae* R6	1312	3RPM[A,B]	GH20-GH20-G5-G5	190-538 635-972
	Streptococcus pneumoniae TIGR4	1312	2YL8[A] 2YL9[A,B,C, D]	GH20-GH20-G5-G5 GH20-GH20-G5-G5	190-538 635-972

For the monomeric proteins, SpHex and ScHex share 94% sequence identity, and their architectures contain the GH20b domain followed by the catalytic GH20 domain (Mark *et al*., 2001; Thi *et al*., 2014). Hex1T consists of four domains -GH20b and GH20, and domains III and IV-. Domains III and IV are described as concanavalin A-like lectin and cellulose-binding domains respectively, according to the Superfamily database (Supfam). The Hex1T crystal structure is a truncated form (residues 1-502), with only the GH20b and GH20 domains. Additional domains could help to interact with lipophilic substrates (Sumida *et al*., 2009). SmCHB was the first GH20 structure to be reported (Tews *et al*., 1996). It is also made up of four domains, where the N-terminal domain (CHB_HEX) is annotated as a carbohydrate binding module (although its function remains unknown) and the C-terminal domain (CHB_HEX_C) as an immunoglobulin-like fold domain (Supfam). The central domains correspond to GH20b and GH20 domains (Prag *et al*., 2000). LnbB (which exhibits less than 25% amino acid sequence identity to all the characterised GH20 enzymes (Wada *et al*., 2008)) is formed by five domains. Domains I and II are the GH20b and GH20 ones. Domain III (residues

546-700) is similar to a lectin domain according to a PSI-Blast search. Domain IV (784-932) has sequence homology to CBM32 (Supfam), and domain V (962-1041) is described as a Immunoglobulin-like domain in PFam. The recent crystal structure of this enzyme corresponds to a C-terminal truncated form (41-663) with the GH20b, GH20 and part of the lectin-like domain (Ito *et al*., 2013). Finally, NahA and Bf3009, with unknown oligomerization state but included in the group of monomeric proteins in Table 4.2, present only the GH20b and GH20 domains.

For the dimeric proteins, each subunit of HexA, HexB (Lemieux *et al*., 2006; Mark *et al*., 2003a) and OfHex1 (Liu *et al*., 2011) have only the two domains, GH20b and GH20, while the GcnA structure contains an additional domain III, which is predominantly α-helical and is reported to form a significant part of the dimer interface (Langley *et al*., 2008).

In general, these accompanying domains are non-catalytic domains involved in binding and substrate specificity but still the mechanism by which this occurs remains uncertain (Cuskin *et al*., 2012; Drickamer and Fadden, 2002).

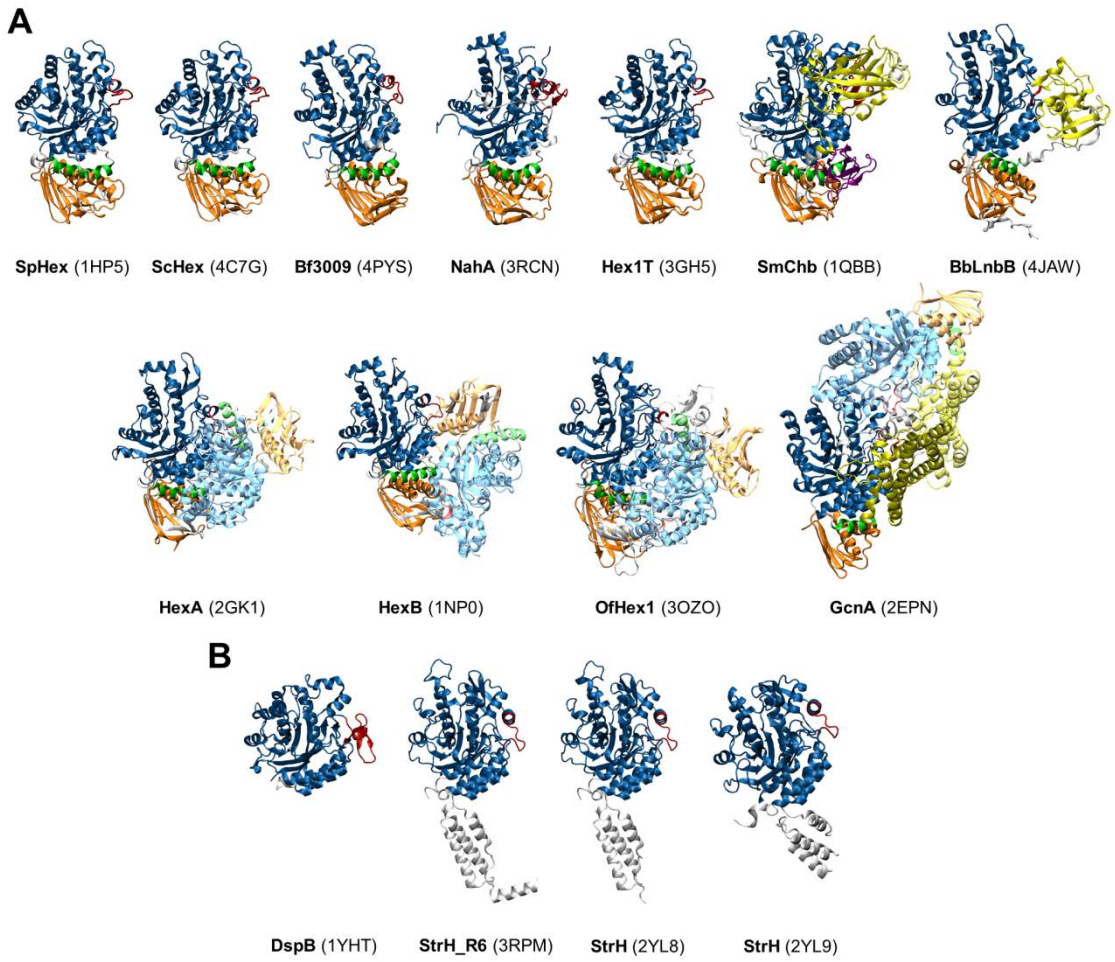


Figure 4.3 Crystal structures of GH20 β-*N*-acetylhexosaminidases.
(A) Structures according to model A architecture. (B) Structures according to model B architecture. GH20b domains are coloured in orange, GH20 domains in blue, α-helix in green, and the rest of accompanying domains in yellow and purple. Loop 2 from GH20 domain is in red. Dimer counterparts follow the same colouring scheme but with lower intensity.

The remaining two GH20 enzymes present structures that are assigned to Model B (Figure 4.3B). These are Dispersin B from *Aggregatibacter actinomycetemcomitans* (DspB) and the β-hexosaminidases from *Streptococcus pneumoniae* (StrH). Both are monomeric proteins in solution. DspB presents only one domain, the unique GH20 domain (Ramasubbu *et al*., 2005), and StrH consists of four domains, where the N-terminus is a tandem repeat of two GH20 domains, called GH20-1 and GH20-2, and the C-terminus is a tandem repeat of two domains tentatively designated as G5 with unknown function (Pfam). Truncated forms containing a single GH20-1 and GH20-2 domains of *S. pneumoniae* were crystallized (Jiang *et al*., 2011). The GH20-1 and GH20-2 domains are quite similar although specific features confer differential substrate specificities.

4.3.2 *B. bifidum* Lacto-*N*-biosidase truncated forms: Construct A, B and Bα

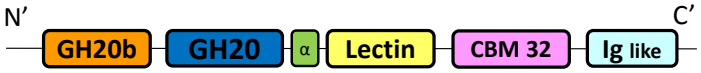
Due to the diversity in the GH20 hexosaminidase domain organization, several questions are raised. Are all the accompanying domains essential for activity? Would the GH20 catalytic domain alone be functional? And, what is the importance of GH20b?

In order to answer these questions, *B. bifidum* Lacto-*N*-biosidase was truncated according to the two different models of structure organization proposed. As the crystal structure of LnbB was not solved by the time of this study, prediction and sequence aligning online tools were used to predict secondary structure of LnbB and, subsequently, determine the different truncations.

LnbB consists of five domains: GH20b and GH20 catalytic domains, lectin-like domain (residues 546-700), CBM32 domain (784-932) and Immunoglobulin-like domain (962-1041), as has been detailed in Chapter 3.

According to the domain architecture of LnbB, and the proposed model organization of GH20 enzymes (section 3.3.1), three different truncated proteins were designed: construct A, B and Bα (Table 4.3). The truncation location was designed in areas of transition from one domain to the next, without secondary structure, as the Jpred3 prediction revealed.

Table 4.3 Truncated forms of *B.bifidum* Lacto-*N*-biosidase; construct A, B and Bα.

	Domain architecture	Residues
Wild type protein		**35-1064**
Construct A		**40-528**
Construct B		**176-497**
Construct Bα		**176-528**

Construct A (residues 40-528) follows the architectural organization of Model A enzymes and is composed of the GH20b and GH20 domains always followed by the extra α-helix. This construct would be expected to be active since several enzymes of model A exhibit just these two domains alone in the wild-type protein (such as SpHex structure and HexA, HexB and OfHex1 subunits). Furthermore, a truncated form of Hex1T (encoding the GH20b-GH20-α domains alone) is reported to be active (Sumida *et al*., 2009).

Construct B (residues 176-497) mimics Model B enzymes and is formed by the single catalytic GH20 domain alone.

Lastly, construct Bα (residues 176-528) was designed to tackle the open question of whether the accompanying GH20b domain can be removed in Model A enzymes. This construct encodes the catalytic GH20 domain with the extra α-helix at the C-terminus. Both constructs, B and Bα, were designed for the study of the behaviour of the GH20 catalytic domain in model A proteins.

Construct A gene was produced as a synthetic gene with codon optimised sequence *for E. coli* expression. It was cloned into a pET28a+ vector and was used as template for PCR amplification of construct B and Bα. The three genes of the constructs genes were successfully obtained and cloned into a pET28a+ vector with NdeI and Bpu1102 as flanking regions to generate N-terminally His-tagged truncated proteins. *E.coli* DH5α and BL21 (DE3) were transformed and expression trials were performed.

Expression trials were performed for construct A, B and Bα with both IPTG and autoinduction methods, at different temperatures and different times of of induction (see an example of Autoinduction trials in Figure 4.4). The trials carried out with autoinduction system showed higher levels of protein expression in the soluble fraction, compare to IPTG induction.

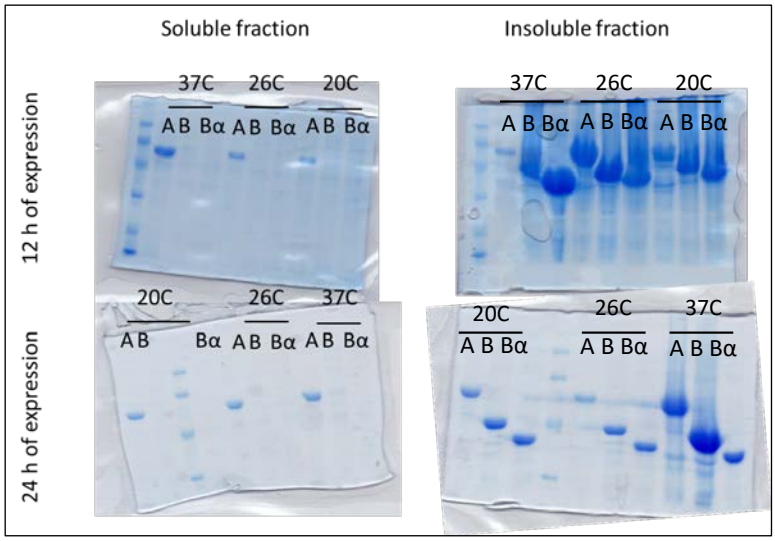


Figure 4.4 SDS-PAGE analysis of autoinduction expression trials for construct A, B and Bα.
12% SDS-PAGE with Coomassie stain analysis reveals that construct B and Bα are always observed in the insoluble fraction, independently of the temperature or time of expression, while construct A is always present in the soluble fraction in all conditions.

Final conditions for the expression of construct A were growing recombinant cells at 30 °C for 24 h with autoinduction medium. After cell lysis, the soluble fraction obtained was purified by Ni^{2+}-charged HiTrap chelating column chromatography. The chromatogram (Figure 4.5A) revealed a single peak of construct A eluted with 100 mM of imidazole (which corresponds to 20% of buffer B). SDS-PAGE analysis revealed that construct A migrated with an approximate molecular mass of 57 kDa, which agrees with the theoretical molecular mass (Figure 4.5C)
Fractions containing construct A were pooled, concentrated, and further purified on a Superdex 200 column (GE Healthcare). In this case the chromatogram also showed a single peak at 71.5 mL (Figure 4.5B). According to protein standards this elution volume corresponds to molecular masses higher than 60-100kDa, so this would indicate that construct A was eluting as a dimer.
Fractions, previously analysed by SDS-PAGE, were again pooled and concentrated with a 10 kDa cut-off membrane Centricon (Merck-Millipore). Final protein concentration was determined spectrophotometrically by the BCA method measuring Abs 590 nm using BSA as standard (Smith *et al.*, 1985). An average yield of 27mg of construct A was obtained per litre of culture.

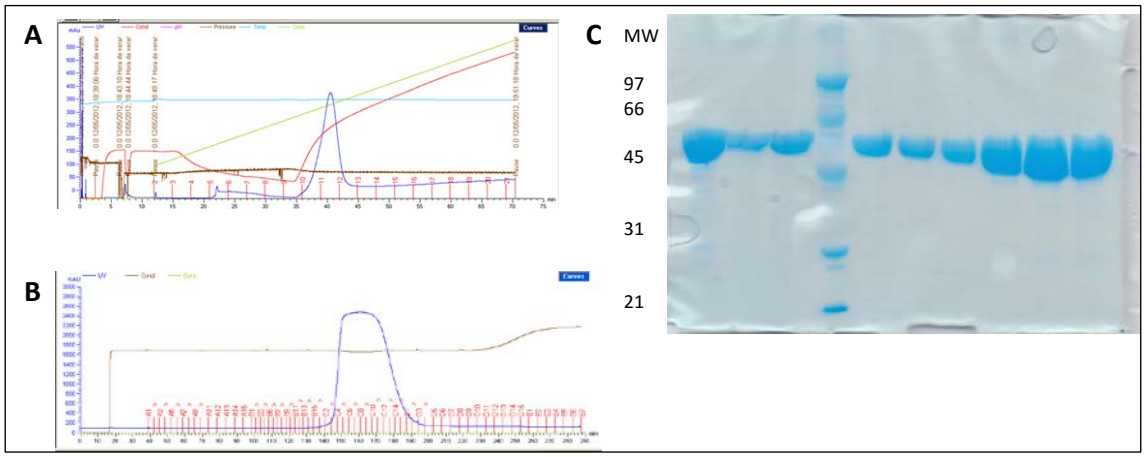


Figure 4.5 Schematic representation of the purification of Construct A.
(A) Chromatogram of affinity chromatography (IMAC). Single peak of construct A eluted with 500 mM Imidazole gradient. (B) Chromatogram of size exclusion chromatography on a Superdex 200 column. Single peak of construct A eluted with buffer A. (C) 12% SDS-PAGE of fractions from size exclusion chromatography shows pure bands of construct A (57kDa).

Construct A in solution was also analysed by mass spectrometry (MALDI-TOFF MS) and analytical size-exclusion chromatography (HPSEC). By contrast to gel filtration, the molecular mass estimation of the protein in solution was consistent with a monomeric 57 kDa protein using HPSEC. Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK). Construct A solutions in buffer A were measured without previous dilution and showed a single peak for monodisperse particles of 7 nm and PDI 0.85, which is also consistent with monomeric species (Figure 4.6).
Peptide mass fingerprint obtained after tryptic digestion by MALDI-TOF mass spectrometry confirmed the presence of peptides from construct A.

Taking into account all the results obtained, it was concluded that construct A was a monomer in solution, and the resolution of the Superdex 200 column was not suitable for molecular mass determination in a range of 60-100 kDa.

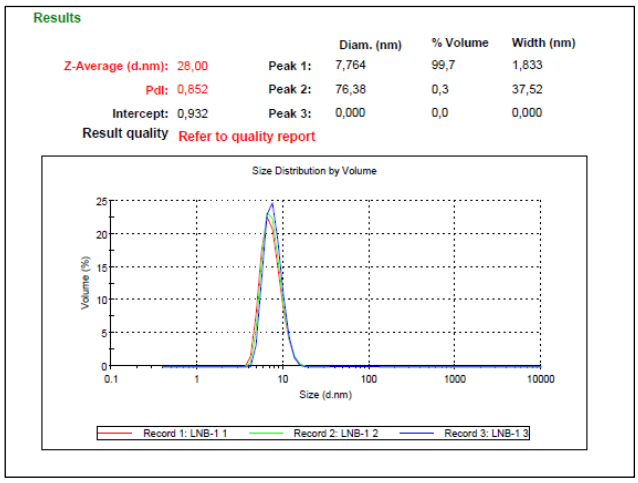


Figure 4.6 Construct A identification analysis.
(A) MALDI-TOF spectrum of construct A in solution and identification of peptides from digested construct A protein. (B) Size distribution plot by Dynamic Light Scattering. 99.7 % of the particles of the assayed sample presented a unique size particle of construct A.

On the other hand, constructs B and Bα, with a predicted molecular masess of 44 kDa and 40 kDa respectively, were only present in the insoluble pellet and no soluble protein was achieved after the different expression conditions mentioned above (see an example in Figure 4.7).

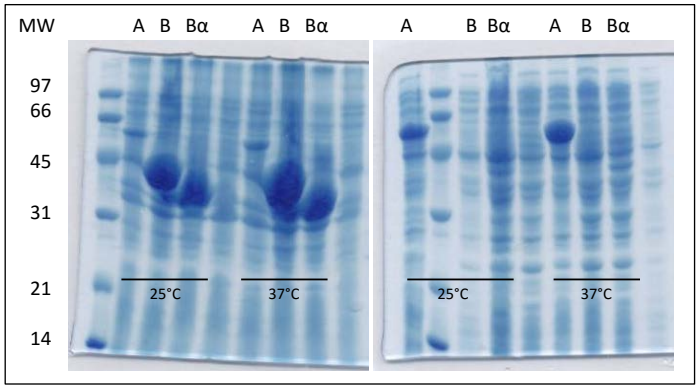


Figure 4.7. Expression trials with IPTG for construct A, B and Bα.
12% SDS-PAGE with Coomassie stain. Gel in the left for insoluble fractions, and gel in the right for soluble fractions. A, B and Bα lines correspond to the respective truncated proteins charged in a gel after 1mM IPTG induction for 3 hours at 25°C and 37°C. Overexpressed construct B (44kDa) and Bα (40kDa) present in gel for insoluble fraction, while overexpressed construct A (57kDa) is present in the gel for soluble fractions.

This seems to indicate that the single GH20 catalytic domain of LnbB, with or without the extra α-helix, is not stable enough and yields insoluble aggregates. Thus, GH20b domain at the N-terminus of GH20 is required for LnbB expression and stability. This domain is present in all model A enzymes, and the relative orientation of GH20b and GH20 catalytic domain is also the same in all structures (Figure 4.3). The number of β-strands of the GH20b domain varies (for example five in GcnA, six in HexA and HexB and seven in SpHEx and LnbB) but the interface between both domains is well conserved: two loops (residues 54-60 of loop 1 and residues 127-130 of loop 4 in LnbB) and three helices (residues 82-99, 146-163 and the extra α-helix of the GH20 domain which is oriented toward an hydrophobic pocket). GH20b domain is not only found in GH20 enzymes but also in some GH84 hexosaminidases and GH67 glucoronidases

annotated in Pfam as GH20b and GH67N (Nurizzo *et al*., 2002) respectively, always located at the N-terminus of the catalytic domain.

Considering this scenario, it is concluded that GH20b-GH20-α architecture of model A enzymes cannot be reduced to that of model B.

4.3.3 Kinetics of the hydrolase activity of construct A

In order to determine whether construct A was the minimal functional unit or not, the catalytic activity of the truncated protein was assayed. Hydrolytic activity assay were performed in saturated substrate conditions: 0.25 mM of substrate, at pH 4.5 at 30 °C. Once the reaction was over, it was quenched with 1.5 volumes of 0.5 M glycine buffer at pH 11, and released *p*-nitrophenol was measured at 400 nm in a plate reader spectrophotometer.

The hydrolytic activity of the soluble construct A was assayed using 0.25 mM of *p*-NP-LNB as substrate. Enzymatic reactions were performed using different protein concentrations (20 nM, 100 nM, 1 µM and 4,5 µM) in a wide range of pH (3.5-7.5) of 50 mM citrate-phosphate buffer, and at 25°C and 30°C. However, construct A was catalytically inactive at the assayed conditions.

In conclusion, GH20b-GH20-α architecture of LnbB is enough to guarantee protein expression, but this architecture does not correspond to the minimal functional unit maybe requiring other accompanying domains from the C-terminus for the enzyme to be active.

4.3.4 Structural requirements in the GH20 domain for functionality

While this minimal GH20b-GH20-α architecture assures the activity of SpHex and Hex1T (Mark *et al*., 1998; Sumida *et al*., 2009) and also of dimeric proteins such as HexA, HexB and OfHex1 (Liu *et al*., 2011; Mahuran *et al*., 1985), it is not enough for LnbB.

While construct A was being studied, in 2013 Ito and co-workers presented an LnbB truncation study in which a truncated protein (37-520) equivalent to construct A also exhibited no catalytic activity. However, a longer truncated protein (37-663) composed of GH20b-GH20-α and a part of the lectin-like domain was reported to be as active as the full-length (Ito *et al*., 2013). The accompanying lectin domain turned out to be the missing element required in construct A to be fully active.

In this thesis, the interaction of the catalytic GH20 domain with other C-terminus domains was carefully analysed in all crystallized protein structures of model A (Figure 4.9). Exploring the structure, important structural features can be assigned. The lectin domain interacts with the catalytic GH20 domain, providing a remote element that folds on the active site (Figure 4.8A).

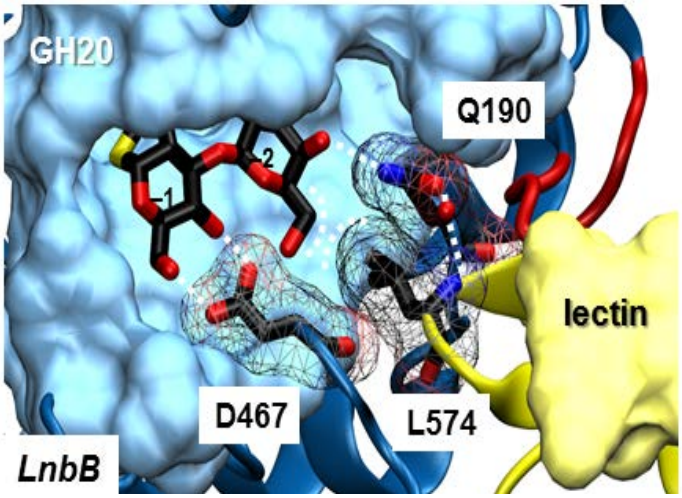


Figure 4.8 Detailed view of the remote elemt of LnbB interaction with the cavity.
Colour assignment: red: loop 2, yellow: lectin domain and blue: catalytic GH20 domain.

This element (residues 570-578) consists of a beta-turn, in which Leu574 completes the substrate binding cavity favouring a hydrophobic interaction with the hydroxymethyl group of the galactose moiety of the substrate. In addition, Leu574 is positioned between two conserved amino acids Gln190 and Asp467 of subsites –1 and –2 forming further interactions between this turn and the active site (Figure 4.8). For instance, through a water mediated hydrogen bond, the backbone nitrogen atom of Leu574 favours the orientation of Gln190 towards the O4 of the galactose unit of the disaccharide ligand. Engineered forms of LnbB in which part of the C-terminus of the lectin domain were even shorter (Ito *et al*., 2013) or completely removed as in our construct A, showed an important decrease in activity of the enzyme, indicating that the interactions observed in the structure between the remote lectin domain and the substrate binding site are essential for catalysis.

Interestingly, it has been detected that this pattern is reproduced in all proteins of model A. There is always a remote element assisting in the definition of the substrate binding cavity by means of direct interactions with conserved amino acids directly involved in protein–substrate interactions. Figure 4.9 shows the three model architectures found in GH20 model A enzyme with regard to the remote element.

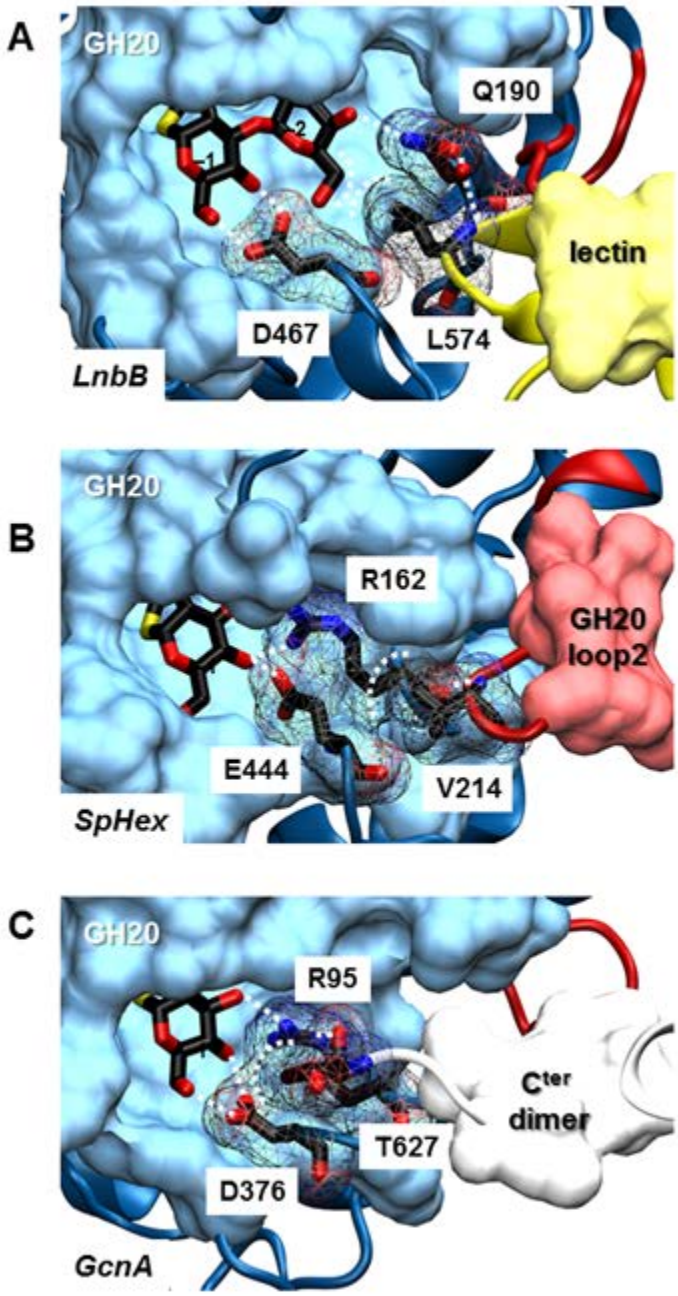


Figure 4.9 Different remote elements and their key residues that complement the active site of GH20 β-*N*-acetylhexosaminidases.
(A) *B. bifidum* lacto-*N*-biosidase. (B) *S. plicatus* and (C) *S. gordonii N*-acetylhexosaminidases. Dot lines represent interactions between residues of the remote element and the active site as indicated in Table 4.4.

Table 4.4 summarizes the exact localization of these remote elements for each member of model A and the interactions it forms with conserved active site residues. In general, this remote element is provided by a long loop of the GH20 domain itself (loop 2). In the cases where this loop is too short, the remote element comes from an accompanying domain (such as the lectin domain in the case of LnbB) or from the C-terminus of the other monomer in the case of dimeric proteins.

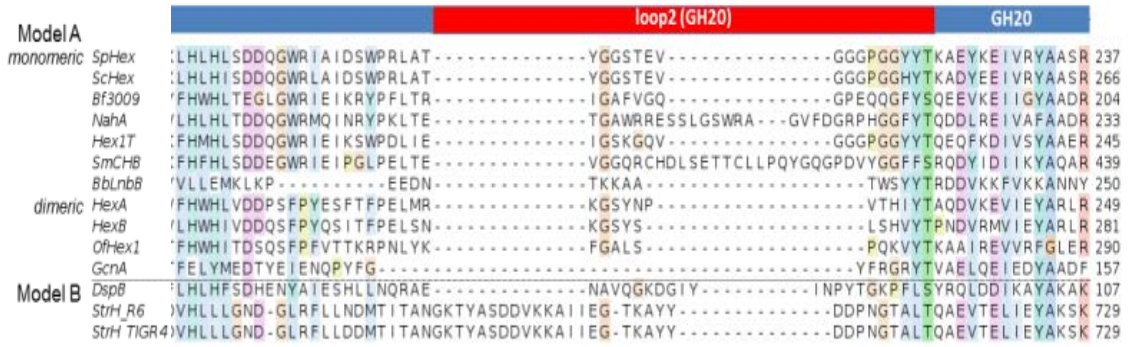


Figure 4.10 Multiple sequence alignment of loop 2 region of β-*N*-acetylhexosaminidases of known structure from family GH20.

Domains distribution is shown as coloured boxes on top. Accompanying domains are not shown. Sequences were downloaded from UniProt entry names: SpHex (O85361_STRPL), ScHex (Q9L068), Bf3009 (Q5LAT3) NahA (A1RBZ5_ARTAT), Hex1T (D2KW09_9BACL), SmCHB (CHB_SERMA), LnbB B3TLD6_BIFBI), HexA (HEXA_HUMAN), HexB (HEXB_HUMAN), OfHex1 (Q06GJ0_OSTFU), GcnA (Q6ST21_STRGN), DspB (Q840G9_AGGAC), StrH_R6 (Q8DRL6_STRR6), StrH_TIGR4 (STRH_STRPN). Alignment was performed with PROMALS. Conserved positions are coloured according to ClustalW colour scheme. Only the sequence fragments covering GH20b, GH20 domains and extra α-helix were considered in this alignment. Full image at Appendix section.

In detail, structures of SpHex, ScHex, NahA, Bf3009, SmCHB and Hex1T present this remote element coming from the extended loop 2 of the GH20 domain itself (Figure 4.10). This loop positions a hydrophobic residue, such as Val214 in SpHex (equivalent to Leu574 in LnbB), into the second shell of the subsite –1 cavity just behind the conserved arginine and aspartate in all known GH20 hexosaminidases, Arg 162 and Asp444 (equivalent to Gln190 and Asp467 in LnbB, see alignment in Figure 4.10). These amino acid residues can form hydrophobic interactions between their side-chains and also, a hydrogen bond between the backbone NH group of Val214 and the carbonyl group of the backbone of Arg162 (Figure 4.9B). These interactions between residues of the remote element and the conserved arginine and aspartate of subsite –1, are present in all the structures with long loop 2 (Table 4.4). ScHex presents Val243 of loop2 next to Arg191 and Asp473 in the same way. NahA and SmCHB show the longest loops 2 as seen in the alignment. They provide the remote elements next to the corresponding arginines, Arg146 and Arg349, and aspartates, Asp456 and Asp739 respectively. One arginine of the remote element, Arg196 in NahA and Arg399 in SmCHB, is hydrogen-bonded to the backbone of the active site arginine. Similarly, Bf3009 presents Gln209 of loop2 next to Arg153 and Asp450. In Hex1T, Val222 from the remote element is positioned behind Arg170 and Asp443 and the backbone NH group of this valine donates a hydrogen bond to the arginine carbonyl oxygen atom. Interestingly, a truncated form of Hex1T with only the GH20b and GH20 domains has been reported to be active (Sumida *et al*., 2009). Consistent with the explanations given in this study, this truncated form does not affect the definition of the active site, because this construct keeps the long loop 2 in the GH20 domain.

Table 4.4 Remote element and conserved interactions with the active site of model A GH20 β-*N*-acteylhexosaminidases.

*Residue in LnbB substituting the function of the conserved arginine but in subsite -2.

Enzyme	Loop 2 length	Remote element	GH20 conserved Arg/Asp or Glu at subsite –1	Remote element interactions with conserved Arg/Asp	
				non-polar interactions	polar interactions
SpHex	Long G209-T223	Loop 2	R162	V214	V214
			E444	V214	-
ScHex	Long G238-T252	Loop 2	R191	V243	V243
			E473	V243	-
NahA	Long G193-T219	Loop 2	R146	-	R196
			E456	-	-
Bf3009	Long G200-S214	Loop 2	R153	-	Q209
			D450	-	
Hex1T	Long G217-T231	Loop 2	R170	V222	V222
			E443	V222	-
SmCHB	Long G396-S425	Loop 2	R349	Y413	R399
			E739	Y413	-
LnbB	Short T231-T236	Lectin G570-T578	Q190*	L574	S233
			D467	L574	-
HexA	Short αG225-αT235	C-terminus subunit β A543-N552	αR178	βA548	βG549
			αE462	βY547	-
HexA	Short βG258-βT267	C-terminus subunit α A514-E523	βR211	αV519	αG520
			βE491	αV519	-
HexB	Short G258-T267	C-terminus dimer A514-E523	R211	A548	G549
			E491	Y547	-
OfHex1	Short G267-T276	C-terminus dimer A577-S594	R220	P582	E583
			E526	P582	-
GcnA	Short G136-T143	C-terminus dimer S622-T627	R95	-	T627
			D376	T627	-

By contrast, when loop 2 is short, this remote element comes from the accompanying domains, as in LnbB, or from the C-terminus domains of the other monomer in dimeric structures, such as, GcnA, HexA, HexB and OfHex1. For example, in the homodimeric structure of GcnA, domain III forms a significant part of the dimer interface (Langley *et al*., 2008). The C-terminus of domain III of chain B (622-627) winds back into the structure and contacts the catalytic GH20 of chain A (Figure 4.9C). Specifically, the C-terminus amino acid residue Thr627 projects into subsite –1 and it is adjacent to the conserved arginine and aspartate (Arg95, Asp376) and is part of the active-site pocket of the dimer partner. Besides, the carboxyl group of this C-terminal Thr627 and the Arg95 side-chain are at less than 3 Å distance, establishing

electrostatic interactions. As in LnbB, any truncated form of this C-terminus domain III is believed to be inactive. In the dimeric structures of HexA, HexB and OfHex1 (Lemieux *et al*., 2006; Liu *et al*., 2011; Mark *et al*., 2003), dimerization occurs exclusively between the GH20 catalytic domains. Dimerization is essential for catalysis since residues of one subunit structurally complete the active site of the other one. For example, Tyr456 of subunit β is part of the active site of subunit α in HexA. In addition, the C-terminus of subunit β provides this remote element very close to the subsite-binding cavity of subunit α and vice versa (Table 4.4). Particularly, the hydrophobic residue Ala548 (subunit β) of this element in HexA forms part of the second shell of the cavity behind the conserved Arg178 (subunit α) and Asp462 (subunit α), and the backbone nitrogen atom of Gly549 (subunit β) interacts by hydrogen bond with the carbonyl oxygen atom of Arg178 (subunit α). The equivalent interactions between subunits are also seen in HexB and OfHex1 crystal structures (Table 4.4).

Therefore, all GH20 structures of model A clearly show the presence of a remote element folding on the substrate-binding cavity providing amino acid residues that either complete the active site, as in the extended subsite –1 of the *S. gordonii* hexosaminidase or in subsite –2 of the LnbB, or participate in the second shell of this cavity behind the conserved arginine and aspartate residues in subsite –1 of hexosaminidases.

4.3.5 *B. bifidum* Lacto-*N*-biosidase truncated forms: Construct D and F

Further studies are presented in this work to determine the details of the minimum requirement for LnbB to be catalytically active. Four new constructs of LnbB were analysed (Table 4.5).

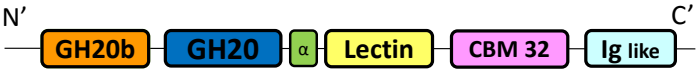
Construct C (residues 40-663) is equivalent to the active truncated protein of Ito and co-workers (residues 37-663). It contains the non-catalytic GH20b domain, the catalytic GH20 domain and a part of the lectin domain, which they named the C-terminal domain. According to the HHPred prediction studies reported in Chapter 3, the lectin-like domain consists of residues 546 to 776. Then, the truncated protein of Ito and collaborators (residues 37-663) does not contain the entire lectin domain in its sequence.

Construct D (residues 40-776) is similar to construct C, but longer. This construct contains the non-catalytic GH20b domain, the catalytic GH20 domain and the entire lectin domain as is predicted.

Construct E (residues 546-663) contains the amino acids that are lacking in construct A to be active. Construct A (residues 40-528) and the truncated protein of Ito and collaborators (residues 37-663) differ in 135 amino acids that correspond to the C-terminal domain designated by Ito and co-workers. This C-terminal domain contains the remote element, but does not consist of an entire domain, according to the HHPred results. Construct E was designed to mimic C-terminal domain from Ito and co-workers.

Construct F (residues 546-776) contains the complete sequence of amino acids for the entire lectin domain predicted by HHpred. Both constructs, E and F, contains the remote element needed for catalysis.

Table 4.5 Truncated forms of B.bifidum Lacto-*N*-biosidase; construct C, D, E and F.

	Domain architecture	Residues
Wild type protein		**35-1064**
Construct C		**40-663**
Construct D		**40-776**
Construct E		**546-663**
Construct F		**546-776**

Engineered constructs C, D, E and F were designed as C-terminal His-tagged proteins while constructs A, B and Bα were designed as N-terminal His-tagged proteins. The position of the His-tags was located according to the proximity of the active centre, in the sense of having the His-tag far away from the active centre of the protein. Then, in the constructs C, D, E and F the His-tag was located in the C-terminal region of the protein since the active centre and the tag are enough distant to each other. In contrast, the GH20 catalytic domain is located in C-terminal of the truncated proteins A, B and Bα, so the His-tag was located in N-terminal of these constructs sequences.

Constructs C, D, E and F were successfully obtained by PCR using wilt type LnbB C-terminal His-tagged DNA as template. They were cloned into a pET28a+ vector with NdeI and Bpu1102 as flanking regions to obtain C-terminally His-tagged truncated proteins. DH5α and BL21 (DE3) were transformed with the different plasmids and the expression trials for the autoinduction method were performed.

Constructs C, D and F were also expressed and purified by following the same procedure as for the wild type protein (see Chapter 3). Construct C, D and F were expressed as soluble proteins and SDS-PAGE analysis revealed that they migrate with approximate molecular masses of 72kDa, 83 kDa and 28kDa respectively (Figure 4.11 for constructs D and F). These results are in line with their theoretical molecular masses. Average yields of 17 mg of construct C, 20 mg of construct D and 15 mg of construct F per litre of culture were obtained.

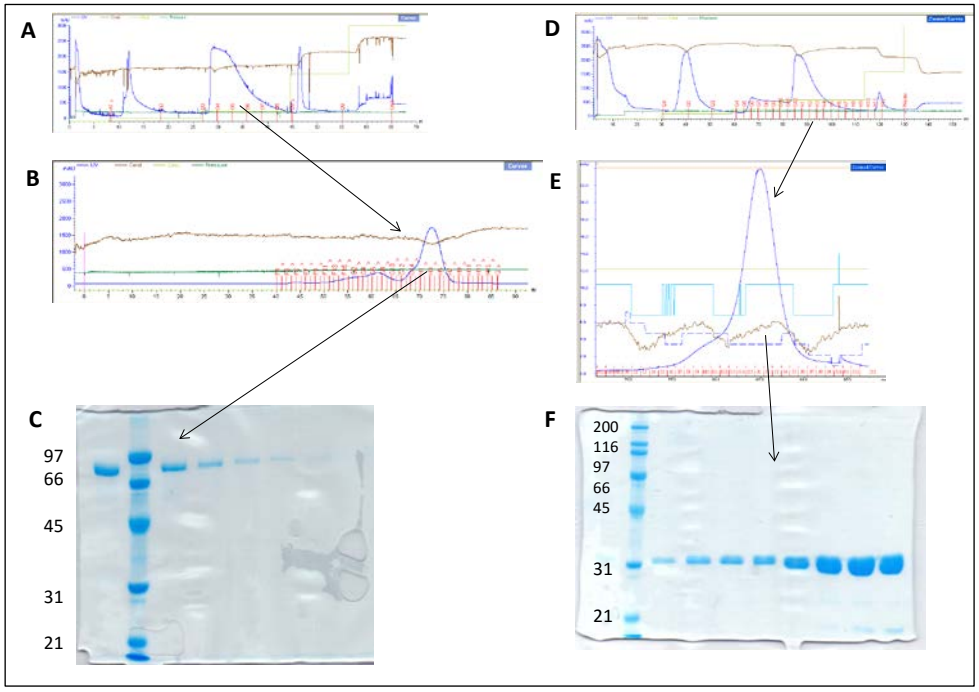


Figure 4.11 Purification and SDS-PAGE analysis of construct D and F.
(A) Chromatogram of affinity chromatography (IMAC). Single peak of construct D eluted with 50 mM Imidazole. (B) Chromatogram of size exclusion chromatography on a Superdex 200 column. Peak at 71.4 mL of construct D eluted with buffer A as a monomer. (C) 12% SDS-PAGE of fractions from size exclusion chromatography, pure bands of construct D (83kDa) are observed. (D) Chromatogram of affinity chromatography (IMAC). Single peak of construct F eluted with 100 mM Imidazole gradient. (E) Chromatogram of size exclusion chromatography on a Superdex 200 column. Single peak at 84.1 mL of construct F eluted with buffer A. (F) 16% SDS-PAGE of fractions from size exclusion chromatography. Pure bands of construct F (28kDa) are observed. Molecular masses in kDa.

By contrast, construct E was not expressed in any of the conditions performed in the expression trials (Figure 4.12). This data would confirm that residues 546-633 do not correspond to a complete domain. At this point, constructs C and E were no longer being considered for this study.

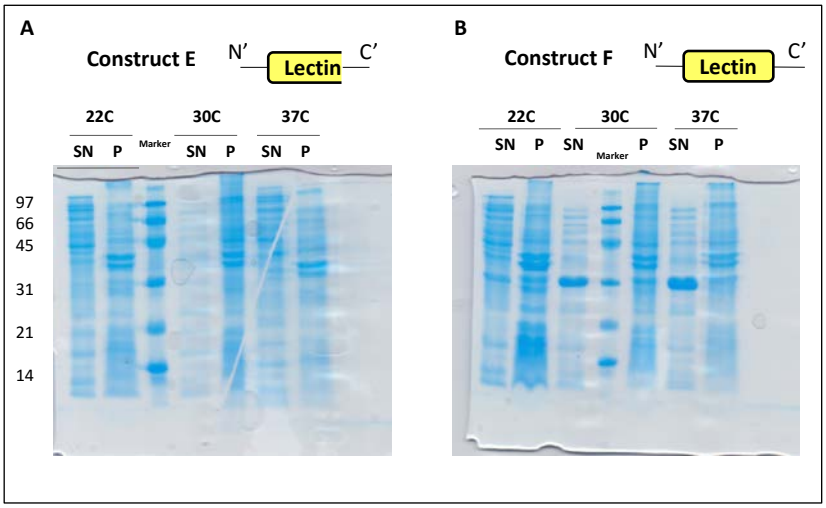


Figure 4.12 Expression trials for construct E and F.
16% SDS-PAGE of the expression of construct E (gel in the left) and F (gel in the right) at different temperatures by following the autoinduction protocol. (SN) for the soluble fraction, (P) for the insoluble fraction. A clearly overexpression of construct F at 30°C and 37°C in the soluble fraction can be observed, while no overexpression is observed for construct E. Molecular masses in kDa.

4.3.6 Kinetics of the hydrolase activity of wild type LnbB and construct D

The hydrolytic activity of the soluble proteins (20 nM) was performed in a reaction mixture containing 0.25 mM of *p*-NP-LNB and 50mM of citrate-phosphate buffer, pH 4.5 at 30 °C. Wild type protein showed a specific activity of 27.2 ± 3.5 s^{-1} and construct D was active with the same specific activity range as the wild type enzyme on the same substrate, 23.3 ± 1.0 s^{-1}.

Therefore, both the C-terminal domains in LnbB, the CBM32 and the Ig-like domains are not important for the catalytic activity, and construct D is the minimal functional unit of LnbB containing the entire domains.

4.3.7 Complementation assays of construct A and F

If construct A (composed of the GH20b and GH20-α domains) is inactive because of the lack of the remote element, would the addition of the lectin domain (as a single protein) reconstitute the active site and restore activity?

4.3.7.1 Complementation assay analysed by hydrolytic activity

To provide experimental evidence of the requirement of this remote element in LnbB to be active, a complementation assay was designed. The complementation experiment was assayed by adding construct F (the single lectin domain) into the inactive construct A solution. After pre-incubation of construct A (50 nM) and F (25 nM-500 nM) for 5 min at 30 °C, the addition of the substrate initiated the reaction. At regular time intervals, reactions were stopped and released *p*-nitrophenol was measured by absorbance at 400 nm.

Activity was restored in a concentration-dependent manner when increasing the concentrations of construct F (Figure 4.13).

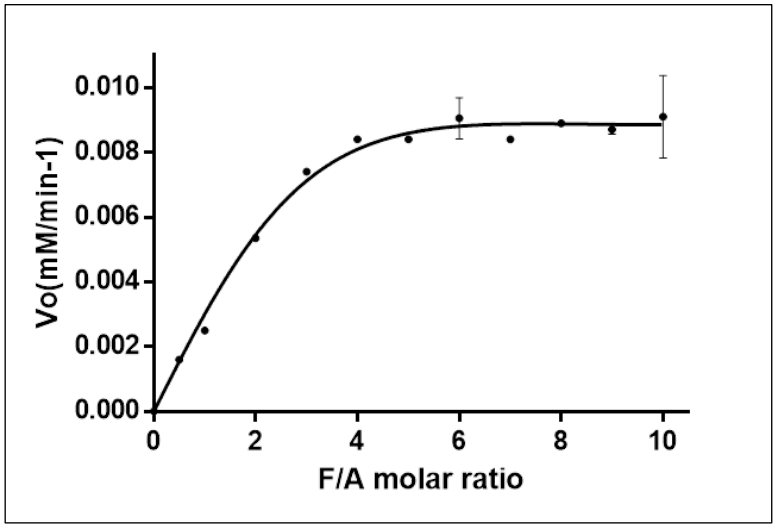


Figure 4.13 Complementation assay of construct A with construct F at different molar ratios.
The relative standard deviation for this kinetic data was less than 15%.

At a F:A molar ratio of 6:1 activity arrived at saturation at level rates of 2.9 ± 0.07 s^{-1}, which is 11% of the hydrolytic activity of the wild type enzyme. The activity was calculated taking into account the concentration of construct A (50 nm), which holds the catalytic centre.

To exclude non-specific interactions, complementation assay was performed with construct A and BSA at different concentration ratios and no activity was recovered.

Then, the restoration of the activity when the lectin domain is added to inactive construct A, indicates that the lectin domain carries the remote element needed for the structural conformation of the active cavity and therefore for LnbB being active.

4.3.7.1 Complementation assays using preparative chromatography

The complementation assay between construct A and F was followed by preparative size exclusion chromatography in order to identify the AF complex formation.

First, the wild type protein and constructs A, D and F were analysed by preparative size exclusion chromatography (Figure 4.14). Proteins (50-250 µM) were loaded to a Superdex 200 column pre-equilibrated with buffer A and calibrated with protein standards. Pure fractions were analysed by SDS-PAGE.

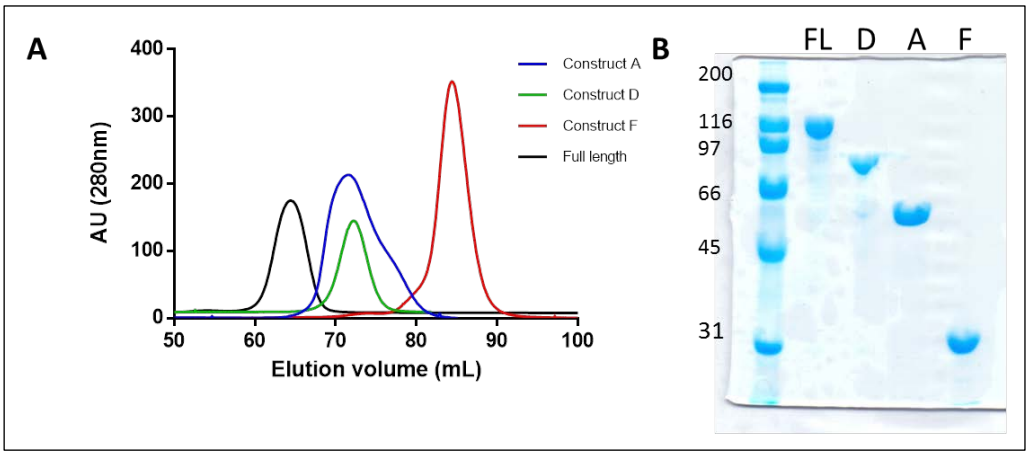


Figure 4.14 Preparative size exclusion chromatography of the wild type protein and the different constructs of Lacto-*N*-biosidase from *B.bifidum*.
(A) Chromatogram of the wild type LnbB: FL (black) and constructs A (blue), D (green) and F (red). (B) 12 % acrylamide SDS-PAGE of the eluted proteins. Molecular masses in kDa.

As observed in Figure 4.14A, construct A and D elute in the same elution volume, while in the SDS-PAGE gel of Figure 4.14B, the four proteins clearly show different molecular masses, which correlates with the theoretical masses. These results indicate that the resolution of the column with Superdex 200 is not enough for molecular masses from 57-90kDa, as was previously shown in section 4.3.2.

Construct A (42 µM), Construct F (245 µM), and the mixture of construct A and F were loaded into the Superdex 200 column. Eluted fractions were recovered for further SDS-PAGE analysis (Figure 4.15C and D). As was expected, due to the weak resolution of the column, no extra peak of the complex formed appears in the chromatogram (Figure 4.15A) when construct A and F are loaded together. According to column calibration with protein standards, the first

peak (71.4 mL) of the chromatogram in Figure 4.15A corresponds to proteins around 57-90 kDa, which includes molecular masses of construct A (57kDa) and the expected AF complex (85kDa). The second peak of the chromatogram (84.1 mL) corresponds to the molecular mass of construct F (28kDa). When the recovered fractions from the interaction between construct A and F are analysed by SDS–PAGE, both proteins, constructs A and F can be seen in the first peak of elution (Figure 4.15C). This means that construct A and F are interacting with each other, and the complex elutes with the corresponding molecular mass (85kDa). From fraction J15 to K6 (69.4-76.7 mL) both proteins are visible at the SDS-PAGE, but not in an equimolar ratio. It is believed that not all of protein A is interacting with F. So in these fractions, free construct A (57kDa) and complex AF (85 kDa) are eluting together. In fractions K7-K9 (76.3-78.4 mL), the elution of the AF complex in a 1:1 molar ratio can be observed. And in factions K11-L4 (80.4-88.7 mL) just free construct F is observed (Figure 4.15D). This means that just a small amount of them are forming the complex and there is still free construct A and F in the analysed mixture.

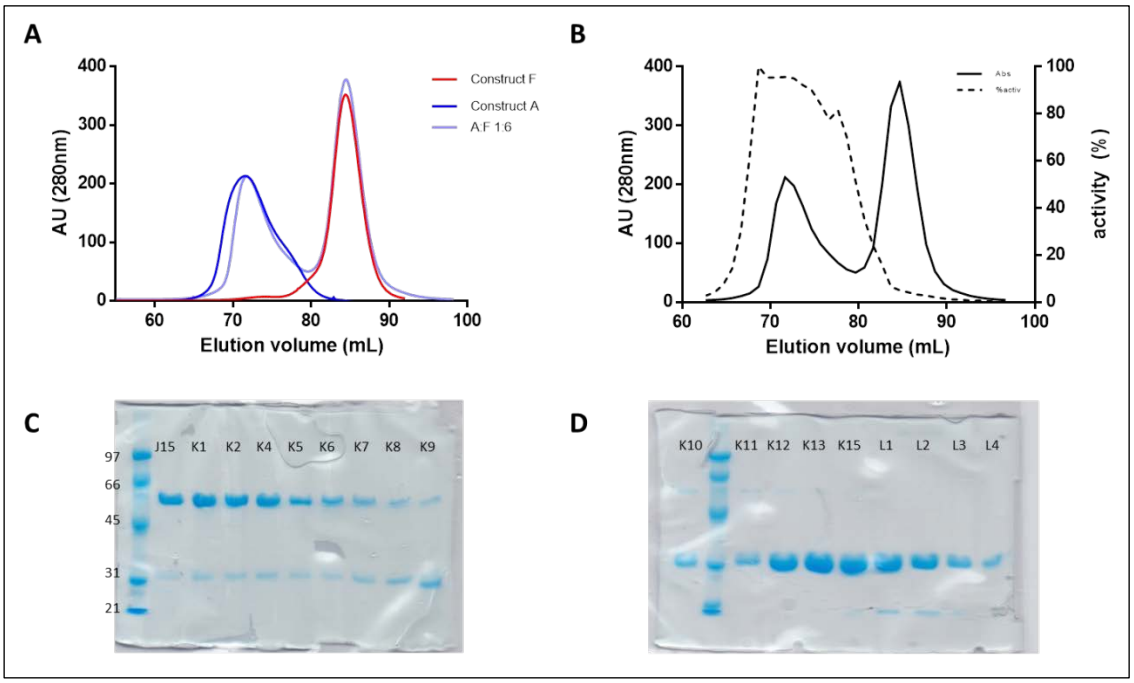


Figure 4.15 Size exclusion chromatography of the complementation assay in a ratio 1:6 of constructs A and F. (A) Chromatogram of constructs A (blue), F (red), and the interaction between both (A:F) in a 1:6 molar ratio(lilac). (B) (A) Chromatogram of the eluted fractions: absorbance at 280nm (-), hydrolase activity (•••). The relative standard deviations for the molecular masses and activities were less than 2% and 15% respectively. (C and D) 12 % acrylamide SDS-PAGE gels of the different fractions; J11-J15, K1-K15, and L1-L10. Molecular masses in kDa.

The hydrolase activity of the collected fractions were analysed (Figure 4.15B). Reaction rates were determined following the procedure cited in Materials and Methods 2.2.11. The hydrolase activity recovered in fractions J14-K9 and the presence of lectin in these fractions (seen in SDS-PAGE in Figure 4.15B, C and D) demonstrate that the complex between A and F is formed and it is active. However, as was previously mentioned, not all of the construct A protein forms a complex with construct F. By densitometry analysis of the SDS-PAGE gel in Figure 4.15B, it is determined that lectin eluted in fractions J14-K9 represents 10% of total protein in the mixture. Assuming a 1:1 interaction, just 10% of Construct A is in complex with Construct F.

4.3.7.3 Complementation assays using analytical chromatography

Since preparative size exclusion chromatography did not afford the desired resolution, new experiments were performed using analytical size exclusion chromatography (HPSEC). Proteins (5-200 µM) were injected into a Shodex-803KW column, previously calibrated with protein standards, and eluted with buffer D (Figure 4.16).

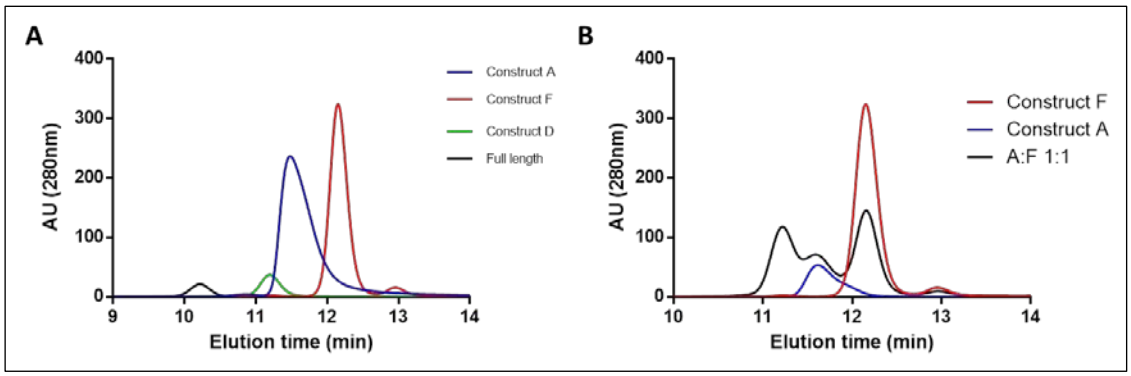


Figure 4.16 Analytical size exclusion chromatography of the wild type protein and the different constructs of Lacto-*N*-biosidase from *B.bifidum.*
(A) Chromatogram of the wild type protein (FL) and constructs A, D and F. (B) Chromatogram of constructs A, F, and the interaction between them in a molar ratio 1:1.

In this case, the column resolution was good enough for the analysis of the proteins as it is observed in Figure 4.16A. Each protein shows a single peak of elution; 10.2 min for wild type, 11.2 min for construct D, 11.7 min for construct A and 12.1 min for construct F, that corresponds to their theoretical molecular masses (112 kDa, 83 kDa, 57 kDa and 28 kDa, respectively), according to calibration with protein standards.

In Figure 4.16B, construct A and F were mixed in a molar ratio 1:1 (50 µM). In the same plot (Figure 4.16B) chromatograms of construct A and F were overlapped. The appearance of a new third peak, at 11.2 min, was observed. According to protein standards calibration, this elution time corresponds to 85 kDa, which clearly demonstrates that a complex among both proteins is formed. Again, the complex formation did not consume the total amount of the different constructs, since the corresponding peaks (11.7 min for construct A and 12.1 min for construct F) also appear in the chromatogram.

A saturation assay of the complementation of construct A and F was also performed by HPSEC (Figure 4.17). Construct A (14 µM) and construct F (14-140 µM) were incubated for 5 minutes at room temperature before being loaded into a Shodex-803KW, previously calibrated with protein standards.

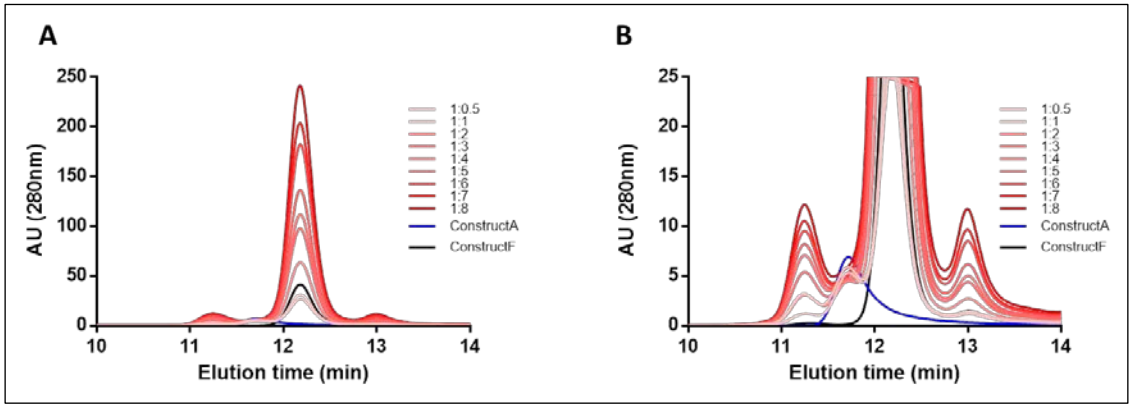


Figure 4.17 Saturation assay of the complementation of construct A and F analysed by HPSEC.
(A) Chromatogram of saturation assay. (B) Zoom of the chromatogram of saturation assay.

The absorbance of the peak corresponding to the AF complex (11.2 min) increases while increasing concentrations of construct F (12.1 min) are added to the reaction, at these conditions. Concentration of construct A is always constant (14 µM) and it can be observed in Figure 4.17B that the peak decreases while the complex is forming. In this case, saturation is not achieved at 1:10 molar ratio of construct A:F.

A peak at 13 min can be observed in Figure 4.17 impurity peak areas are constant during the assay. Figure 4.18A shows an injection of purified construct F alone, and the peak at 13 min appears. When purified fractions of construct F are analysed by SDS-PAGE (Figure 4.18B), a smaller band than 21 kDa is observed. It may correspond to the impurity peak observed at 13 min of elution time.

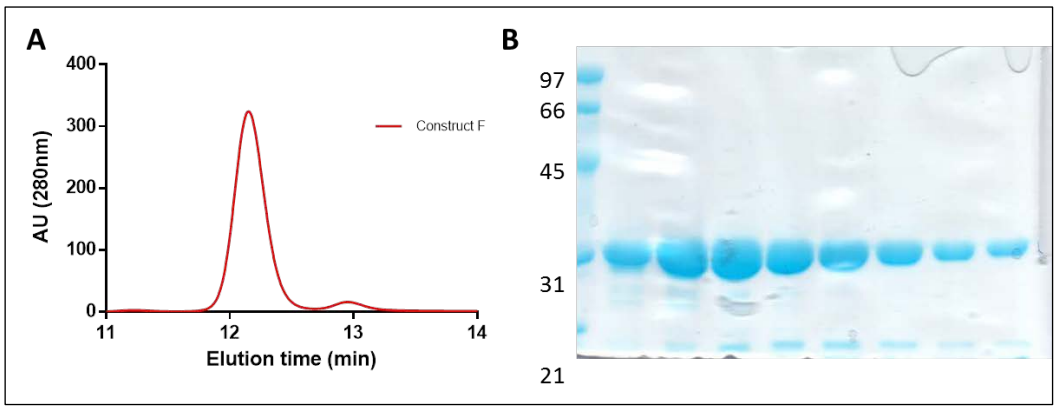


Figure 4.18 Evidence of the presence of an impurity in construct F purification.
(A) Chromatogram of construct F. (B) SDS-PAGE of fractions from construct F purification. Molecular masses in kDa.

4.3.8 Isolation of the complex formed by construct A and F

At this point, pure AF complex has not been obtained yet. In previous reactions, a mixture with free proteins of construct A and F were always present together with the complex. In order to characterize the AF complex, an isolation assay of the complex was designed.

A different tag purification strategy was designed for isolating the AF complex. Taking advantage of affinity purification methodology, a different two tag strategy, His and Strep tags, was proposed for AF complex isolation (Figure 4.19A). While maintaining the same construct A

as for the rest of the experiments cited (N-terminal His-tagged truncated protein), a new recombinant protein of construct F was obtained, in this case with a Strep tag in the carboxyl terminus. Briefly, the strep-tag was introduced by PCR using construct F gene with C-terminal His-tag as the template and the reverse primer carrying the streptavidin sequence. PCR product was introduced into a pET28a+ vector with NdeI and Bpu1102 as flanking regions. *E.coli* DH5α and BL21 (DE3) were transformed with a positive sequenced plasmid. Cells were grown at 30°C for 24 h in autoinduction medium. After cell disruption, the clarified lysate was loaded onto 1 mL Strep-Taq column (GE Healthcare). The protein was eluted with 2.5 mM d-desthiobiotin buffer. Collected fractions were analysed by SDS-PAGE (16% acrylamide) that revealed a protein migrating at, approximately, 28kDa of molecular mass. Fractions containing the proteins were pooled and concentrated before being purified on a Superdex 200 column (GE Healthcare). Fractions were then pooled again and concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore). Final protein concentration was determined spectrophotometrically by the BCA method measuring A590 nm using BSA as standard (Smith at al., 1985). The yield of the purification was 17mg of construct F (Strep-tagged) per litre of initial culture.

The idea was to obtain pure AF complex from the mixture of construct A and F tagged with His and Strep tag, respectively. When proteins interacted with each other, the complex was formed carrying both tags. Sequential different affinity purifications, with a HiTrap and a StrepTrap column, allowed the pure AF complex to be obtained. In Figure 4.19A, a detailed scheme of the process is shown. Starting from pure construct A and F, both proteins are mixed in a 1:10 molar ratio (26 µM:254 µM, construct A:F, respectively). In the reaction tube there were N-terminal His-tagged construct A, C-terminal Strep-tagged construct F and N-terminal His-tagged and C-terminal Strep-tagged AF complex. After the HiTrap purification with imidazole elution, Strep-tagged construct F (which was not interacting in the complex) was removed. N-terminal His-tagged construct A and N-terminal His-tagged and C-terminal Strep-tagged AF complex were eluted together. Collected fractions were then loaded into a StrepTrap column. In this case, N-terminal His-tagged construct A, which was not interacting in the complex, was removed, and pure AF complex was obtained after the elution with d-desthiobiotin.

Proteins eluted after every step of the scheme was analysed by SDS-PAGE (Figure 4.19B). Lane 1 and 2 correspond to pure N-terminal His-tagged construct A and pure C-terminal Strep-tagged construct F, respectively. The mixture of both proteins was charged in lane 3 and lane 4 (1:4 dilution of lane 3). And pure complex, obtained after sequential HiTrap and StrepTrap affinity purifications, was loaded into lane 5.

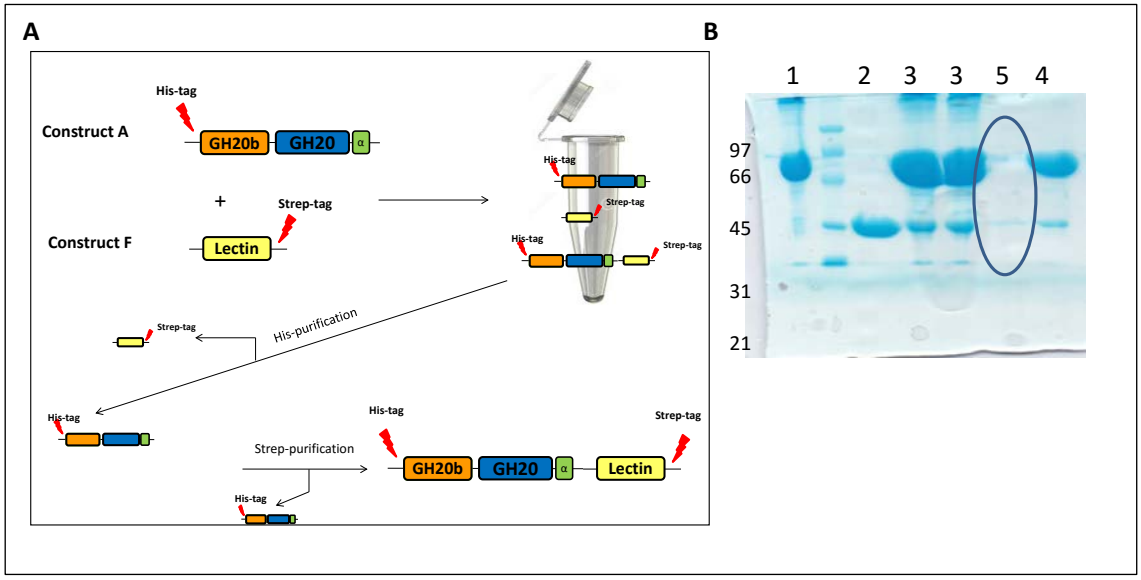


Figure 4.19 Design of AF complex isolation.
(A) Diagram of the design of the assay. (B) SDS-PAGE analysis of the different steps of the diagram; 1: pure construct A after first His purification, 2: pure construct F after first Strep purification, 3 and 4: mix of pure construct A and F, 5: pure complex after second His and Strep purification. Molecular masses in kDa.

The pure AF complex solution (shown in line 5) was concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore) and the concentration was analysed by BCA. A solution of 7.5 µM of pure AF complex was obtained.

The hydrolytic activity was analysed with 100 nM of AF complex and 0.25 mM of *p*-NP-LNB as substrate at 30°C. At regular time intervals, reactions were stopped with quenching buffer and released *p*-nitrophenol was measured by absorbance at 400 nm. A specific activity of 0.73 ± 0.19 s^{-1} was obtained. This specific activity represents 3% of the wild type protein.

This new result differs from the specific activity of 2.9 ± 0.07 s^{-1}, which was determined at 1:6 A:F interaction. In that case, the specific activity was calculated taking into account the overall amount of construct A added into the reaction (50 nM), but at that time, there was no data about how much complex was formed.

By contrast, the specific activity of 0.73 ± 0.19 s^{-1} was measured with the pure complex. Although, at the time of measuring the concentration of the complex, it was possible that the equilibrium with the free forms of construct A and construct F would be established. Nevertheless, it can be confirmed that specific activity is at least 0.73 ± 0.19 s^{-1}.

It is concluded that pure AF complex can be obtained. It seems that the interaction between both proteins has low affinity, resulting in a low K_D (dissociation constant). Despite having excess of both, construct A and F, a small amount of AF complex is formed.

4.3.9 LnbB loop mutants

Different mutants were designed to explore the interactions between the remote element of the lectin domain and the active site. The remote element (residues 570-578) consists of a beta-turn which folds on the active site of LnbB (Figure 4.20). Leu574 completes the substrate binding cavity interacting with the hydroxymethyl group of the galactose moiety of the

substrate, at -2 subsite, and it is positioned between two conserved amino acids in the GH20 domain, Gln190 and Asp467, forming further interactions between this turn and the active site. Residues His573 and Asp575 are also interacting with the cavity by backbone interactions (Table 4.**4**), while the rest of the remote element (570-578) is providing the structure needed for the interactions to be done.

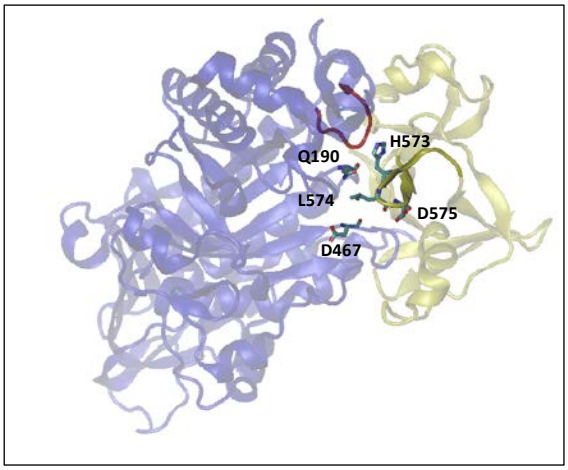


Figure 4.20 Residues involve in remote element and cavity interactions in LnbB.
Colour representation; GH20b and GH20 domains in transparent blue, lectin domain in transparent yellow, loop 2 in solid red and remote element in solid yellow.

4.3.9.1 Cloning of LnbB loop mutants

Four single mutants and one double mutant were constructed. His573, Leu574 and Asp575 were mutated to alanine to remove possible side-chain interactions. The double mutant to alanine was produced by the methodology in residues Leu574 and Asp575. In addition, Leu574 was also mutated to arginine in order to introduce a positive charge that creates repulsion between the remote element and the cavity.

Mutagenesis was carried out by site-directed mutagenesis using wild type pET24b construct as a template (C-terminal His-tagged). Mutations were confirmed by DNA sequencing and the final constructs were used to transform *E. coli* BL21 (DE3) cells for the expression and purification of mutant recombinant proteins (detailed sequence primers and procedure are cited in section 3.2.4).

4.3.9.2 Expression and purification of LnbB loop mutants

C-terminal His-tagged mutated proteins were obtained growing recombinant cells at 30°C for 24 h with autoinduction medium. Purification was done by HiTrap chromatography using imidazole as eluent. SDS-PAGE analysis revealed that all mutants migrated with an approximate molecular mass of 112 kDa, which agrees with the theoretical molecular mass. Fractions containing the mutants were pooled and concentrated with a 10 kDa cut-off membrane centricon (Merck-Millipore). Final protein concentration was determined spectrophotometrically by the BCA method measuring Abs 590 nm using BSA as standard (Smith at al., 1985). Expression and purification were very similar for all the mutated proteins, and the average yields were 19-21 mg of protein per litre of culture. All proteins were considered to be >90% pure by Coomassie staining after SDS-PAGE (Figure 4.21).

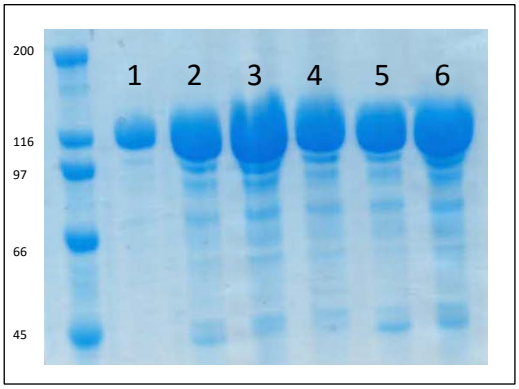


Figure 4.21 SDS-PAGE analysis of the wild type and loop mutant proteins.
Lane 1: wild type protein, Lane 2: H573A protein, Lane 2: L574A protein, Lane 4: L574R protein, Lane 5: D575A protein, Lane 6: L574A_D575A protein.

4.3.9.3 Differential scanning fluorimetry (Thermofluor)

Protein stability of the mutated proteins was evaluated by ThermoFluor assay. Fluorescence-based thermal stability assay or differential scanning fluorimetry (DSF) (Pantoliano *et al*., 2001) is a temperature-based assay to assess the stability of proteins. The methodology takes advantage of the fact that the fluorescence of many non-specific protein-binding dyes increases with increasing hydrophobicity of their environment. It is based on the detection of changes in the exposure of hydrophobic areas upon heat denaturation. The dye is quenched in aqueous solutions, but when the aromatic moieties of the dye intercalates into a hydrophobic pocket, it regains its fluorescence. A sharp sigmoidal curve allows for the calculation of a melting temperature (Tm), which corresponds to the temperature where the protein is 50% unfolded (Niesen *et al*.; Vivoli *et al*., 2014). A positive shift in the melting temperature ΔTm can be coupled to an increase in structural order and a reduced conformational flexibility, whereas a negative ΔTm, indicates that induced protein structural changes lead to a more disordered protein conformation (Boivin *et al*., 2013).

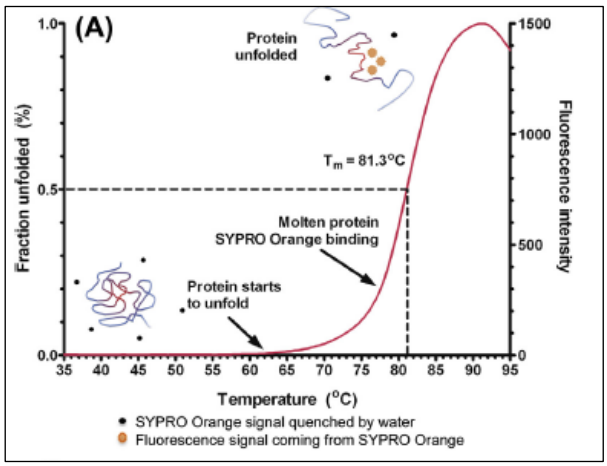
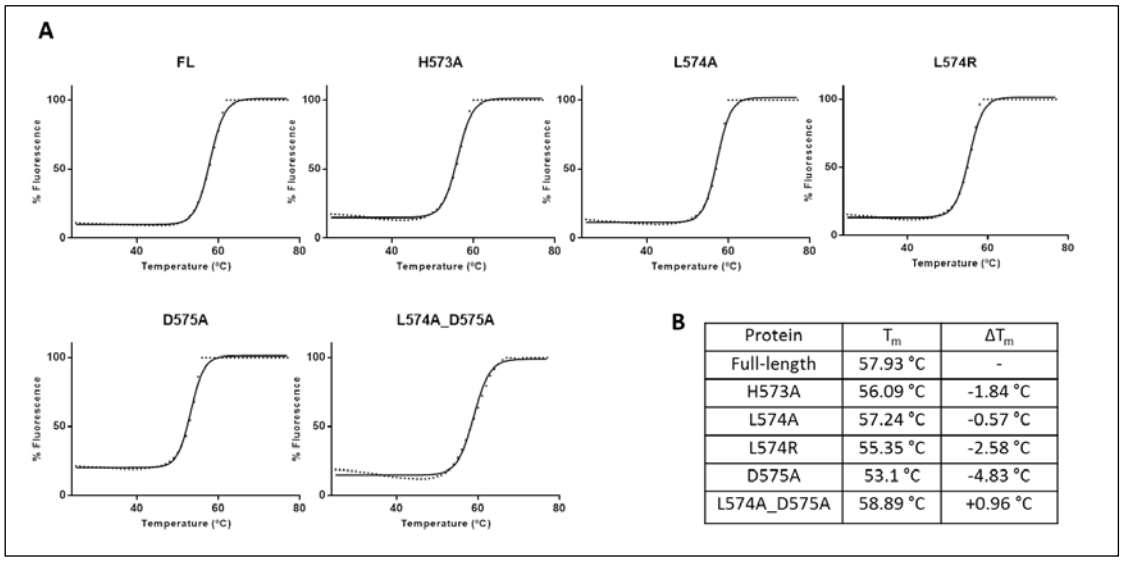


Figure 4.22 Typical thermal denaturation assay using Thermofluor.
Image from Boivin *et al*., 2013.

Different dyes can be used, and in this case SYPRO orange dye was chosen, which has a maximal absorption at 470 nm and maximal emission at 569 nm (Boivin *et al*., 2013).

SYPRO Orange dye was added to the corresponding proteins (5 µM), wild type LnbB and loop mutant proteins. Once the reactions were over, the results of fluorescence were plotted as a function of temperature (Figure 4.23).



Protein	T_m	ΔT_m
Full-length	57.93 °C	-
H573A	56.09 °C	-1.84 °C
L574A	57.24 °C	-0.57 °C
L574R	55.35 °C	-2.58 °C
D575A	53.1 °C	-4.83 °C
L574A_D575A	58.89 °C	+0.96 °C

Figure 4.23 Graphical representation of the analysis by differential scanning fluorimetry of wild type LnbB and loop mutant proteins.
(A) Denaturation curves fitting two state model. (B) T_m corresponds to different melting temperatures, and ΔT_m are related to wild type protein T_m.

All proteins followed the expected shapes of a thermal denaturation profile. The characteristic sigmoidal shape of the raw data shows a progressive rise in fluorescence to a maximum, where the protein is denatured. As soon in Figure 4.23B, melting temperatures of the mutated proteins do not significantly differ from the wild type LnbB which means that the structural stability of the protein is not affected by the exchange of the amino acids.

These minor differences in temperature support a well-folded structure for the mutant proteins, and hence, changes of the catalytic properties can be assigned to the replacement of the target amino acid residues, and not to a different folding of the mutated proteins

4.3.9.4 Kinetics of LnbB loop mutants

The hydrolytic activity of all mutants was analysed with 50 nM of protein and 0.25 mM of substrate.

All mutated proteins exhibited hydrolytic activity (Table 4.6), with rates lower than wild type protein. While mutants to alanine reduce their specific activity at maximum of 10 times less than the wild type protein, the mutation of Leu574 to arginine resulted in a significant decrease of the specific activity (118-fold less) compared to the wild type. This means that the introduction of a positive charge in the side chain of the residue (Arg) instead of the hydrophobic existing side chain of Leu disturbs the interaction between the remote element and the active centre, and then, a significantly decrease of the catalytic activity is shown.

Table 4.6 Kinetic parameters of the wild type LnbB and loop mutant proteins.
The reactions were done at 20 nM of wild type and 50 nM of loop mutants, 0.25 mM of substrate, 30 °C in buffer C at pH 4.5.

Protein	Specific activity (s^{-1})
Wild type	27.2 ± 3.5
H573A	6.0 ± 1.32
L574A	2.8 ± 1.89
L574R	0.23 ± 0.09
D575A	3.78 ± 0.97
L574A_D575A	4.53 ± 0.75

It is concluded that the implication of the remote element is not strictly dependent on the specific interaction of residues 573-575 within the active centre, but it depends on the overall structure of the entire lectin domain.

4.4 CONCLUSIONS

This structural-functional analysis allows us to come to different conclusions. First, it has been identified two levels of domain organization in GH20 enzymes: model A with GH20b-GH20-α architecture and model B with the single catalytic GH20 domain alone. As Lacto-*N*-biosidase from *B. bifidum* constructs show, model A enzymes cannot be reduced to model B, since GH20b domain is a structural requirement in the N-terminus of GH20 to assure protein stability. Secondly, GH20b-GH20-α architecture is not always the minimal functional unit since the isolated model A domain of LnbB is inactive. Based on these results and analysis of the structures of GH20 enzymes, it has been proposed a broader mechanism by which important interactions in the substrate binding cavity of the catalytic GH20 domain are provided by a remote element in all structures, which must be preserved to ensure a proper definition of the active site. The remote element can directly be provided by a long loop 2 of the GH20 domain itself. In this case, GH20b-GH20-α can act as a minimal functional unit and large multidomain enzymes can be engineer to reduced forms with a minimal model A architecture. In other circumstances, when loop 2 is short, proteins dimerize or present an accompanying domain in the C-terminus of GH20 that provide this remote element.

In this thesis the remote element of LnbB has been studied, and in this case, loop 2 from GH20 domain is not long enough to act as the remote element, and this element is provided by the consecutively C-terminal domain, a lectin like domain. It has been restored the hydrolytic activity of construct A (inactive truncated form of LnbB) when the lectin domain (construct F, which contains the remote element) was added to the reaction. It is confirmed that the lectin domain assists in the definition of the substrate binding cavity providing an essential loop (remote element) to shape the active site.

Moreover, a mutational study of the residues directly implied in the interactions of the remote element with the catalytic GH20 domain was also performed. Although mutations of the remote element did not yield a complete inactive enzyme, there was an important decrease of the activity for L574R mutant. This suggests that modifications in the interaction area of both domains yield less active proteins. These overall results confirm that the lectin structure of the lectin is required for the catalytic activity of Lacto-*N*-biosidase.

This work suggests that protein engineering of GH20 enzymes must be designed according to the remote element. The proposed model can stimulate further studies on other members of the superfamily.

4.5 REFERENCES

Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. Protein Expr. Purif. *91*, 192–206.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. *37*, D233–D238.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Cuskin, F., Flint, J.E., Gloster, T.M., Morland, C., Baslé, A., Henrissat, B., Coutinho, P.M., Strazzulli, A., Solovyova, A.S., Davies, G.J., *et al*. (2012). How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc. Natl. Acad. Sci. U. S. A. *109*, 20889–20894.

Dowd, P.F., Johnson, E.T., and Pinkerton, T.S. (2007). Oral toxicity of beta-*N*-acetyl hexosaminidase to insects. J. Agric. Food Chem. *55*, 3421–3428.

Drickamer, K., and Fadden, A.J. (2002). Genomic analysis of C-type lectins. Biochem. Soc. Symp. *72*, 59–72.

Gloster, T.M., and Vocadlo, D.J. (2012). Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. Nat. Chem. Biol. *8*, 683–694.

Hattie, M., Debowski, A.W., and Stubbs, K.A. (2012). Development of tools to study Lacto-*N*-Biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. ChemBioChem *13*, 1128–1131.

Henn-Sax, M., Höcker, B., Wilmanns, M., and Sterner, R. (2001). Divergent evolution of (betaalpha)8-barrel enzymes. Biol. Chem. *382*, 1315–1320.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. J. Mol. Graph. *14*, 33–38.

Intra, J., Pavesi, G., and Horner, D.S. (2008). Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family. BMC Evol. Biol. *8*, 214.

Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., Ashida, H., Wakagi, T., Yamamoto, K., Stubbs, K.A., *et al*. (2013). Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-*N*-biosidase from Bifidobacterium bifidum. J. Biol. Chem. *288*, 11795–11806.

Jiang, Y.-L., Yu, W.-L., Zhang, J.-W., Frolet, C., Di Guilmi, A.-M., Zhou, C.-Z., Vernet, T., and Chen, Y. (2011). Structural basis for the substrate specificity of a novel β-*N*-acetylhexosaminidase StrH protein from Streptococcus pneumoniae R6. J. Biol. Chem. *286*, 43004–43012.

Langley, D.B., Harty, D.W.S., Jacques, N. a, Hunter, N., Guss, J.M., and Collyer, C. a (2008). Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen Streptococcus gordonii and its complex with the mechanism-based inhibitor NAG-thiazoline. J. Mol. Biol. *377*, 104–116.

Lemieux, M.J., Mark, B.L., Cherney, M.M., Withers, S.G., Mahuran, D.J., and James, M.N.G. (2006). Crystallographic Structure of Human β -Hexosaminidase A: Interpretation of Tay-Sachs Mutations and Loss of GM2 Ganglioside Hydrolysis. J. Mol. Biol. *359*, 913–929.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. *8*, 91.

Liu, T., Zhang, H., Liu, F., Wu, Q., Shen, X., and Yang, Q. (2011). Structural determinants of an insect β-*N*-acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. J. Biol. Chem. *286*, 4049–4058.

Liu, T., Yan, J., and Yang (2012). Comparative biochemistry of GH3, GH20 and GH84 B-*N*-acetyl-D-hesoxaminidases and recents progress in selctive inhibitor discovery. Curr. Drug Targets *13*, 512–525.

Mahuran, D., Novak, A., and Lowden, J.A. (1985). The lysosomal hexosaminidase isozymes. Isozymes Curr. Top. Biol. Med. Res. *12*, 229–288.

Mark, B.L., Wasney, G. a, Salo, T.J., Khan, a R., Cao, Z., Robbins, P.W., James, M.N., and Triggs-Raine, B.L. (1998). Structural and functional characterization of Streptomyces plicatus beta-*N*-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. J. Biol. Chem. *273*, 19618–19624.

Mark, B.L., Vocadlo, D.J., Knapp, S., Triggs-Raine, B.L., Withers, S.G., and James, M.N.G. (2001). Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. J. Biol. Chem. *276*, 10330–10337.

Mark, B.L., Mahuran, D.J., Cherney, M.M., Zhao, D., Knapp, S., and James, M.N.G. (2003a). Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. J. Mol. Biol. *327*, 1093–1109.

Mark, B.L., Mahuran, D.J., Cherney, M.M., Zhao, D., Knapp, S., and James, M.N.G. (2003b). Crystal Structure of Human β-Hexosaminidase B: Understanding the Molecular Basis of Sandhoff and Tay–Sachs Disease. J. Mol. Biol. *327*, 1093–1109.

Niesen, F., Berglund, H., and Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. *2*, 2212–2221.

Nurizzo, D., Nagy, T., Gilbert, H.J., and Davies, G.J. (2002). The structural basis for catalysis and specificity of the Pseudomonas cellulosa ??-glucuronidase, GlcA67A. Structure *10*, 547–556.

Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. a, Lane, P., *et al*. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. J. Biomol. Screen. Off. J. Soc. Biomol. Screen. *6*, 429–440.

Pei, J., and Grishin, N. V. (2007). PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. Bioinformatics *23*, 802–808.

Pluvinage, B., Higgins, M. a., Abbott, D.W., Robb, C., Dalia, A.B., Deng, L., Weiser, J.N., Parsons, T.B., Fairbanks, A.J., Vocadlo, D.J., *et al*. (2011). Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. Structure *19*, 1603–1614.

Prag, G., Papanikolau, Y., Tavlas, G., Vorgias, C.E., Petratos, K., and Oppenheim, a B. (2000). Structures of chitobiase mutants complexed with the substrate Di-*N*-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. J. Mol. Biol. *300*, 611–617.

Ramasubbu, N., Thomas, L.M., Ragunath, C., and Kaplan, J.B. (2005). Structural Analysis of Dispersin B, a Biofilm-releasing Glycoside Hydrolase from the Periodontopathogen Actinobacillus actinomycetemcomitans. J. Mol. Biol. *349*, 475–486.

Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006). MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics *7*, 382.

Shane Rountree, J.S., Butters, T.D., Wormald, M.R., Boomkamp, S.D., Dwek, R.A., Asano, N., Ikeda, K., Evisnson, E.L., Nash, R.J., and Fleet, G.W.J. (2009). Design, Synthesis, and Biological Evaluation of Enantiomeric β-*N*-Acetylhexosaminidase Inhibitors LABNAc and DABNAc as Potential Agents against Tay-Sachs and Sandhoff Disease. ChemMedChem *4*, 378–392.

Slámová, K., Bojarová, P., Petrásková, L., and Křen, V. (2010). β-*N*-Acetylhexosaminidase: What's in a name…? Biotechnol. Adv. *28*, 682–693.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. Anal. Biochem. *150*, 76–85.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244–W248.

Studier, W. (2005). Protein production by autoinduction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sumida, T., Ishii, R., Yanagisawa, T., Yokoyama, S., and Ito, M. (2009). Molecular cloning and crystal structural

analysis of a novel beta-*N*-acetylhexosaminidase from Paenibacillus sp. TS12 capable of degrading glycosphingolipids. J. Mol. Biol. *392*, 87–99.

Sumida, T., Stubbs, K. a., Ito, M., and Yokoyama, S. (2012). Gaining insight into the inhibition of glycoside hydrolase family 20 exo-β-*N*-acetylhexosaminidases using a structural approach. Org. Biomol. Chem. *10*, 2607.

Tews, I., Vincentelli, R., Vorgias, C.E., Vorgias, C.C.E., and Wilson, K.S. (1996). N-Acetylglucosaminidase (chitobiase) from Serratia marcescens : gene sequence, and protein production and purification in Escherichia coli. Gene *170*, 63–67.

Thi, N.N., Offen, W. a, Shareck, F., Davies, G.J., and Doucet, N. (2014). Structure and activity of the Streptomyces coelicolor A3(2) beta-*N*-acetylhexosaminidase provides further insight into GH20 family catalysis and inhibition. Biochemistry *53*, 1789–1800.

Vivoli, M., Novak, H.R., Littlechild, J.A., and Harmer, N.J. (2014). Determination of protein-ligand interactions using differential scanning fluorimetry. J. Vis. Exp. 51809.

Vocadlo, D.J., and Davies, G.J. (2008). Mechanistic insights into glycosidase chemistry. Curr. Opin. Chem. Biol. *12*, 539–555.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T., and Yamamoto, K. (2008). Bifidobacterium bifidum Lacto-*N*-Biosidase, a Critical Enzyme for the Degradation of Human Milk Oligosaccharides with a Type 1 Structure. Appl. Environ. Microbiol. *74*, 3996–4004.

Wielgat, P., Walczuk, U., Szajda, S., Bień, M., Zimnoch, L., Mariak, Z., and Zwierz, K. (2006). Activity of lysosomal exoglycosidases in human gliomas. J. Neurooncol. *80*, 243–249.

Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G., and Withers, S.G. (2002). Aspartate 313 in the Streptomyces plicatusHexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. J. Biol. Chem. *277*, 40055–40065.

CHAPTER 5.
ENGINEERED LACTO-*N*-BIOSIDASE FOR LACTO-*N*-TETRAOSE SYNTHESIS

5.1 INTRODUCTION

There is a potential commercial use of HMO, specially, for supplementation of infant formulas, and to date, access to HMO, in a large-scale use, has not been possible by any synthetic methodology. By use of biocatalytic approaches, some simple HMO have been produced. However, the use of HMO as bioactive substances in infant food, pharmaceuticals or for clinical trials is still limited by the ability to synthesise more complex structures and, in sufficient quantity and purity (Bode, 2015; Kobata, 2003; Kunz *et al*., 2000).

In contrast to conventional chemical synthesis, enzymatic glycosylation is proposed to tackle the challenge of glycosynthesis of these complex molecules since offers many advantages. Enzymatic synthesis, without the need of protecting groups, is characterized by a high catalytic activity, lack of undesirable side-reactions, mild reaction conditions, and high regio- and stereoselectivity (Planas *et al*., 2015; Armstrong and Withers, 2013).

In this context, we aim to engineer *Bifidibacterium bifidum* Lacto-*N*-biosidase (LnbB) activity to be able to modulate the transglycosylation/hydrolysis ratio and synthesise our target Lacto-*N*-tetraose (type 1 HMO core). Two rational-design strategies have been considered. An enhanced transglycosylation strategy in which a new generation of mutants on the acceptor subsites of the enzyme will be able to favourably accommodate a sugar acceptor to compete with water, and then, increase transglycosylation activity. In addition, reaction will be engineer to favour the transglycosylation reaction, and *p*-NP-LNB will be used as substrate. And a glycosynthase-like strategy in which a new mutant on the catalytic residue of the enzyme will be able to exhibit transglycosylation activity without further hydrolysis of the product. Since LnbB, classified as a member of family 20 glycosyl hydrolases, uses a two-step catalytic mechanism involving substrate-assisted catalysis (Figure 5.1A). Thus, the glycosyl donor for the glycosynthase technology is an activated sugar oxazoline, which binds the enzyme to form a new glycosidic linkage by transglycosylation (Planas *et al*., 2015).

Therefore, we explore the feasibility of these new engineered LnbB proteins as biocatalysts for the glycosynthase strategy using oxa-LNB (Figure 5.1C) and lactose as substrates, or for the transglycosylation strategy using *p*-NP-LNB and lactose (Figure 5.1B) to achieve the target lacto-*N*-tetraose.

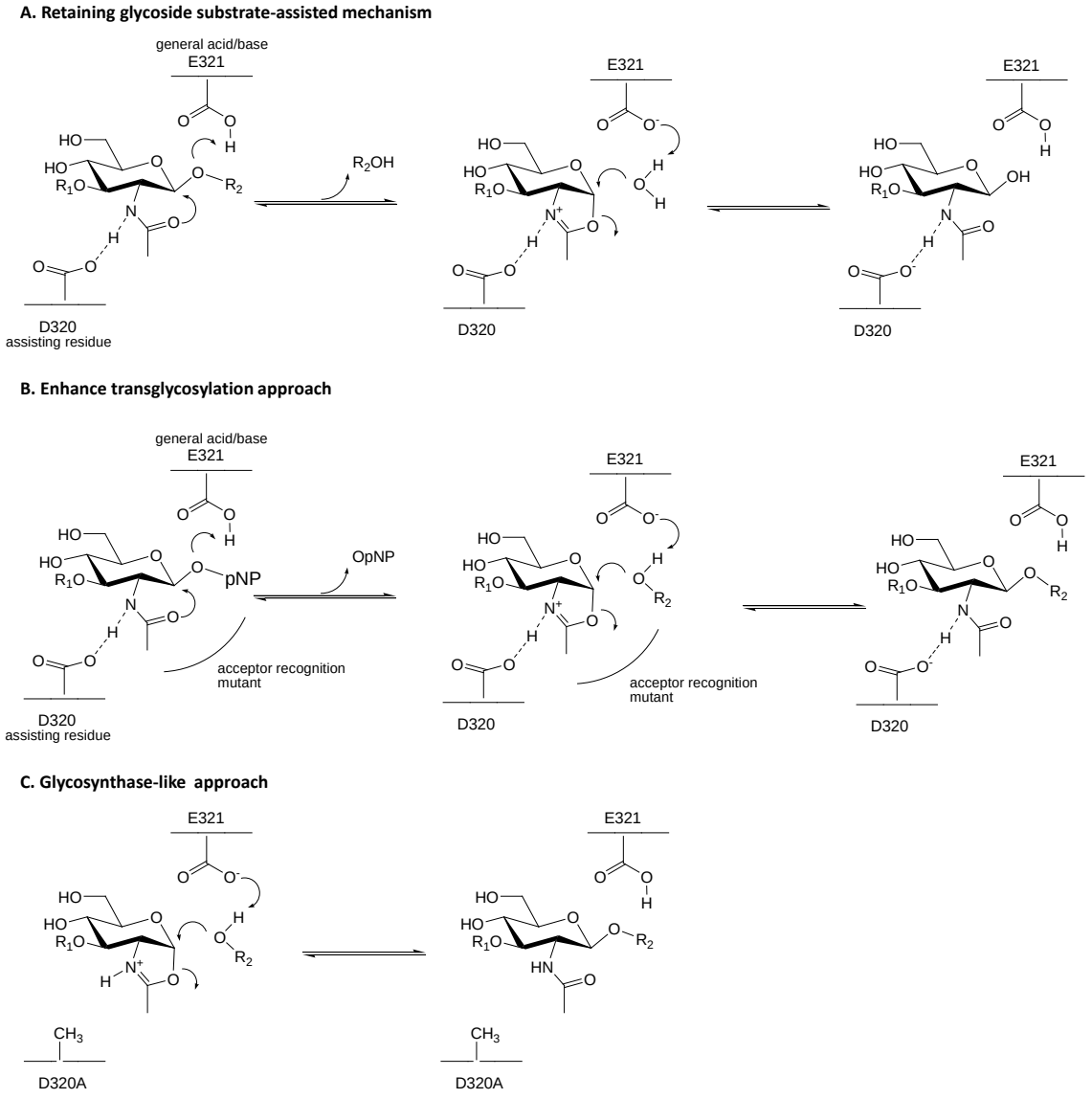


Figure 5.1 Lacto-*N*-biosidase mechanism *via* substrate-assisted mechanism.
(A) Wild type LnbB catalytic mechanism (B) enhanced transglycosidase approach using *p*-NP-LNB as donor (C) glycosynthase approach using oxa-LNB as donor (B). R_1: galactose, R_2: lactose.

5.2 MATERIALS AND METHODS

5.2.1 Bioinformatics

Protein sequences were retrieved from UniProt (http://www.uniprot.org). Sequence alignment was performed with PROMALS (Pei and Grishin, 2007) and ClustalW with default options.

5.2.2 General materials and experimental procedures

All reagents are from Sigma-Aldrich unless otherwise designated.

General material and methods are detailed in Chapter 3.

Specific substrates for both enzymatic approaches (*p*-NP-LNB and oxa-LNB) were synthesised according to Chapter 2 (see experimental section).

5.2.3 Site-directed mutagenesis of active site mutants of LnbB

Plasmid pET24b-LnbB-FL, which encodes for wild type full length LnbB with C-terminal His-tagged was used as template for mutagenesis. PCR reactions were carried out using IProof polymerase (2 units/µL, Bio-Rad) and a thermocycler PTC-200 (MJ research). Amplification was carried out in a 20-50 µl final volume reaction mixture containing plasmidic DNA (~150 ng), 0.02- 0.05 µM of sense primer, 0.02-0.05 µM of antisense primer, nuclease-free water and I-proof PCR master mix (Ref. 172-5310, Bio-Rad). A 3 min initial denaturation at 95 °C was followed by 30 cycles: 95 °C 30 s, 30 s of the corresponding annealing temperature for the different primers, 72 °C 4 min, and a 10 min final extension at 72°C. After PCR reaction, the samples were digested with DpnI for 1 hour at 37 °C, and then, transformed into *E. coli* DH5α. Positive transformants were confirmed by DNA sequencing and the final constructs were used to transform *E. coli* DH5α and BL21 (DE3) cells.
Following are listed primers used are list below for D320E and Y419F mutant proteins (mutated nucleotides are underlined, and codons affected are lowercase letter):

D320E sense	5’- TGGGCGCGgaaGAATACATGATTGGTACC-3’
D320E antisense	5’-CATGTATTCttcCGCGCCCATGTGCCAATATTTCG-3’
Y419F sense	5’-AGGCTCTGttcTGGTCACGTTCGGCACAAG-3’
Y419F antisense	5’-ACGTGACCAgaaCAGAGCCTGCGTTGCATTCATCAGG-3’

5.2.4 Expression and purification of enzymes

Proteins were expressed and purified as described in Chapter 3. Protein concentration was determined by Bradford (Bradford, 1976)using BSA as standard.

5.2.5 Kinetics of the hydrolase activity

Hydrolytic activity of soluble enzymes was performed as described in Chapter 3. Briefly, activity assays were performed in a reaction mixture containing 0.25 mM substrate in citrate-phosphate buffer at pH 4.5. After pre-incubation for 5 min at 30 °C, reactions were initiated by addition of the enzyme (20 nM). At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release

was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed.
Initial rates (v_o) were obtained from the linear progress curve of release of *p*-NP, and they were expressed as $v_o/[E]$ in s^{-1}.

5.2.6 Stopping method of enzymatic reactions

The following stopping method was implemented for HPLC analysis. After performing hydrolytic reaction at standard conditions, reactions were stopped with different volumes of 50% 1-propanol in water at different time intervals, 30 and 60 min. Different dilutions were studied: 1:2, 1:4, 1:6 and 1:8 ratio (reaction:1-propanol (v/v)). At 30 min one of the stopped duplicates was quenched with 100µL of glycine buffer pH 11. At 60 min the rest of the reactions were quenched with 100µL of glycine buffer pH 11. Release of *p*-nitrophenol was measured at 400 nm with a microplate-reader MRX (Cultek). The molar extinction coefficients of *p*-nitrophenol at 400 nm were also determined with the addition of 50% 1-propanol in water.

5.2.7 Glycosynthase reaction

Glycosynthase reaction conditions involved 1 mM of oxa-LNB (donor), 2 mM of *p*-NP-lactose and 1 µM of protein in a final volume of 1 mL. The reactions took place in 50mM of citrate-phosphate buffer at 30 °C. At different times (0-10 min every minute, and then, at 1, 2, 3, 6 and 24 hours), reaction samples of 20 µL were stopped by adding 60 µL volumes of 1-propanol (stored at 4°C), and then, were injected for HPLC-UV analysis.

Analysis of the glycosynthase reactions was performed using HPLC-UV (Agilent-1200 with a Nova-Pak® C18 column (Waters), 2 µL injection, and isocratic elution of 14% methanol in water at 25 °C. The detection of the *p*-NP-glycosids was done by UV at 300 nm. The validation of the method was done using *p*-NP-GlcNAc as standard. The limit of detection (LOD) and the limit of quantification (LOQ) for them were 1.5 µM and 5 µM respectively, and coefficient of variation was 8%.

5.2.8 DMSO tolerance analysis

Lacto-*N*-biosidase hydrolytic activity was determined with *p*-NP-LNB as substrate in the presence of DMSO. The DMSO tolerance was analysed from 0 to 15 % (v/v). Reactions containing 0.25 mM substrate in citrate-phosphate buffer at pH 4.5, and the corresponding amount of DMSO were pre-incubated for 5 min at 30 °C. Then, reactions were initiated by addition of the enzyme (20 nM). At regular time intervals, reactions were quenched by addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 11) and *p*-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed.

5.2.9 Transglycosylation reaction

Reaction conditions involved 2.5 mM of *p*-NP-LNB (donor), 0.25 M of lactose and 500 nM of protein in a final volume of 300 µL. The reactions took place in 50mM of citrate-phosphate buffer, pH 4.5 (Buffer C) with 15% of DMSO at 30 °C. At different times (0-16 h), reaction

samples of 20 µL were stopped by adding 60 µL volumes of 50% 1-propanol in water (stored at 4°C). LNT standard dilutions (0.005-0.5 mM) were injected in each analysis.

The analysis was performed using HPLC-MS (Agilent Infinity 1260 HPLC) with a XBridge UPLC BEH Amide 1.8 µm 2.1x100 mm Column (Waters). Chromatographic conditions involved 1 µL sample injection, gradient elution of acetonitrile/water with 0.1% formic acid at 50 °C and at a flow rate of 0.3 ml/min.

Table 5.1 Gradient elution conditions for HPLC.MS analysis

Time (min)	Acetonitrile (%)	H_2O (%)
0	10	90
5	30	70
14	30	70
15	10	90
21	10	90

MS conditions consist of using a SQ 6130 MS system (Agilent) in positive ionization (electrospray ionization (ESI+), single quadrupole MS detector, drying gas flow: 11 L/min, drying gas temperature: 325 °C, nebulizer pressure: 30 psi, capillary voltage: 4000 V, SCAN range (m/z): 100 – 900 (40%), fragmentor: 100 V.

Table 5.2 Reference masses used to follow the transglycosylation reaction.

Channel	Compound	SIM ion
2	Lacto-*N*-biose	384 $[M+H]^+$; 406 $[M+Na]^+$
3	*p*-NP-Lacto-*N*-biose	360 $[M+NH_4]^+$, 365 $[M+Na]^+$
	Lactose	522 $[M+NH_4]^+$, 527 $[M+Na]^+$
4	Lacto-*N*-tetraose	708 $[M+H]^+$; 730 $[M+Na]^+$

The validation of the method was done using Lacto-*N*-tetraose as standard. The limit of detection (LOD) and the limit of quantification (LOQ) were 2.5 µM and 5 µM respectively, and coefficient of variation was less than 15 %.

5.3 RESULTS AND DISCUSSION

With the aim of engineering *Bifidibacterium bifidum* Lacto-*N*-biosidase (LnbB) to modulate transglycosylation/hydrolysis ratio for the synthesis of Lacto-*N*-tetraose (type 1 HMO core), two rational-design strategies were evaluated.

5.3.1 Glycosynthase-like strategy

In the substrate-assisted mechanism, a conserved Asn or Asp residue play a role in orientating the 2-acetamido group and ultimately promoting the nucleophilic attack of the 2-acetamido group at the anomeric position to form the oxazolinium ion intermediate in the hydrolase mechanism. The first glycosynthase from retaining GHs acting via substrate-assisted catalysis was the endo-β-*N*-acetylhexosaminidase from *Mucor hiemalis* (Endo-M) (Umekawa *et al*., 2008). It belongs to family GH85 enzymes which are endo-β-*N*-acetylhexosaminidases that cleave the chitobiose core (GlcNAcβ-1,4GlcNAc) of *N*-linked glycans. Asn175 residue of Endo-M assists the appropriate orientation of the 2-acetamido group of the substrate for catalysis. Its mutation to alanine abolishes the hydrolytic activity, but is able to take the sugar oxazoline as donor substrate for transglycosylation, acting like a glycosynthase. The methodology has been extended to other GH85 enzymes (Endo-A, Endo-D) for the chemoenzymatic synthesis of N-glycoproteins (Wang and Huang, 2009; Yamamoto, 2013). In addition, family GH18 enzymes, which comprise chitinases and endo-β-*N*-acetylglucosaminidases, also operate by substrate-assisted catalysis and were also successfully converted to glycosynthases (Huang *et al*., 2012; Martinez *et al*., 2012).

Since Asp320 of LnbB was identified as the assisting residue and D320A mutated protein was 3100 times less hydrolytically active than the wild type (see Chapter 3), D320A mutant is proposed as the glycosynthase-like LnbB biocatalyst.

Glycosynthase reaction of D320A was studied using Lacto-*N*-biose oxazoline derivative (oxa-LNB) as donor and *p*-NP-lactose as acceptor to yield *p*-NP-Lacto-*N*-tetraose (Figure 5.2).

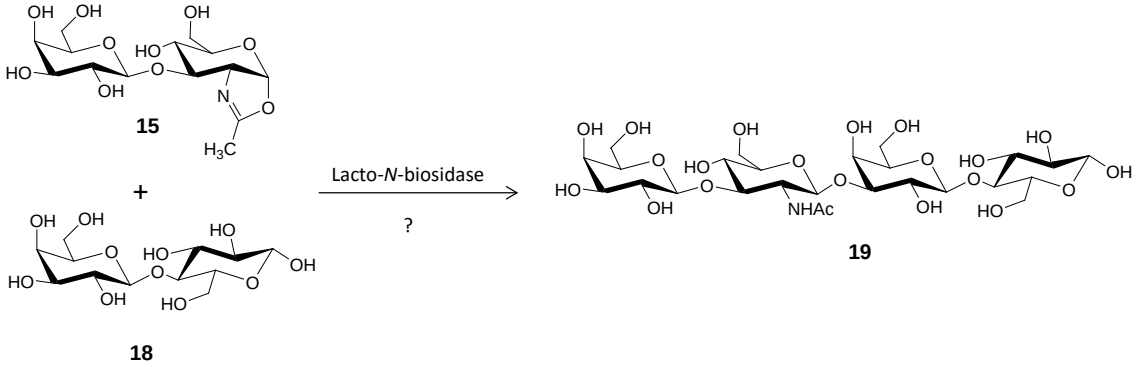


Figure 5.2 Glycosynthase reaction to produce Lacto-*N*-tetraose using Lacto-*N*-biosidase mutant protein.

Although pH 4.5 corresponds to the maximum hydrolytic activity of the LnbB wild-type, glycosynthase reactions will be explored at pH higher than 6 to guarantee the stability of oxazoline derivative donor (see Chapter 2). Similarly to D320A, the equivalent mutant of SpHex (D313A) also had a reduced hydrolytic activity and interestingly, the optimum pH for hydrolytic

activity was shifted from pH 4.5 to pH 7.0 (Williams *et al*., 2002). Consequently, LnbB glycosynthase experiments were performed under different pH conditions: pHs 6, 7 and 8.

Final conditions involved 1 mM of LNB-oxa (donor), 2 mM of lactose and 1 µM of protein. The reactions took place in 50mM of citrate-phosphate buffer at different pH and at 30 °C. At different times, reactions were stopped by adding 4-fold volumes of 1-propanol (50% in water), and then, were analysed using HPLC-UV.

The use of 1-propanol to stop the enzymatic reaction instead of 1.5 M glycine buffer at pH 11 was due to the hydrolysis of the LNT product when it was incubated at basic conditions. Moreover, glycine buffer was not suitable for further analysis using MS equipment. Different stopping methods were tried, and the final conditions involved adding 4 volumes of 1-propanol to the reaction aliquots prior to HPLC-MS analysis.

Unfortunately, at the conditions tested, no synthetic activity was detected. LNT was not formed in any of the reactions. The fact that the hydrolytic activity was not completely abolished by the mutation of the assisting residue suggested that hydrolysis of oxa-LNB donor could occur, and prevent LNT formation.

In retaining glycosynthases with two essential residues, the removal of the nucleophilic residue renders completely inactive enzymes. Thus, when performing glycosynthase reaction, the corresponding activated fluoride donor transfer to an acceptor is kinetically favoured over hydrolysis. The synthesised product is no longer hydrolysed because of the lack of the nucleophile, and it renders high transglycosylation yields (Planas *et al*., 2015). The initial concept of glycosynthase preparation based on mutated glycosidases at catalytic nucleophile and the use of glycosyl fluoride donor appears to be, in principle, a simple strategy. However, this has not led to a large variety of different enzymes, since, nowadays, the glycosynthases available are, so far, limited to 16 glycoside families out of the more than 100 glycosidase families described in Cazy database (Cantarel *et al*., 2009; Cobucci-Ponzano *et al*., 2008).

The additional fact that, to date, there is no glycosynthase acting enzyme from family GH20, suggests a more complex strategy. More research is needed to unravel whether LnbB can become a glycosynthase, and in that sense, synthesise the added-value oligosaccharide, lacto-*N*-tetraose.

5.3.2 Enhanced transglycosylation strategy

In order to explore transglycosylation activity of LnbB, Endo-M protein was also used as reference. Umekawa *et al*. generated two more interesting mutants with superior transglycosidase activity; N175Q and Y217F. The N175Q mutant was found to possess enhanced transglycosidase-like activity with “natural” N-glycan compared to the wild type (3-fold higher), and even higher glycosynthase-like activity with sugar oxazoline in comparison with N175A. Authors suggested that the extension of the side-chain from Asn to Gln by one carbon might hindered the hydrogen bonding between the amide group of the assisting residue (Asn or Gln) and the acetamide of the GlcNAc of the substrate. These results also implicated the significance of amide side-chain in the Asp175 of Endo-M for promoting oxazoline transglycosylation in the second step of the catalysis (Umekawa *et al*., 2010). Finally,

the authors mutated the residue Tyr217 to Phe. Tyr217 is located in the active-site, is a highly conserved residue among GH85, and was found to be essential for transglycosylation activity, but not for hydrolytic activity of the wild type Endo-M. The Y217F mutant catalysed the transglycosylation reaction more efficiently than the wild type Endo-M (5-fold higher). Besides, the removal of the hydroxyl group of Tyr residue resulted in a diminished hydrolysis activity. Moreover, replacement of Tyr217 to other amino acids than Phe, resulted in abolishment of both activities, hydrolysis and transglycosylation. These results suggested that the phenyl group at this position is important for, both, acceptor recognition and for the catalysis and may be involved in the activation of a water molecule for hydrolysis (Umekawa *et al*., 2008, 2010).

Accordingly, two mutated proteins of LnbB were generated: D320E and Y419F. Asp320 was substituted by Glu since this longer side-chain might allow more favourable interaction to facilitate oxazoline substrate processing similarly to Endo-M. Finally, Tyr419, which is highly conserved in GH20 enzymes, was changed to Phe to improve acceptor recognition. It is known that the role of the residues located in positives subsites are important for transglycosylation activity by contributing to higher binding free energy of the incoming acceptor (Bissaro *et al*., 2015).

Again, the genes of the corresponding mutants were successfully obtained by site-directed mutagenesis using wild type Full-length pET24b construct as template (C-terminal His-tagged). Mutations were confirmed by DNA sequencing and the final constructs were used to transform *E. coli* BL21 (DE3) cells for the expression and purification of mutant recombinant proteins (see Chapter 3 (material and methods section).

The protein concentration was determined by Bradford (Bradford, 1976) using BSA as standard. Yields of 3-7 mg of proteins per litre of culture were obtained. All proteins were considered to be >90% pure by Coomassie staining after SDS-PAGE (Figure 5.3, lane 4 and 5).

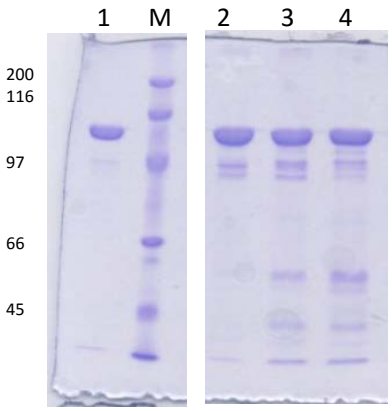


Figure 5.3 SDS-PAGE analysis of the wild type and mutant LnbB proteins.
Lane 1: full-length protein, M: marker, Lane 2: D320A protein, Lane 3: D320E protein, Lane 4: Y419F protein. Molecular masses in kDa.

The hydrolytic activity of the mutated proteins was evaluated. In addition, a new batch of wild type LnbB was obtained and analysed as the control. Briefly, the hydrolytic activity of the soluble proteins was performed in a reaction mixture containing *p*-NP-LNB as substrate and

citrate-phosphate buffer at pH 4.5. The reaction was stopped at different interval times and the release of *p*-NP was measured (see Chapter 3, materials and methods).

Table 5.3 Kinetic parameters of the hydrolytic reaction of wild type and mutant LnbB proteins. The reactions were done at 20 nM of proteins, 0.25 mM of substrate, 30 °C in buffer C at pH 4.5.

Protein	Specific activity (s^{-1})
Wild type	31
D320E	0.24
Y419F	0.23

Both mutated proteins showed similar decreased hydrolytic activity (130-fold decrease) compared to the wild type enzyme (Table 5.3). In the case of the D320E mutant, the introduction of a longer side-chain seemed to disturb the interaction with the 2-acetamido group of the substrate, although the mutated enzyme showed residual activity. Similarly, in Y419F mutant, the removal of the hydroxyl group showed a worsening of the hydrolytic capacity, possibly due to a loss of an hydrogen bond with the substrate. As both mutants were designed to evaluate its possible enhanced transglycosylation activity compared to the wild type LnbB, the decreased but not abolished hydrolytic activity, postulated them as good candidates.

Condensation reactions between *p*-NP-LNB (donor) and lactose (acceptor) were performed in order to study the transglycosylation activity of D320E and Y419 mutants, as well as the wild type LnbB. In addition to LNT formation, other side reactions such as the hydrolysis of the donor *p*-NP-LNB to yield LNB and *p*-NP, and the hydrolysis of the product LNT to LNB and lactose could occur (Figure 5.4).

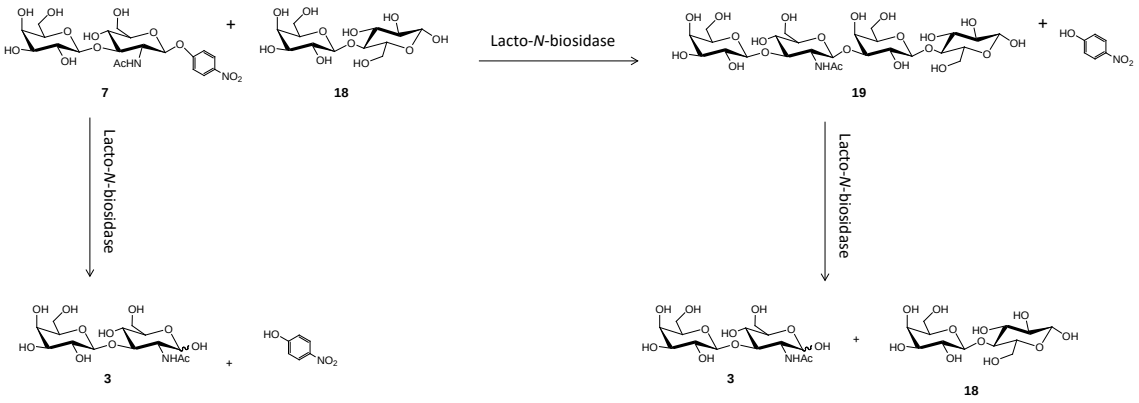


Figure 5.4 Trasnglycosylation reaction of LnbB to synthesise Lacto-*N*-tetraose (19).

Transglycosylation reaction catalysed by LnbB was followed using HPLC-MS in order to identify LNT ((m/z) 708 $[M+H]^+$ and 730 $[M+Na]^+$), LNB (384 $[M+H]^+$ and 406 $[M+Na]^+$), *p*-NP-LNB (360 $[M+NH_4]^+$ and 365 $[M+Na]^+$) and lactose (522 $[M+NH_4]^+$ and 527 $[M+Na]^+$). The procedure involved an Acquity UPLC BEH Amide Column and a gradient with acetonitrile/water with 0.1% formic acid. The mass spectrometer was used in positive ionization (electrospray ionization (ESI+) with a single quadrupole MS detector).

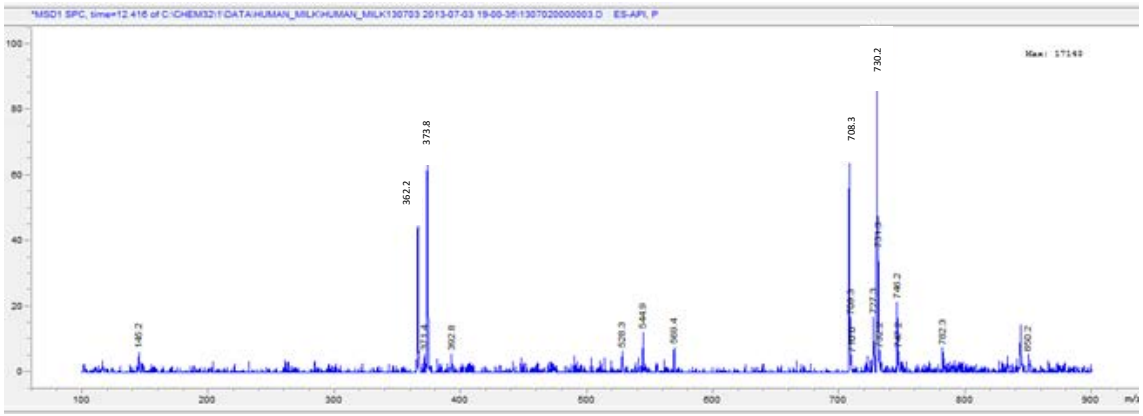


Figure 5.5 MS spectrum of LNT.

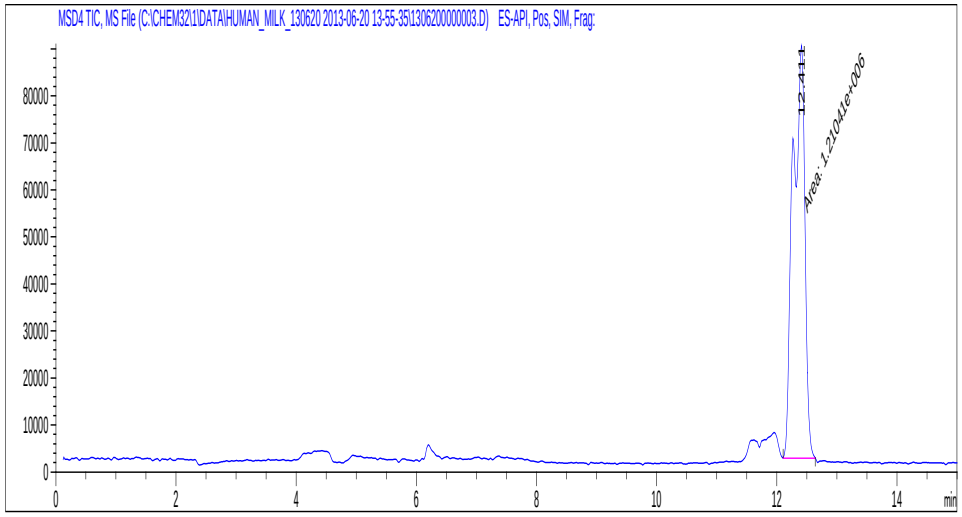


Figure 5.6 Chromatogram of LNT.

The method was validated with LNT. It was determined that the limit of detection (LOD) and the limit of quantification (LOQ) were 2.5 µM and 5 µM respectively, and coefficient of variation was less than 15 %. Figure 5.5 shows the mass spectrum and the HPLC-MS chromatogram of LNT. As is observed, LNT elutes at 12.7 min. LNT standard was used in each reaction and the corresponding response factor was applied of $3.7 \cdot 10^{6} \pm 5.8 \cdot 10^{4}$ mM^{-1} with a relative error of 1.6%).

5.3.2.1 Synthesis of LNT by transglycosylation reaction

With the aim of enhancing the transglycosylation/hydrolysis ratio, an excess of acceptor (compared to donor) was used to minimise the presence of water molecules within the active centre of the protein. In addition, an organic solvent was added (Dimethylsulfoxide, DMSO) to reduce water activity.

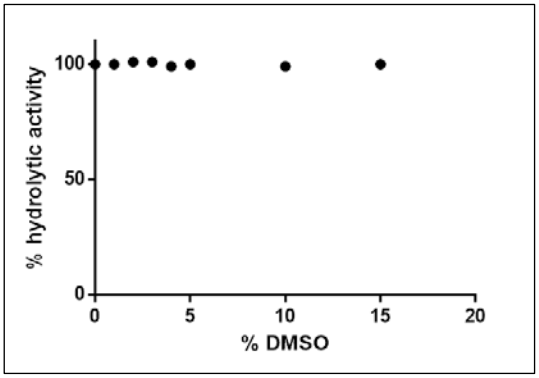


Figure 5.7 DMSO tolerance of LnbB.
Reaction mixtures containing 20 nM of LnbB, 0.25 mM of p-NP-LNB, 0-15 % DMSO and 50mM of citrate-phosphate buffer pH 4.5 (Buffer C) in a final volume of 100 µL, were incubated at 30 °C. At different interval times, the different reactions were stopped by adding 150 µL of Glycine buffer pH 11, and released *p*-nitrophenol was measured at 400 nm in a plate reader spectrophotometer.

Since, in some cases, the use of hydro-organic solvents in enzymatic reactions faces with the instability of enzymes in those organic solvents, the DMSO tolerance of LnbB was analysed in the hydrolytic activity of wild type enzyme with *p*-NP-LNB as substrate. As observed in Figure 5.7, the hydrolytic activity of LnbB was not affected by a range of 0-15 % of DMSO, therefore, 15 % DMSO was used for transglycosylation reaction.

Transglycosylation reaction catalysed by the wild type LnbB and D320E and Y419F mutants was analysed. Initial conditions were as follows: *p*-NP-LNB and lactose at donor:acceptor ratio of 1:5 or 1:10, 200 nM of protein, in 50mM of citrate-phosphate buffer, citrate-phosphate buffer at pH 4.5 with 15% of DMSO at 30 °C. Since very low yields of LNT were obtained using wild type LnbB at these conditions, new reactions were done with a molar excess of 100-fold acceptor compared to donor and also with higher amount of protein. Final reactions were performed at 2.5 mM of *p*-NP-LNB (donor), 0.25 M of lactose and 500 nM of protein. At different times, reactions were stopped by adding 4-fold volumes of 1-propanol (50% in water), and then, were injected into the amide column for HPLC-MS analysis.

The wild type LnbB showed transglycosylation activity, however, the LNT synthesis were too low to quantify it and in less than 10 min it quickly disappeared by hydrolysis. Similar results were reported by Wada and coworkers (Wada et al., 2008).

LNT formation by transglycosylation occurred when Y419F mutant was used as biocatalyst in the same conditions. Figure 5.8 shows that peak at 12.7 min corresponding to LNT evolved during the course of the reaction. From 60 min to 90 min of reaction, LNT increased to a maximum of 0.24 mM (0.9% yield), and from then, it decreased to its completely hydrolysis (Figure 5.9A).

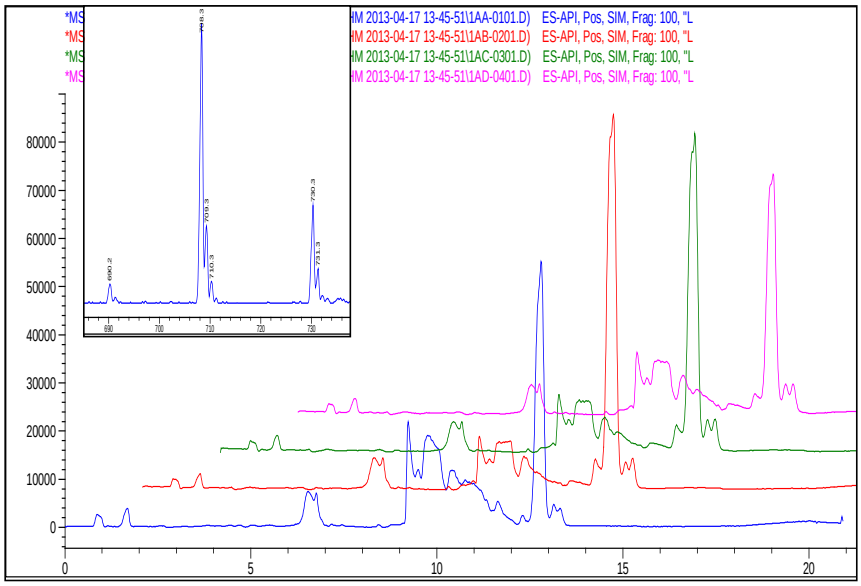


Figure 5.8 Progression of the transglycosylation reaction of Y419F mutant followed using HPLC-MS.
Transglycosylation at different times of reaction: Blue: 60 min, Red: 90, Green: 120 min, and Pink: 180 min.

D320E mutant showed higher transglycosylation activity. The progress curve of the reaction was very similar to Y419F, but, in this case, LNT was synthesised in higher yields. As observed in Figure 5.9B, at time 0 min, LNT was already formed. The maximum production was at 90 min with 3.3% yield. From then, LNT concentration was decreased till 16 h that LNT was completely hydrolysed.

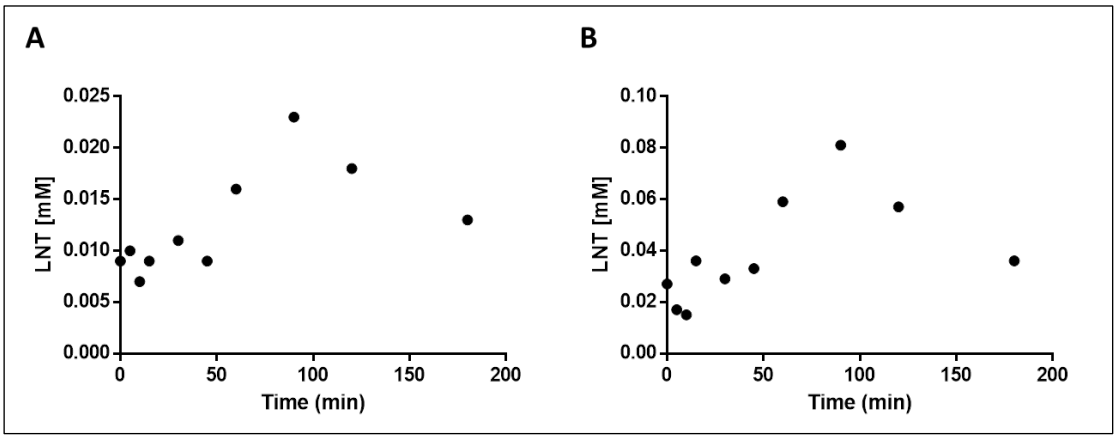


Figure 5.9 LNT synthesis by transglycosylation of (A) Y419F protein and (B) D320E protein.
Values were calculated using a response factor of $3.7 \cdot 10^{6} \pm 5.8 \cdot 10^{4}$ mM^{-1} with a relative error of 1.6%.

The analysis of *p*-NP-LNB revealed that it was under its quantification limit in all reactions, which indicates that the hydrolysis of the *p*-NP-LNB substrate was the main reaction accomplished when transglycosylation was performed.

These results, although being unpretentious, position LnbB as an interesting protein for further engineering studies in order to synthesise LNT. And, a second generation of mutants are proposed for future research.

Recently, Bissaro *et al*. studied the molecular basis of transglycosylation in engineering retaining glycosidases, mainly exo-acting enzymes. They observed that mutations of conserved negative subsite residues (donor subsites) produce retention time prolongation of the donor glycosyl moiety, and thus, enhance the transglycosylation activity (Bissaro *et al*., 2015).

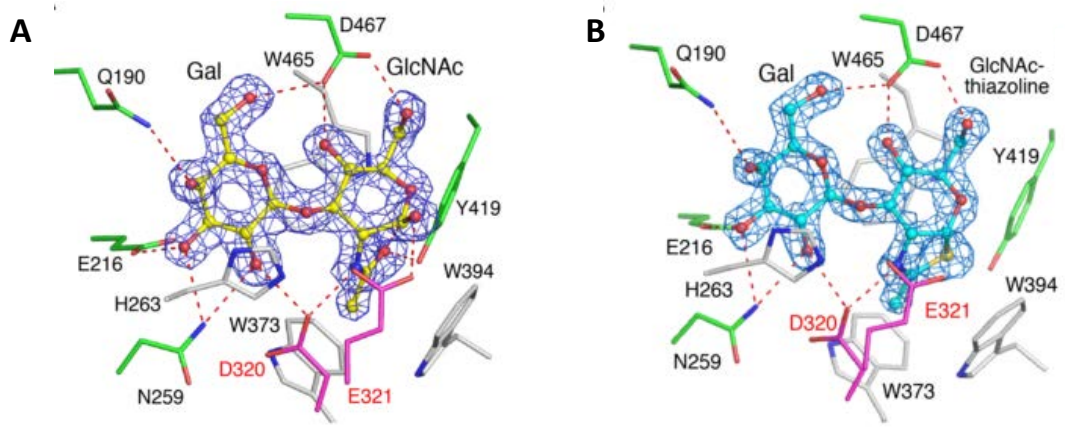


Figure 5.10 Stereoviews of LNB (A) and LNB-thiazoline (B) bound to the active site of LnbB.
Catalytic residues, residues forming hydrogen bonds, and residues forming hydrophobic interactions are shown in magenta, green, and white, respectively. Hydrogen bonds are shown as red dashed lines. Images from Ito et a., 2013. PDB entry (A) 4H04 (B) 4JAW.

We have identified some conserved residues among GH20 proteins with the aim of favouring transglycosylation over hydrolysis ratio of LnbB. His263 which establishes hydrophobic interactions with the galactose moiety of LNB (Figure 5.10). Trp373, Trp394 and Trp465 also form hydrophobic interactions with both, Gal and GlcNAc, of LNB. Asp467 which is also hydrogen bonding with the O4- and O6-hydroxyl groups of the GlcNAc residue. Gln190 (which corresponds to Arg in others GH20 enzymes) forms hydrogen bonds with the O4- hydroxyl group of Gal and defines the characteristic -2 subsite of LnbB. And finally, Asn259, which is hydrogen binding the galactose unit of LNB according to the crystal structure (Ito *et al*., 2013).

Then, His263, Trp373, Trp394, Trp465, Gln190, Asn259 and Asp467 residues are proposed for future work as a rational starting point for enzyme engineering to enhance transglycosylation activity of LnbB. These residues are interacting with the donor substrate and their mutations could result in the modulation of transglycosylation/hydrolysis ratio of LnbB.

5.4 CONCLUSIONS

The first generation of mutants for the synthesis of LNT were designed: D320A, D320E and Y419F. These mutants showed decreased hydrolytic activity compared to the wild type enzyme, but in any case, the activity was abolished.

The glycosynthase reaction with D320A mutant in the presence of the oxazoline derivative lacto-*N*-bioside and lactose did not occur. Mutation of the assisting residue is not enough to redesign the catalytic activity of LnbB as a glycosynthase-like biocatalyst.

By contrast, the transglycosylation reactions between of *p*-NP-lacto-*N*-bioside and lactose using D320E and Y419F mutants showed an increased transglycosylation yield compared to the wild type LnbB. However, future protein engineering is proposed for both glycosynthase and transglycosylation approaches to be able to enhance the transglycosylation/hydrolysis ratio.

5.5 REFERENCES

Armstrong, Z., and Withers, S.G. (2013). Synthesis of glycans and glycopolymers through engineered enzymes. Biopolymers *99*, 666–674.

Bissaro, B., Monsan, P., Fauré, R., and O'Donohue, M.J. (2015). Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. Biochem. J. *467*, 17–35.

Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 1–42.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. *37*, D233–D238.

Cobucci-Ponzano, B., Conte, F., Mazzone, M., Bedini, E., Corsaro, M.M., Rossi, M., and Moracci, M. (2008). Design of new reaction conditions for characterization of a mutant thermophilic α - l -fucosidase. Biocatal. Biotransformation *26*, 18–24.

Huang, W., Giddens, J., Toonstra, C., and Wang, L.-X. (2012). Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions. J Am Chem Soc. *134*, 12308–12318.

Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., Ashida, H., Wakagi, T., Yamamoto, K., Stubbs, K.A., et al. (2013). Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-N-biosidase from Bifidobacterium bifidum. J. Biol. Chem. *288*, 11795–11806.

Kobata, A. (2003). Possible application of milk oligosaccharides for drug development. Chang Gung Med J *26*, 620–636.

Kunz, C., Rudloff, S., Baie, W., Klein, N., and St, S. (2000). O LIGOSACCHARIDES IN HUMAN MILK : Structural , Functional , and Metabolic Aspects. Annu. Rev. Nutr. 699–722.

Martinez, E.A., Boer, H., Koivula, A., Samain, E., Driguez, H., Armand, S., and Cottaz, S. (2012). Engineering chitinases for the synthesis of chitin oligosaccharides: Catalytic amino acid mutations convert the GH-18 family glycoside hydrolases into transglycosylases. J. Mol. Catal. B Enzym. *74*, 89–96.

Pei, J., and Grishin, N. V. (2007). PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. Bioinformatics *23*, 802–808.

Planas, A., Faijes, M., and Codera, V. (2015). When Enzymes Do It Better: Enzymatic Glycosylation Methods. In Carbohydrate chemistry:State of the Art and Challenges for Drug Development, p. Chapter 9.

Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L.X., and Yamamoto, K. (2008). Mutants of Mucor hiemalis endo-B-N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. J. Biol. Chem. *283*, 4469–4479.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T., and Yamamoto, K. (2008). Bifidobacterium bifidum Lacto-N-Biosidase, a Critical Enzyme for the Degradation of Human Milk Oligosaccharides with a Type 1 Structure. Appl. Environ. Microbiol. *74*, 3996–4004.

Wang, L.X., and Huang, W. (2009). Enzymatic transglycosylation for glycoconjugate synthesis. Curr. Opin. Chem. Biol. *13*, 592–600.

Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G., and Withers, S.G. (2002). Aspartate 313 in the Streptomyces plicatusHexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. J. Biol. Chem. *277*, 40055–40065.

Yamamoto, K. (2013). Recent advances in glycotechnology for glycoconjugate synthesis using microbial endoglycosidases. Biotechnol. Lett. *35*, 1733–1743.

CONCLUSIONS

CONCLUSIONS

1. Lacto-*N*-biose (LNB), synthesised using *Bacillus circulans* β-Gal-3 enzyme, was used as starting compound for the synthesis of the substrates of *Bifidobacterium bifidum* Lacto-*N*-biosidase (LnbB). A three steps method was performed to synthesise *p*-NP-LNB. It involved the peracetylation and chlorination of the anomeric carbon of LNB, glycosylation with *p*-NP, and final deacetylation. The synthesis of oxa-LNB was achieved via oxazoline ring formation using CDMBI. The study of different methodologies, which are extensively used in carbohydrate synthesis, suggested a high lability of the β-1,3 linkage when an acetamide group is located in position C-2.

2. The kinetic characterization of the recombinant *B. bifidum* LnbB showed optimum catalytic efficiency at pH 4.5, and hydrolytic activity followed saturation kinetics. By sequence alignment analysis of GH20 enzymes, Asp320 and Glu321 of LnbB were studied, and their mutations supported their role as catalytic residues.

3. After the structural-functional analysis of GH20 enzymes, two models of organization were identified: model A with GH20b-GH20-α architecture, and model B with the single catalytic GH20 domain alone.

4. *B. bifidum* LnbB protein belongs to Model A domain organization. Truncated forms of LnbB showed that model A enzymes cannot be reduced to model B, since GH20b domain is a structural requirement in the N-terminus of GH20 to assure protein stability.

5. Model A architecture (GH20b-GH20-α) is not always the minimal functional unit. Important interactions in the substrate binding cavity of the catalytic GH20 domain are provided by a remote element in all structures, which must be preserved to ensure a proper definition of the active site. The remote element can directly be provided by a long loop 2 of the GH20 domain itself. In this case, GH20b-GH20-α can act as a minimal functional unit, and large multidomain enzymes can be engineered to reduced forms with a minimal model A architecture. In other circumstances, when loop 2 is short, proteins dimerize or present an accompanying domain in the C-terminus of GH20 that provide this remote element. In the latter cases, protein engineering of these enzymes must be designed maintaining the accompanying domains.

6. The remote element in *B. bifidum* LnbB is provided by the lectin domain, which is an accompanying domain in the C-terminus. The hydrolytic activity recovery of the inactive truncated form of LnbB in the presence of the lectin domain confirmed that the lectin domain assists in the definition of the active site and thus, is essential for catalysis.

7. The results of D320A mutant as glycosynthase suggested that mutation of the assisting residue is not enough to redesign the catalytic activity of *B. bifidum* LnbB to behave as glycosynthase-like biocatalyst. By contrast, the ability of D320E and Y419F mutants to

perform synthesis of LNT by transglycosylation provides a rational starting point in order to modulate the transglycosylation/hydrolysis activity ratio of in *B. bifidum* LnbB.

APPENDIX

APPENDIX

Appendix 1

Appendix 2

Appendix 3

Appendix 4

Appendix 5

APPENDIX 1

Table A1.1. Monosaccharide building blocks for HMO.
Glycan structures are depicted according to the recommendations of the Consortium of Functional Glycomics (http://www.functionalglycomics.org).

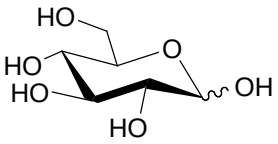
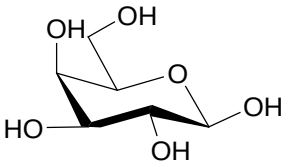
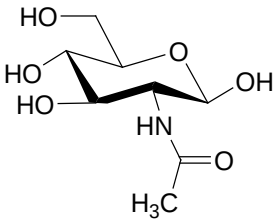
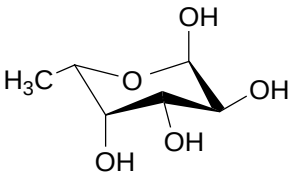
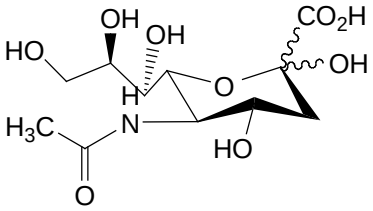
Monosaccharide	Abbreviation	Symbol	Structure
Glucose	Glc		
Galactose	Gal		
N-Acetylglucosamine	GlcNAc		
Fucose	Fuc		
N-Acetylneuraminic acid	Neu5Ac		

Table A1.2. Twelve glycosidic Linkages that constitute HMO.

Glycosidic linkages	**Structure**
Galactosidic bonds	Galβ1-4Glc
	Galβ1 (3GlcNAc)
	Galβ1-4GlcNAc
N-Acetyl-glucosaminidic bonds	GlcNAcβ1 (3Galβ1)
	GlcNAcβ1 (6βGal)
Fucosidic bonds	Fucα1 (Gal2)
	Fucα1 (3Glc)
	Fucα1 (3GlcNAc)
	Fucα1-4GlcNAc
Sialidic bonds	Neu5Acα2 (3Gal)
	Neu5Acα2 (6Gal)
	Neu5Acα2 (6GlcNAc)

Table A1.3. Core structures that occur in HMO.

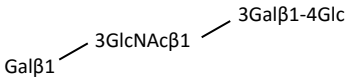
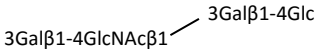
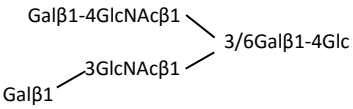
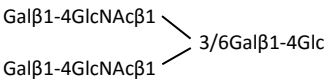
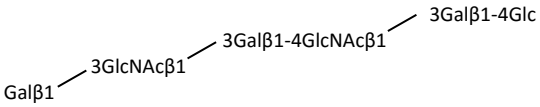
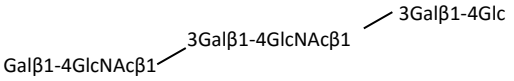
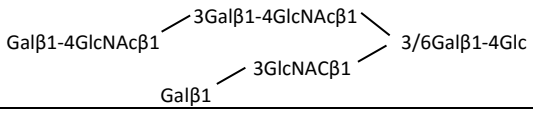
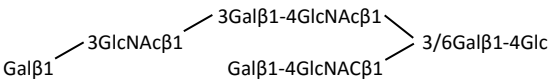
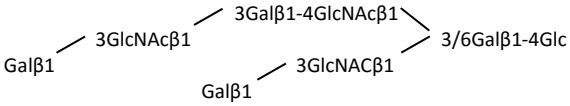
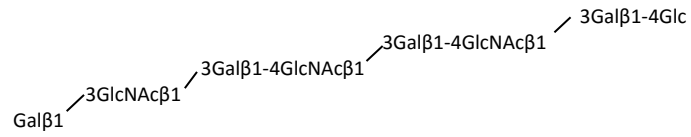
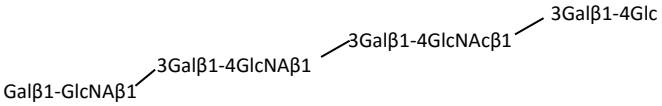
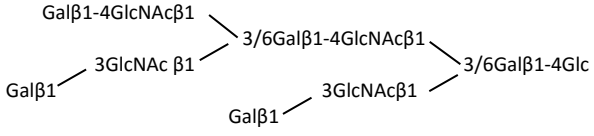
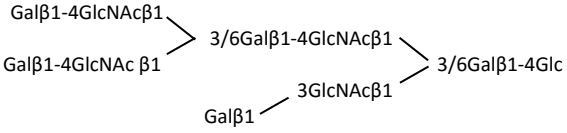
Oligosaccharide		Structure
Lacto-*N*-tetraose	LNT	
Lacto-*N*-neotetraose	LNnT	
Lacto-*N*-hexaose	LNH	
Lacto-*N*-neohexaose	LNnH	
para-Lacto-*N*-hexaose	*p*LNH	
para-Lacto-*N*-neohexaose	*p*LNnH	
Lacto-*N*-octaose	LNO	
Lacto-*N*-neooctaose	LNnO	
iso-Lacto-*N*-octaose	*i*LNO	
para-Lacto-*N*-octaose	*p*LNO	
para-Lacto-*N*-neooctaose	*p*LNnO	
Lacto-*N*-decaose	LND	
Lacto-*N*-neodecaose	LNnD	

Table A1.4. Structures of HMO that appear in this thesis (I).

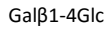
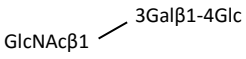
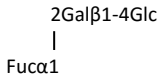
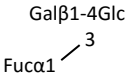
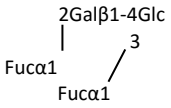
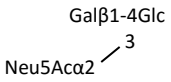
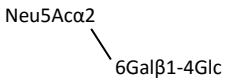
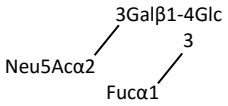
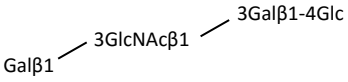
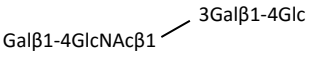
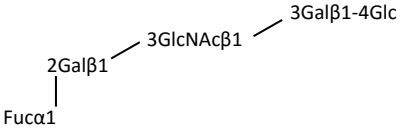
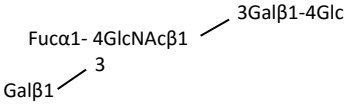
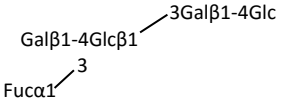
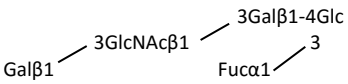
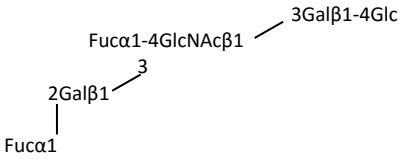
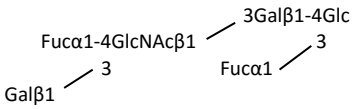
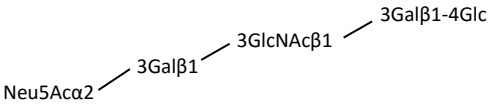
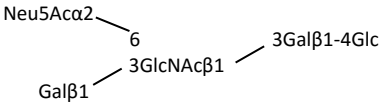
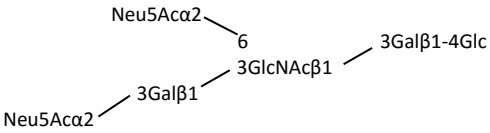
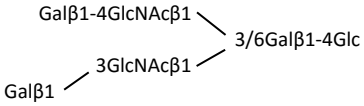
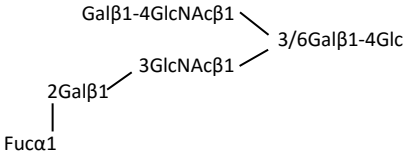
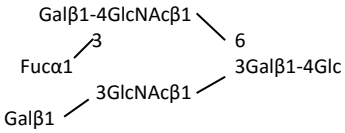
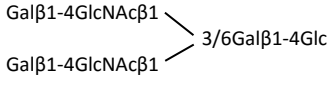
Oligosaccharide		Structure
Lactose	Lac	
Lacto-*N*-triose	LNT-II	
2'-Fucosyllactose	2'FL	
3'-Fucosylalactose	3'FL	
Lactodifucotetratose	LDFT	
3'-sialyllactose	3'SL	
6'-sialyllactose	6'SL	
3'-Sialyl-3-fucosyllactose	3'S3FL	
Lacto-*N*-tetraose	LNT	
Lacto-*N*-neotetraose	LNnT	
Lacto-*N*-fucopentaose I	LNFP I	
Lacto-*N*-fucopentaose II	LNFP II	
Lacto-*N*-fucopentaose III	LNFP III	
Lacto-*N*-fucopentaose V	LNFP V	

Table A1.5. Structures of HMO that appear in this thesis (II).

Oligosaccharide		Structure
Lacto-*N*-difuco-hexaose I	LNDFH I	
Lacto-*N*-difuco-hexaose II	LNDFH II	
Sialyllacto-*N*-tetraose a	LSTa	
Sialyllacto-*N*-tetraose b	LSTb	
Disialyllacto-*N*-tetraose	DSLNT	
Lacto-*N*-hexaose	LNH	
Fucosyllacto-*N*-hexaose I	FLNH I	
Fucosyllacto-*N*-hexaose II	FLNH II	
Lacto-*N*-neohexaose	LNnH	

APPENDIX 2

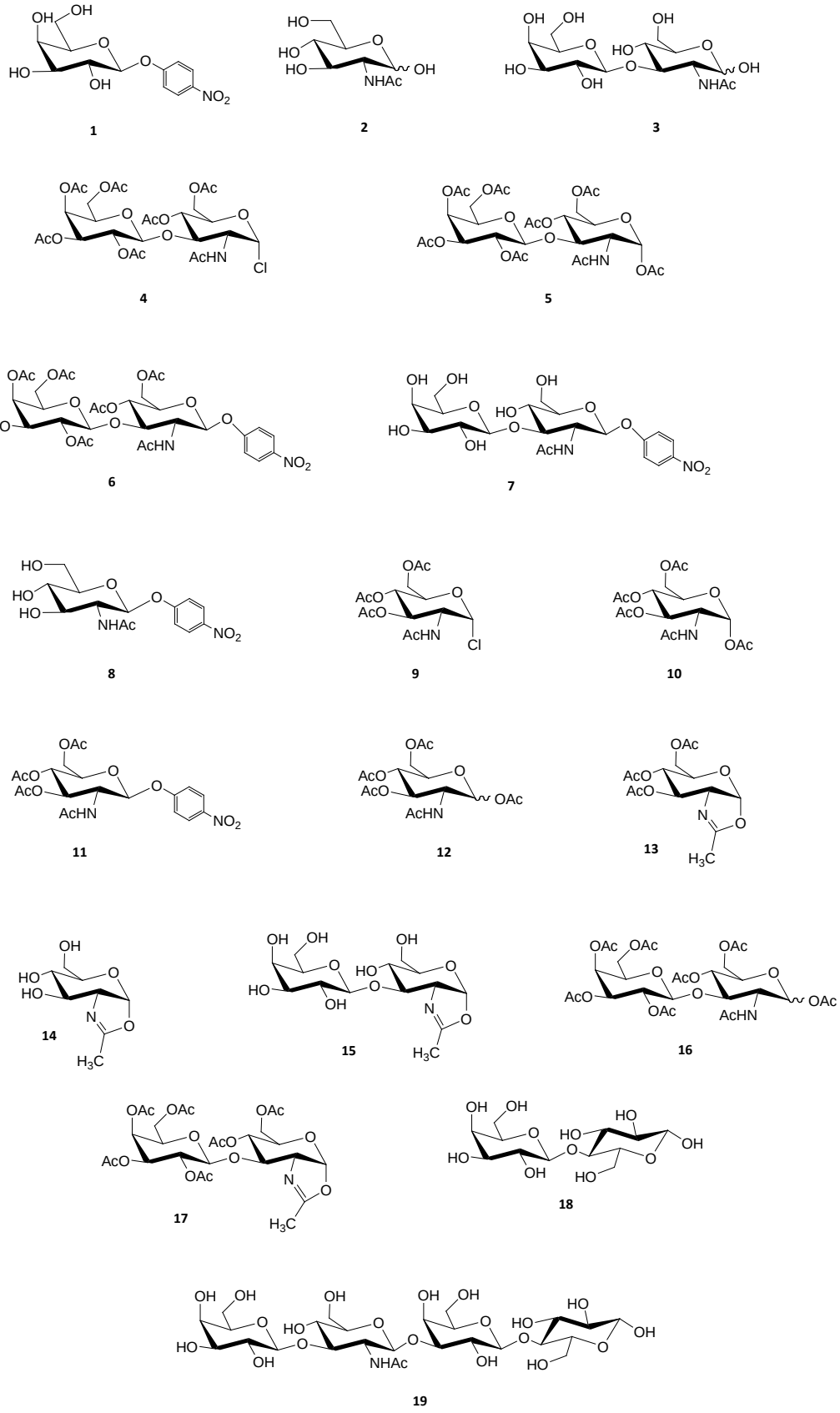


Figure A2.1. Chemical structures of the molecules appeared in Chapter 2.

Compound **3**

2-Acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (Lacto-*N*-biose)

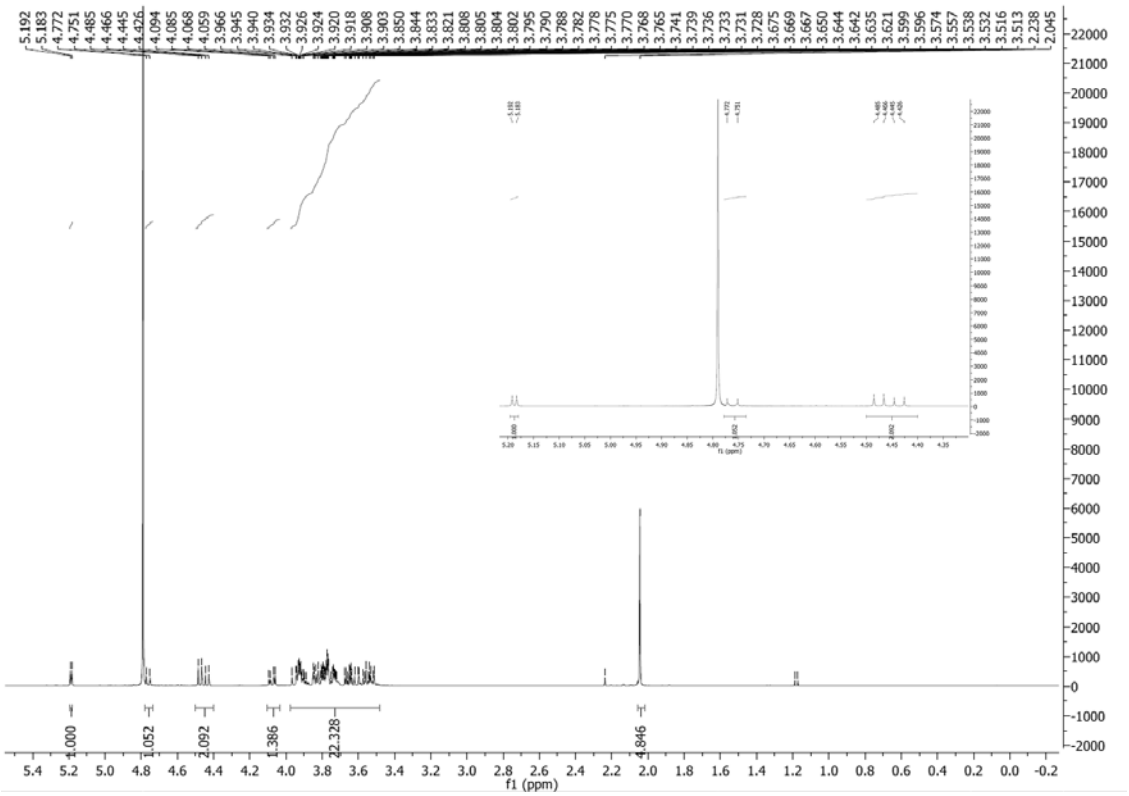


Figure A2.2. ^{1}H NMR of 2-Acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose.

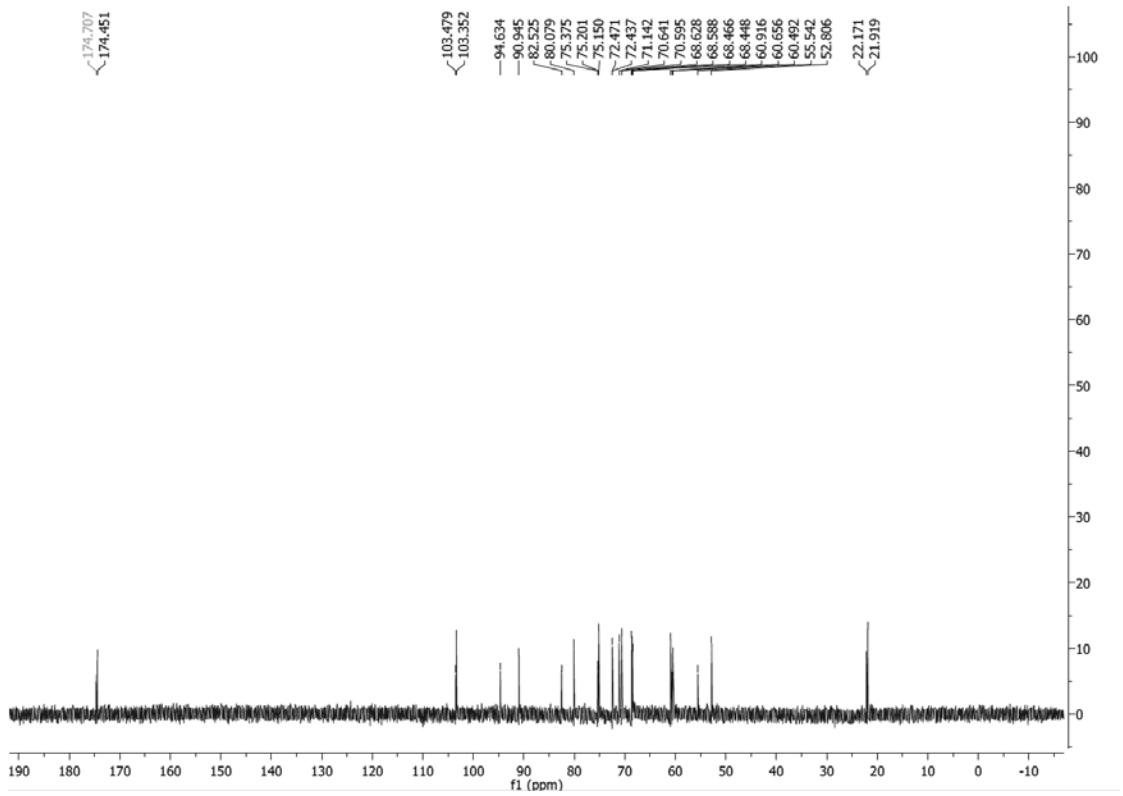


Figure A2.3. ^{13}C NMR of 2-Acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose.

Compound **16**

4-Nitrophenyl 2-acetamido-4,6-di-*O*-acetyl-2-deoxy-3-*O*-(2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside

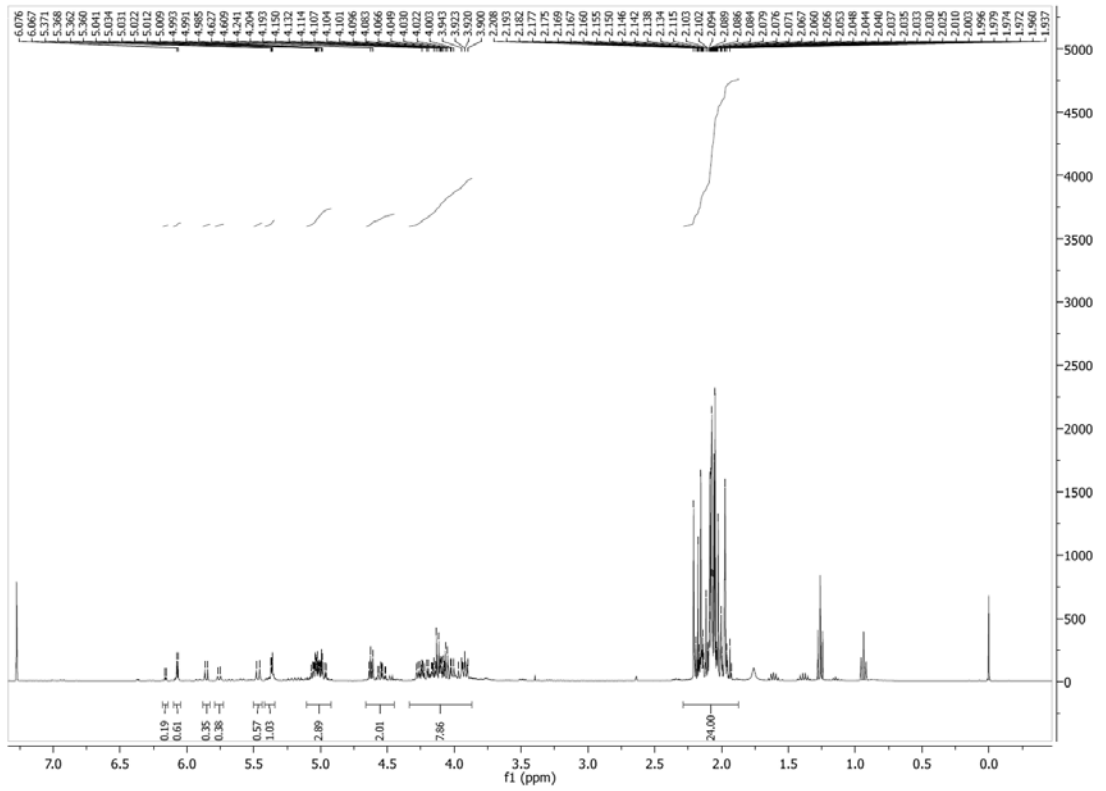


Figure A2.4. ^{1}H NMR of 4-Nitrophenyl 2-acetamido-4,6-di-*O*-acetyl-2-deoxy-3-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside.

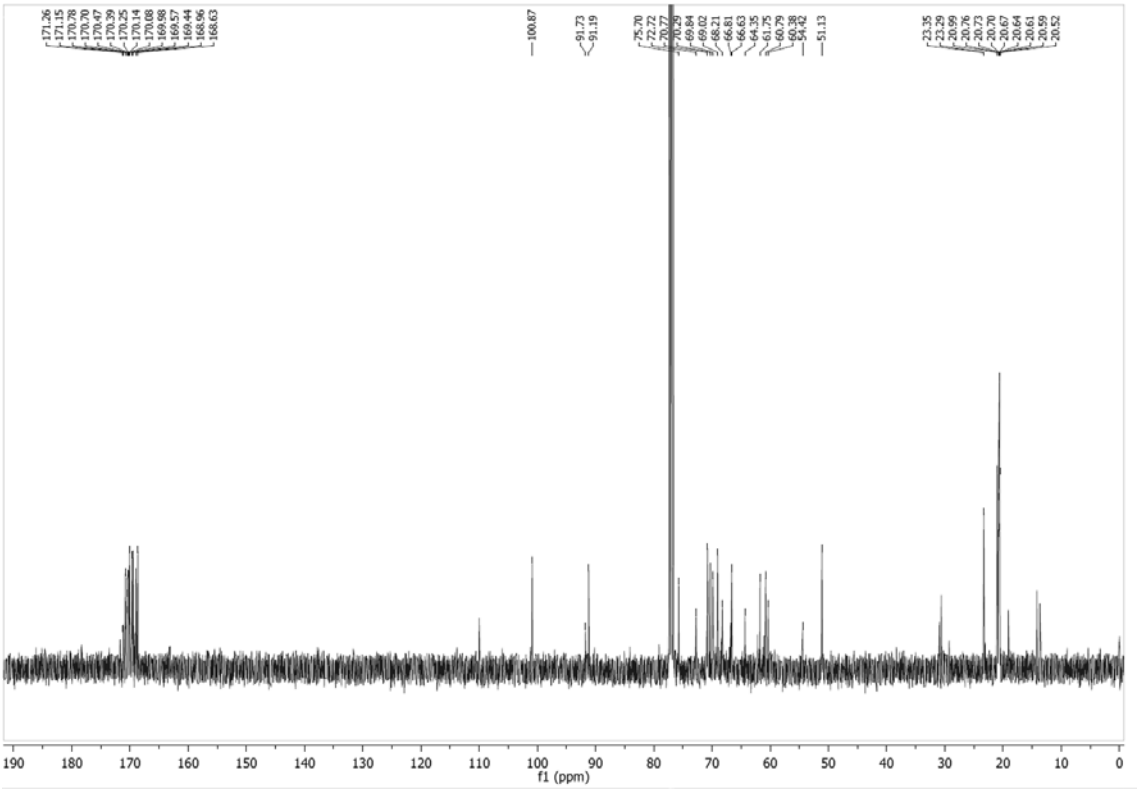


Figure A2.5 ^{13}C NMR of 4-Nitrophenyl 2-acetamido-4,6-di-*O*-acetyl-2-deoxy-3-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside.

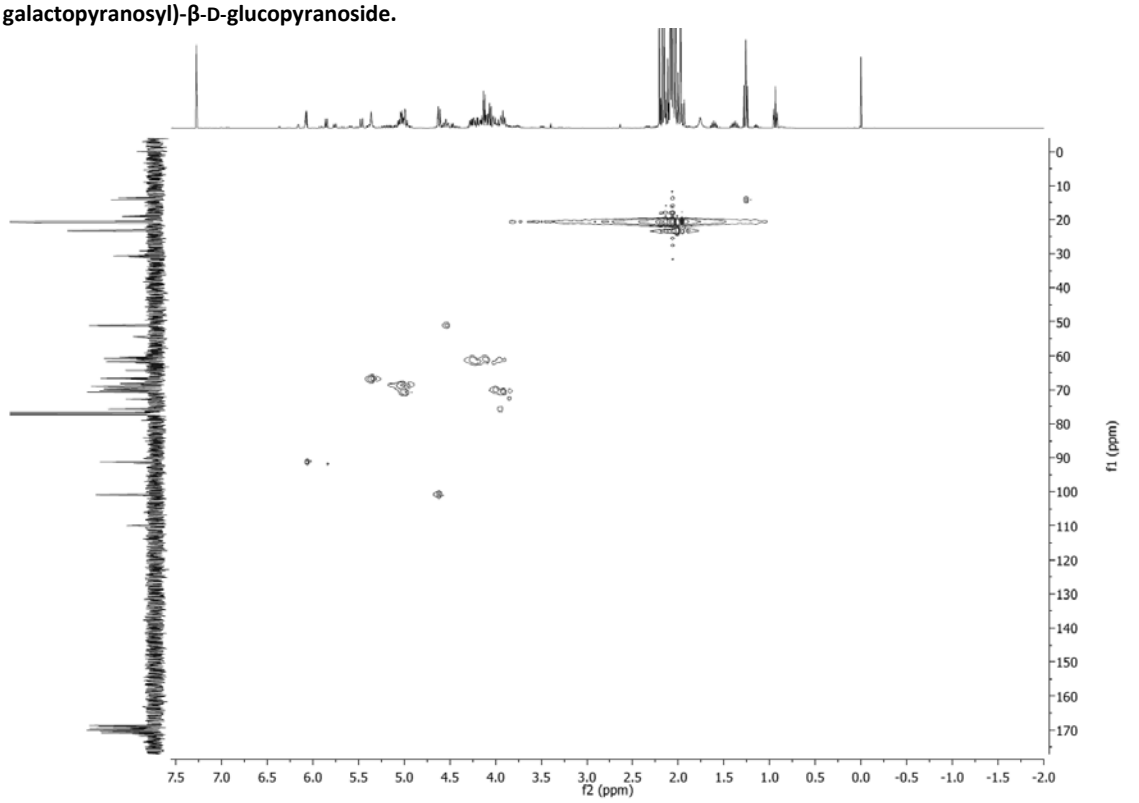


Figure A2.6 HSQC of 4-Nitrophenyl 2-acetamido-4,6-di-*O*-acetyl-2-deoxy-3-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-β-D-glucopyranoside.

Compound **9**

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride

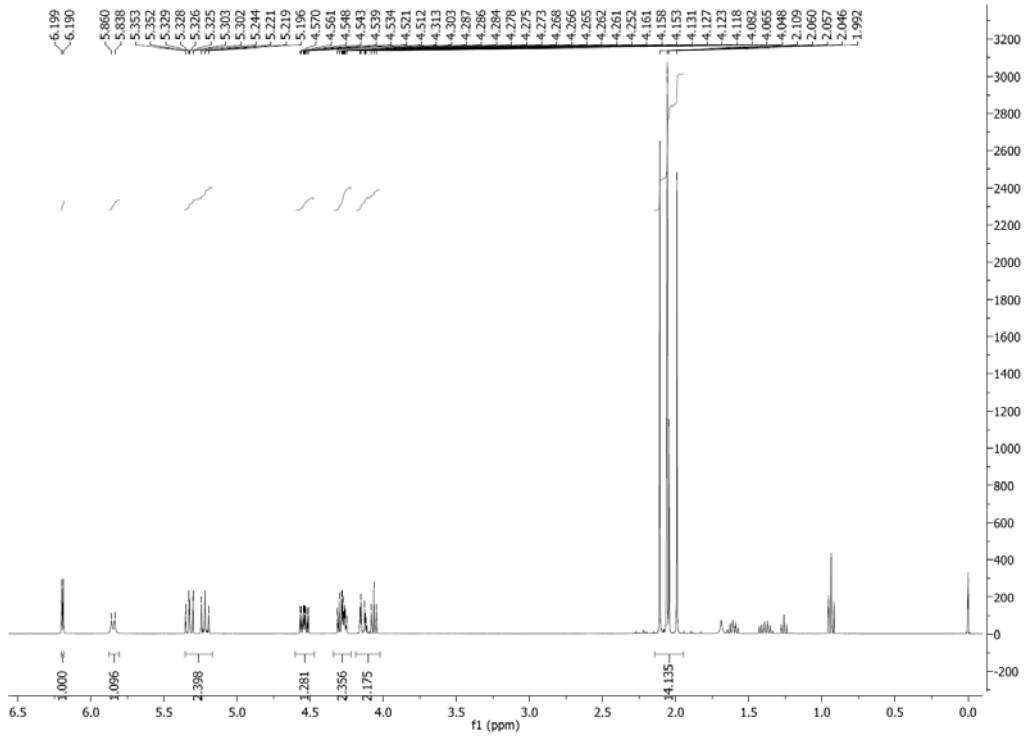


Figure A2.7. ^{1}H NMR of 2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride.

Compound **10**

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl acetate

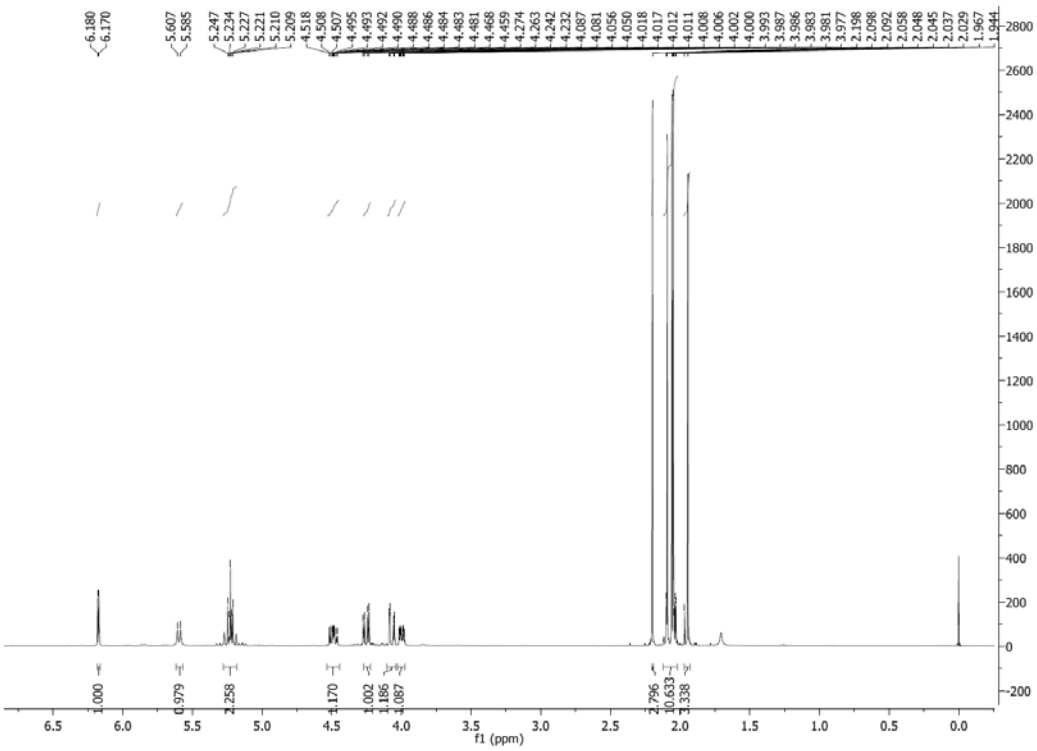


Figure A2.8. ^{1}H NMR of 2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl acetate.

Compound **11**

4-Nitrophenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranoside

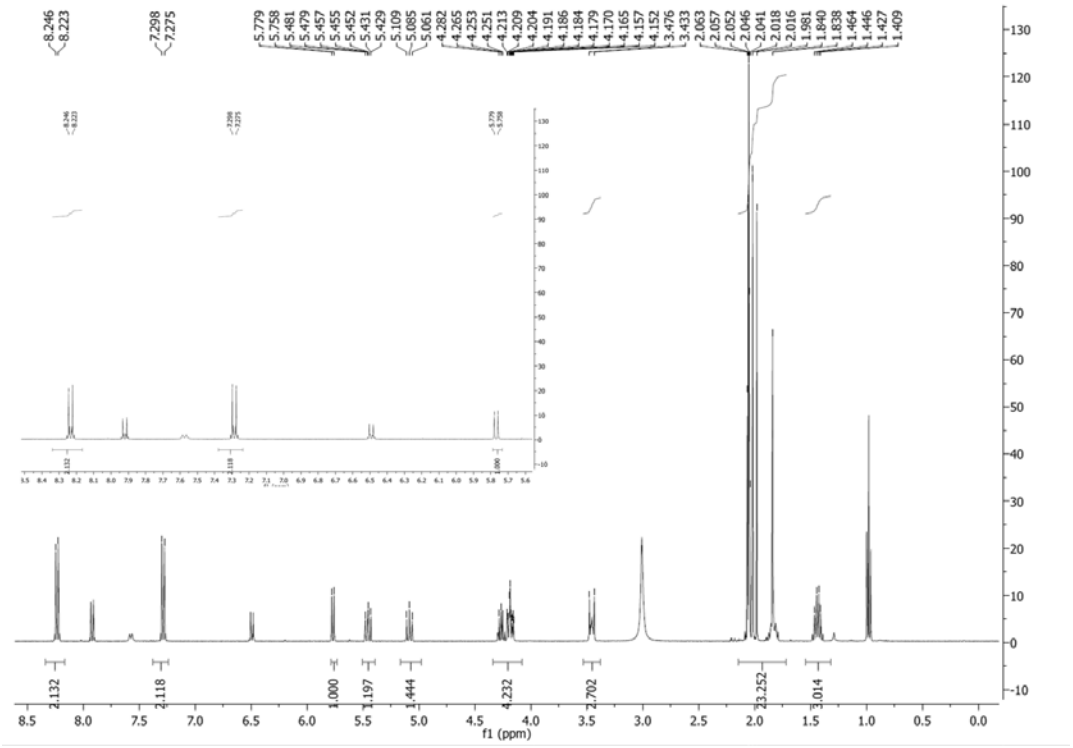


Figure A2.9. ^{1}H NMR of 4-Nitrophenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranoside.

Compound **8**

4-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside

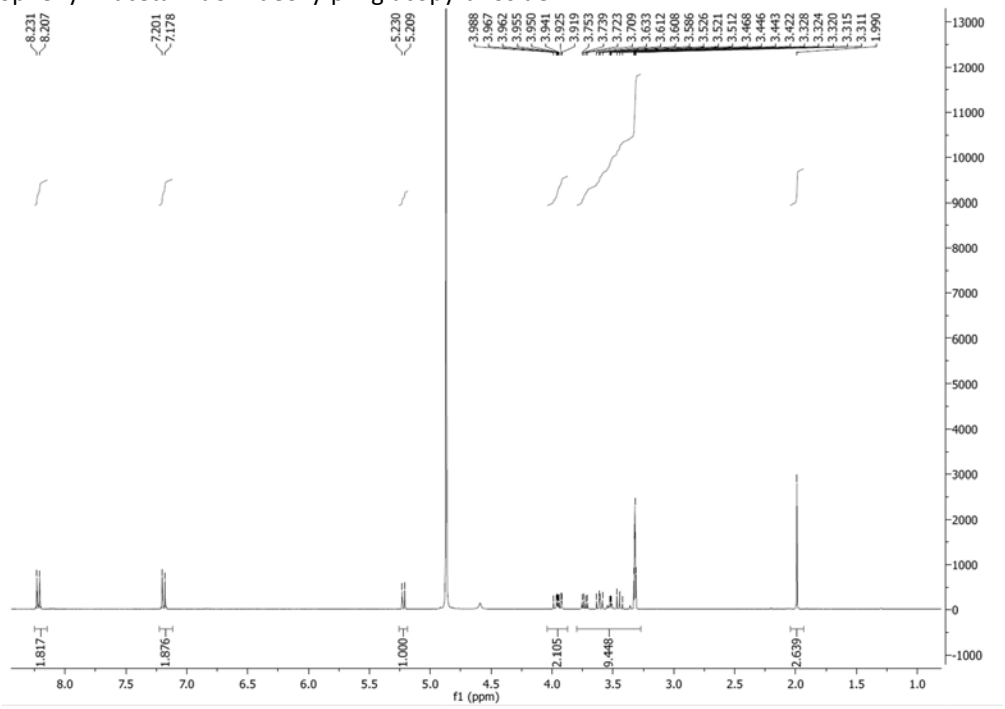


Figure A2.10. ^{1}H NMR of 4-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside.

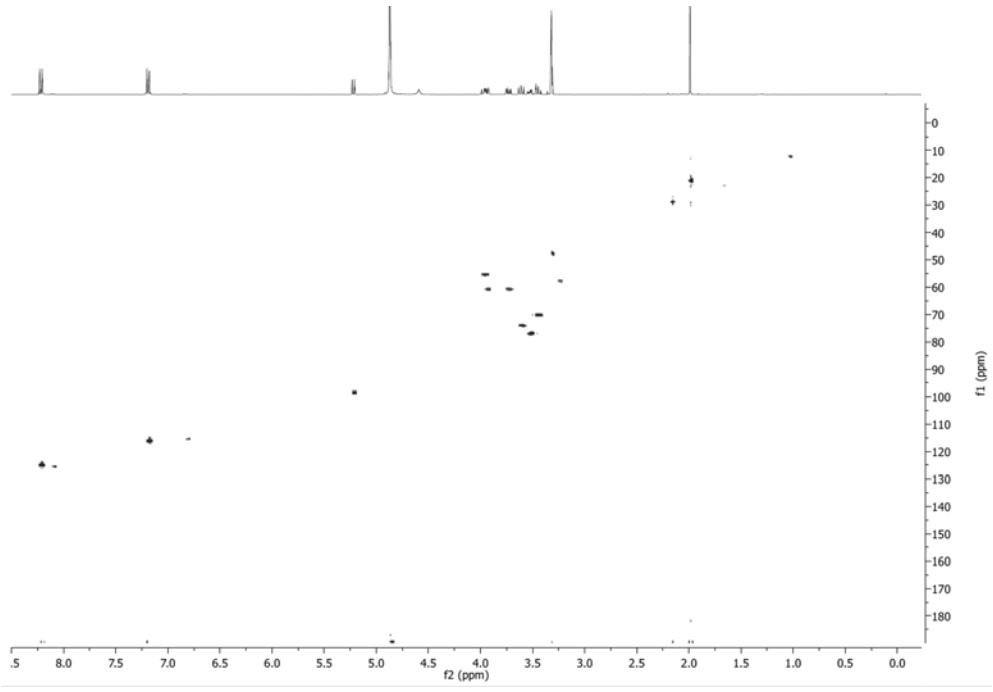


Figure A2.11. HSQC of 4-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside.

Compound **7**

4-nitrophenyl 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside

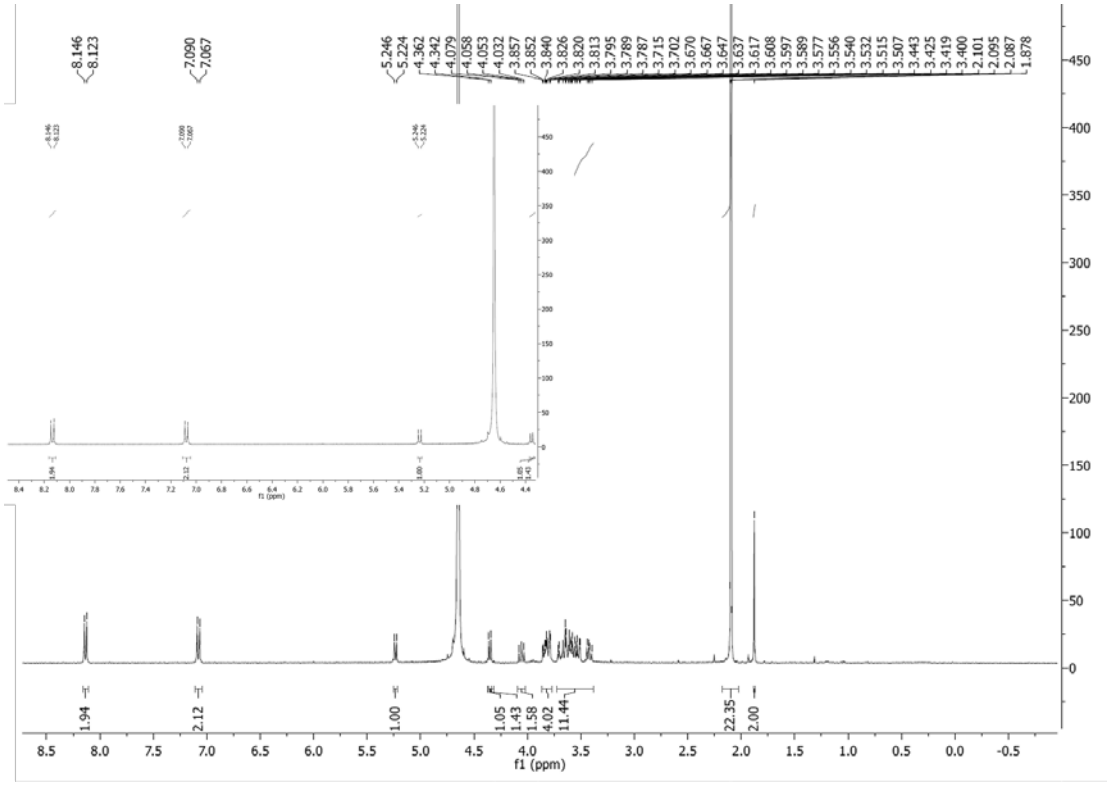


Figure A2.12. ^{1}H NMR of 4-nitrophenyl 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside.

Compound **12**

Per-*O*-acetated *N*-acetyl-D-glucosamine

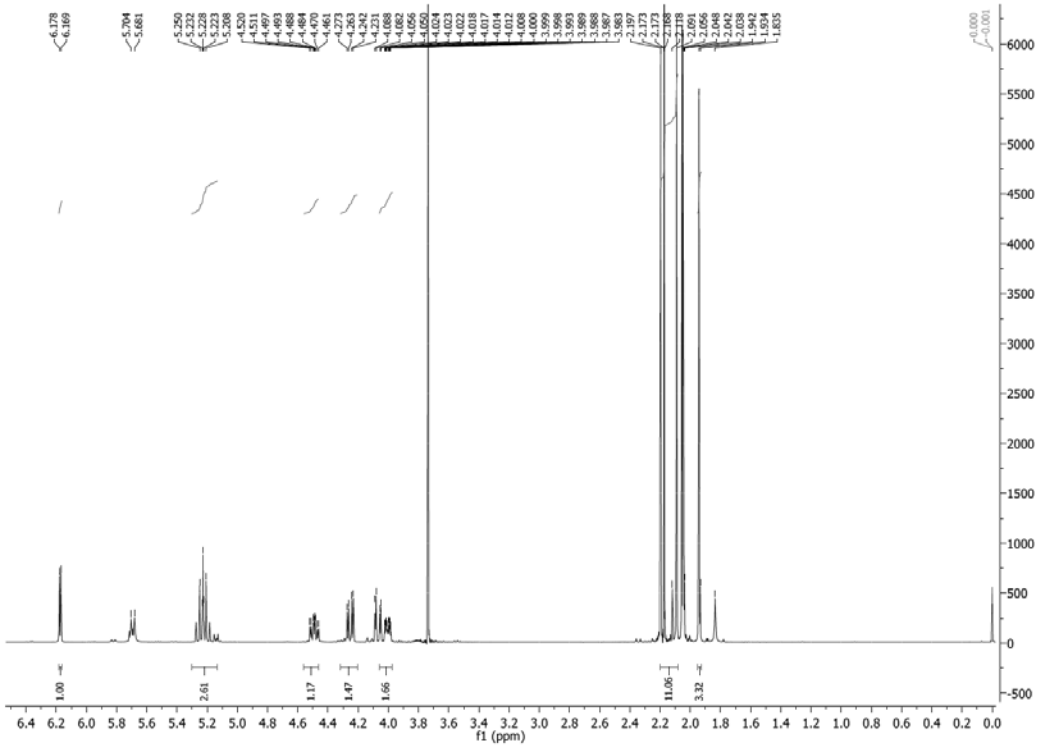


Figure A2.13. ^{1}H NMR of Per-*O*-acetated *N*-acetyl-D-glucosamine.

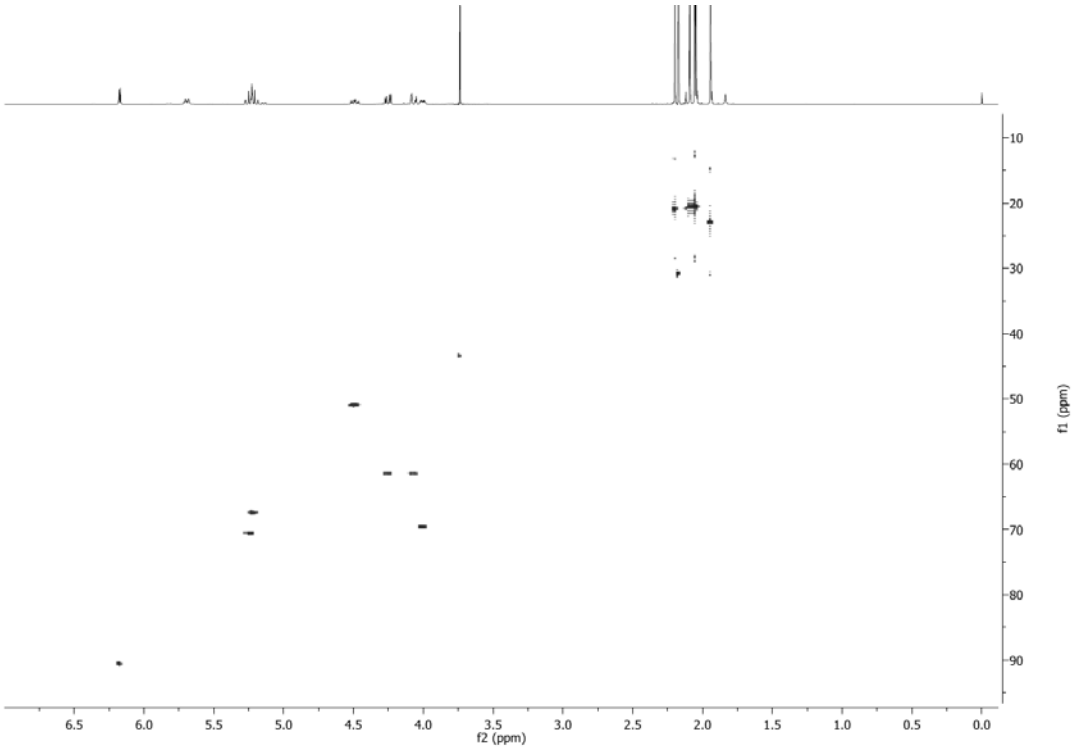


Figure A2.14. HSQC of Per-*O*-acetated *N*-acetyl-D-glucosamine.

Compound **13**

Per-*O*-acetated 1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine

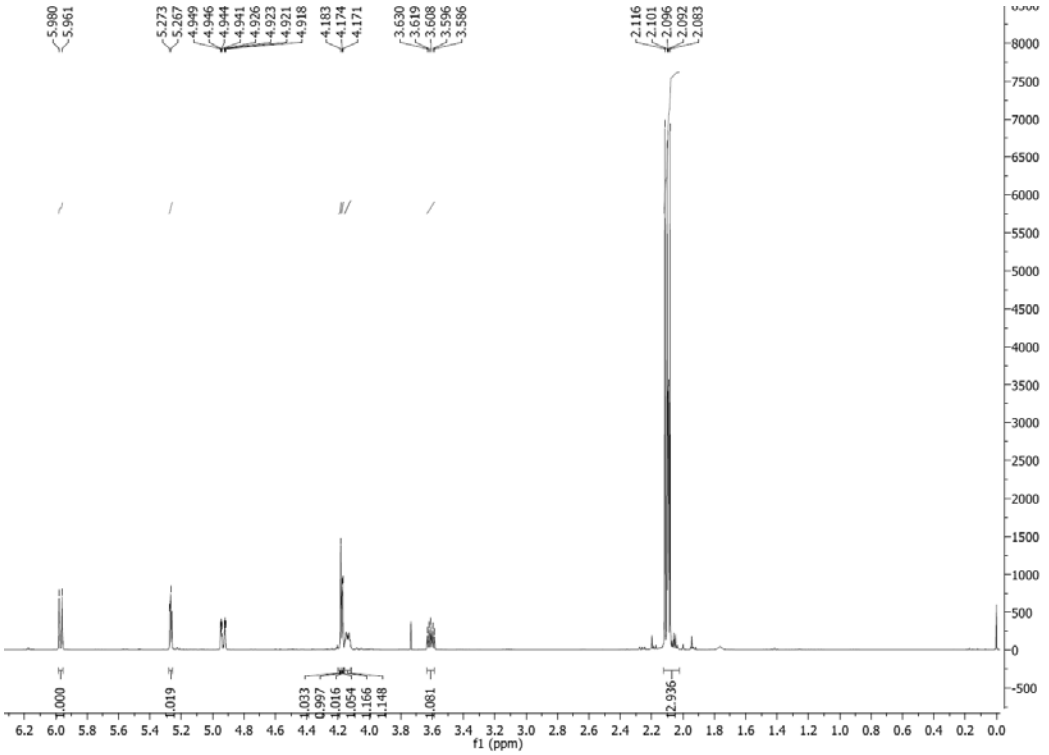


Figure A2.15. ^{1}H NMR of per-*O*-acetated 1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine.

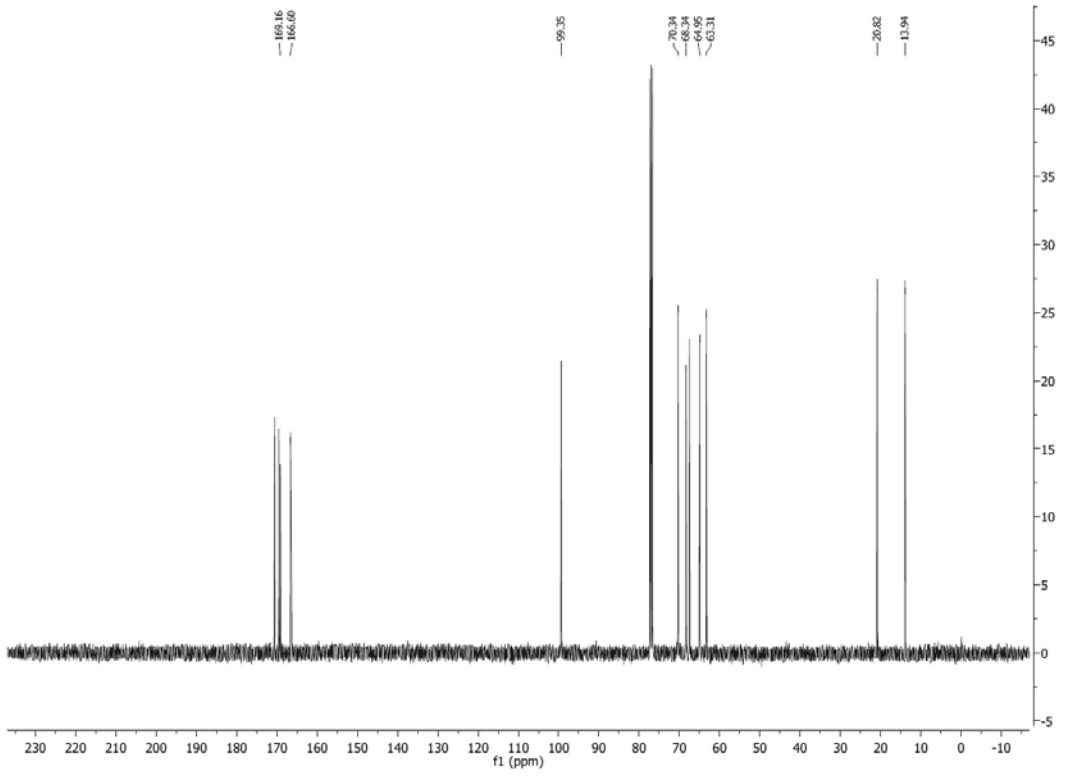


Figure A2.16. ^{13}C NMR of per-*O*-acetated 1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine.

Compound **14**

1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine

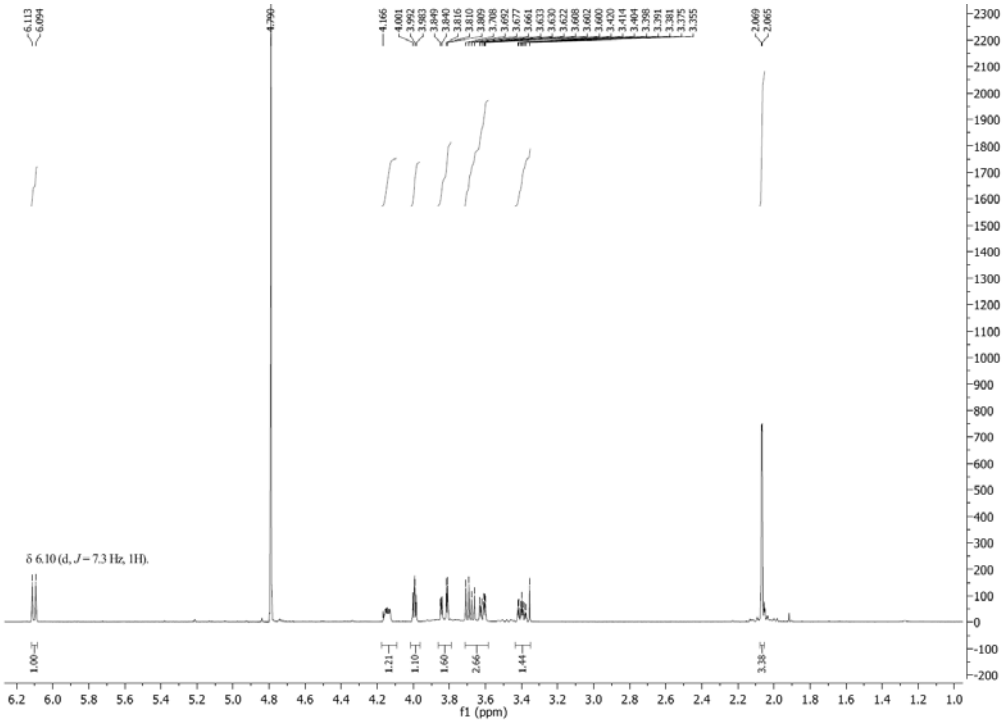


Figure A2.17. ^{1}H NMR of 1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine.

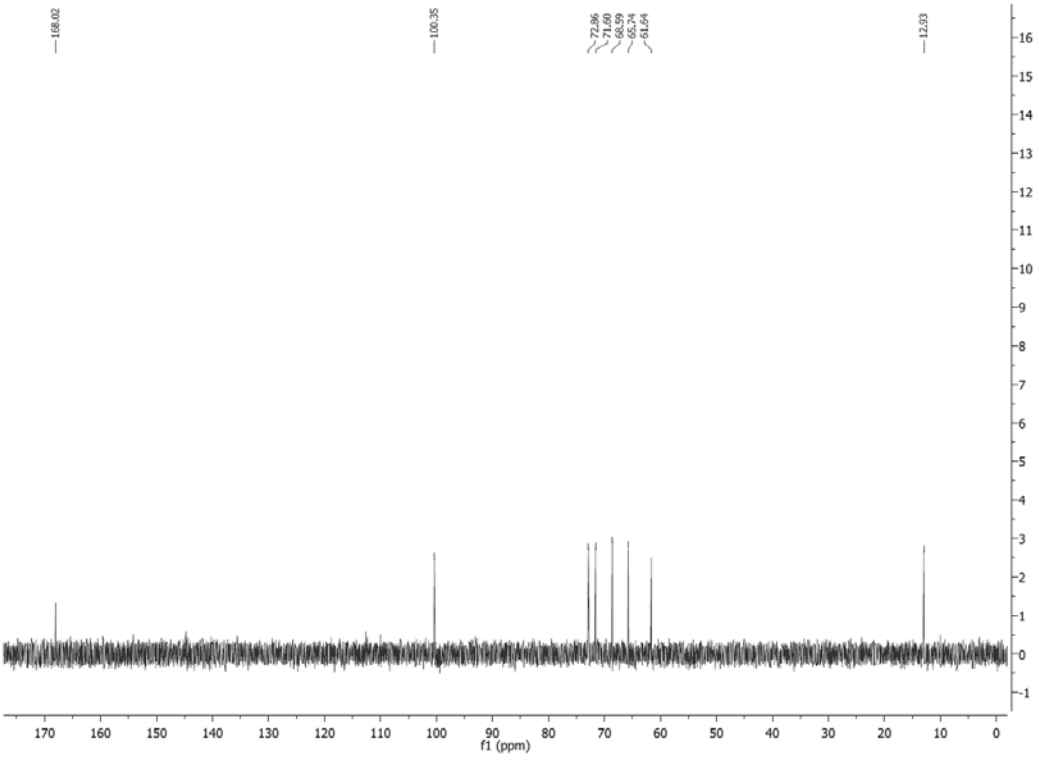


Figure A2.18. ^{13}C NMR of 1,2-ozaxoline derivative of *N*-acetyl-D-glucosamine.

Compound **15**

1,2-ozaxoline derivative of 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose (Lacto-*N*-biose)

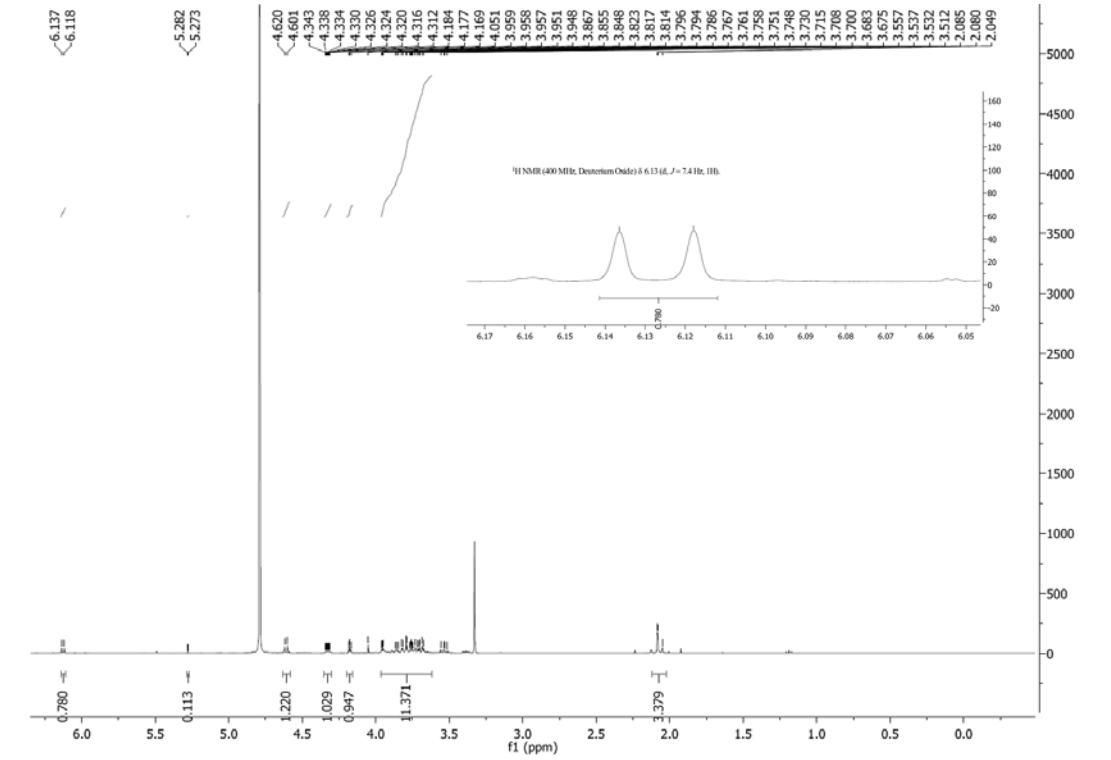


Figure A2.19. ^{1}H NMR of 1,2-ozaxoline derivative of 2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose.

APPENDIX 3

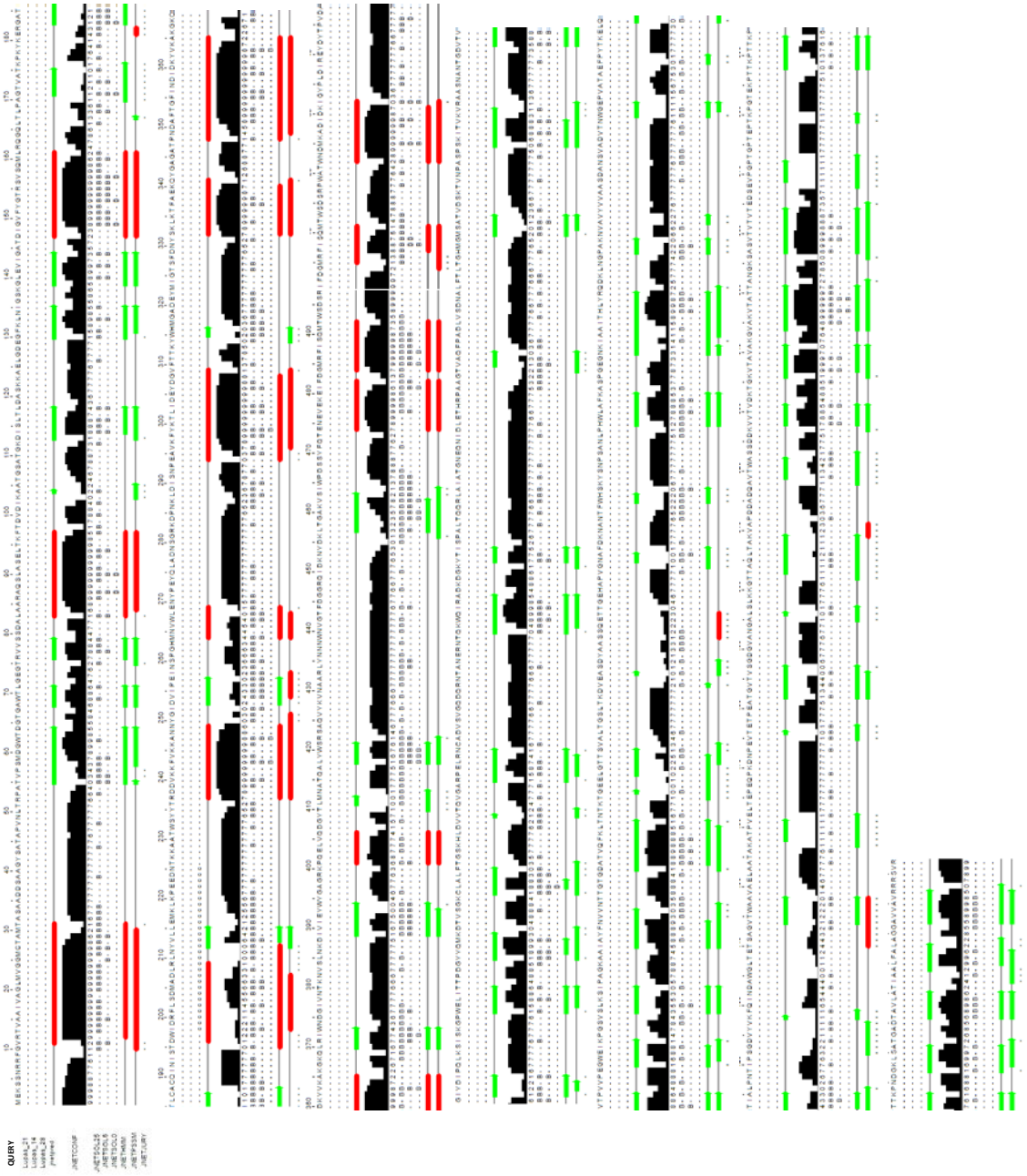


Figure A3.1. Prediction of the secondary structure of full length LnbB using JPred3.
QUERY shows the amino acidic sequence of the protein. Jnetpred shows the prediction of the secondary structure. JNETCONF indicates de probability of the prediction

```
Conservation:
sp_P49610_STRH_STRPN_Beta       1  MKH--------EKQQ-----RFSIRKYAVGAASVLIGFAFQAQTVAADGVTPTTTENQPTIHTVSDSPQS   57
tr_Q8DRL6_Q8DRL6_STRR6_Be       1  MKH--------EKQQ-----RFSIRKYAVGAASVLIGFAFQAQTVAADGVTPTTTENQPTIHTVSDSPQS   57
tr_D2KW09_D2KW09_9BACL_He       1  ----------------------------MMS-----------------------------------------    3
tr_O85361_O85361_STRPL_B-       1  MT-------------------------TGAA-----------------------------------------    6
tr_B3TLD6_B3TLD6_BIFBI_La       1  MEK--------SSNR-----RFGVRTVAAIVAGLMVGGMCTAMTA-------------------------   32
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ----------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito       1  MN-------------------AFKLSALARLTATMGFLGGMGSAMADQQLVDQLSQLKLNVKMLDNRAG   50
tr_Q840G9_Q840G9_AGGAC_Ds          ----------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-       1  MA-----------------------------------------------------------------    2
tr_Q06GJ0_Q06GJ0_OSTFU_N-       1  MW-------------------SRRIPLFIFGVLVLILSVAAE-----------DVVWRWSCDNGKCV   37
sp_P07686_HEXB_HUMAN_Beta       1  MELCGLGLPRPPMLLALLLATLLAAMLALLTQVALVVQVAEA----------------------------   42
sp_P06865_HEXA_HUMAN_Beta       1  MT----------------------SSRLWFSLLL---------------------------------   12
Consensus_ss:                                               hh

Conservation:
sp_P49610_STRH_STRPN_Beta      58  SENRTEETPKAVLQPEAPKTVETETPATDKVASLPKTEEKPQEEVSSTPSDKAEVVTPTSAEKETANKKA  127
tr_Q8DRL6_Q8DRL6_STRR6_Be      58  SENRTEETPKAELQPEAPKTVETETPATDKVASLPKTEEKPQEEVSSTPSDKAEVVTPTSAEKETANKKE  127
tr_D2KW09_D2KW09_9BACL_He          ----------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-          ----------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La          ----------------------------------------------------------------------
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ----------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito      51  ENGVDCAALGADWASCNRVLFTLSNDGQAIDGKDWVIYFHSPRQTLRVDNDQFKIAHLTGDLYKLEPTAK  120
tr_Q840G9_Q840G9_AGGAC_Ds          ----------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ----------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-      38  K-----------------------------------------------------------------    38
sp_P07686_HEXB_HUMAN_Beta          ----------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ----------------------------------------------------------------------
Consensus_ss:

Conservation:
sp_P49610_STRH_STRPN_Beta     128  EEASPKKEEAKEVDSKESNTDKTDKDKPAKKDEAKAEADKPATEAGKERAATVNEKLAKKKIVSIDAGRK  197
tr_Q8DRL6_Q8DRL6_STRR6_Be     128  EEASPKKEEAKEVDSKESNTDKTDKDKPAKKDEAKAEADKPETETGKERAATVNEKLAKKKIVSIDAGRK  197
tr_D2KW09_D2KW09_9BACL_He          ----------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-          ----------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La          ----------------------------------------------------------------------
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ----------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito     121  FSGFPAGKAVEIPVVAEYWQLFRNDFLPRWYATSGDAKPKMLANTDTENLDQFVAPFTGDQWKRTKDDKN  190
tr_Q840G9_Q840G9_AGGAC_Ds          ----------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ----------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ----------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ----------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ----------------------------------------------------------------------
Consensus_ss:

Conservation:
sp_P49610_STRH_STRPN_Beta     198  YFSPEQLKEIIDKAKHYGYTDLHLLVGNDGLRFMLDDMSITANGKTYASDDVKRAIEKGTNDYYNDPNGN  267
tr_Q8DRL6_Q8DRL6_STRR6_Be     198  YFSPEQLKEIIDKAKHYGYTDLHLLVGNDGLRFMLDDMSITANGKTYASDDVKRAIEKGTNDYYNDPNGN  267
tr_D2KW09_D2KW09_9BACL_He       4  ------------FIPESASASTS-----------------------------------------------   14
tr_O85361_O85361_STRPL_B-       7  ------------PDRKAPVRPTP-----------------------------------------------   17
tr_B3TLD6_B3TLD6_BIFBI_La      33  ----SAADDSAAGYSATAPVNLT-----------------------------------------------   51
tr_A1RBZ5_A1RBZ5_ARTAT_Be       1  --------------------MSA-----------------------------------------------    3
sp_Q54468_CHB_SERMA_Chito     191  ILMTPASRFVSNADLQTLPAGAL-----------------------------------------------  213
tr_Q840G9_Q840G9_AGGAC_Ds          ----------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ----------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-      39  LKNDPRSSEPALSLEACKMFCNE-----------------------------------------------   61
sp_P07686_HEXB_HUMAN_Beta      43  -------------ARAPSVSAKP-----------------------------------------------   52
sp_P06865_HEXA_HUMAN_Beta      13  ---------------AAAFAGR-----------------------------------------------   19
Consensus_ss:

Conservation:                                  9 6
sp_P49610_STRH_STRPN_Beta     268  HLTESQMTDLINYAKDKGIGLIPTVNSPGHMDAILNAMKELGIQNPNFSYFGKKSARTVDL---------  328
tr_Q8DRL6_Q8DRL6_STRR6_Be     268  HLTESQMTDLINYAKDKGIGLIPTVNSPGHMDAILNAMKELGIQNPNFSYFGKKSARTVDL---------  328
tr_D2KW09_D2KW09_9BACL_He      15  -----------------QPSILPKPVS-------------YTVGSGQFVLT-KNASIFVAGNNV---GET   50
tr_O85361_O85361_STRPL_B-      18  -----------------LDRVIPAPAS-------------VDPGGAPYRIT-RGTHIRVDDS--------   48
tr_B3TLD6_B3TLD6_BIFBI_La      52  -----------------RPATVPSMDG-------------WTDGTGAWTLG-EGTRVVSSD---------   81
tr_A1RBZ5_A1RBZ5_ARTAT_Be       4  -----------------FHQLIPAPWS-------------FSAGTGELELD-AGTSVGADP---------   33
sp_Q54468_CHB_SERMA_Chito      214  -----------------RGKIVPTPMQ-------------VKVHAQDADLR-KGVALDLST---------  243
tr_Q840G9_Q840G9_AGGAC_Ds          ----------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-       3  ------------------------------------------TFLGLSSKQE-KALVRLDK----------   20
tr_Q06GJ0_Q06GJ0_OSTFU_N-      62  -----------------YGLLWPRPTG-------------EADLGNFLSKI-NLNSIEVKILKK--GATD   98
sp_P07686_HEXB_HUMAN_Beta      53  -----------------GPALWPLPLS-------------VKMTPNLLHLA-PENFYISHSPN-------   84
sp_P06865_HEXA_HUMAN_Beta      20  -----------------ATALWPWPQN-------------FQTSDQRYVLY-PNNFQFQYDVSSAAQPGC   58
Consensus_ss:                                          e             eee   eeee    eeeeee

Conservation:
sp_P49610_STRH_STRPN_Beta     329  -DNEQAVAFTKALIDKYAAYFAKKTEIFNIGLDEYANDATDAKGWSVLQADKYYPNEGYPVKGYEKFIAY  397
tr_Q8DRL6_Q8DRL6_STRR6_Be     329  -DNEQAVAFTKALIDKYAAYFAKKTEIFNIGLDEYANDATDAKGWSVLQADKYYPNEGYPVKGYEKFIAY  397
tr_D2KW09_D2KW09_9BACL_He      51  DELFNIGQALAKKLNASTGYT--------------------------------------------------   71
tr_O85361_O85361_STRPL_B-      49  REARRVGDYLADLLRPATGYR--------------------------------------------------   69
```

```
tr_B3TLD6_B3TLD6_BIFBI_La      82  -ALAARAQSLASELTKFTDVD-------------------------------------------------  101
tr_A1RBZ5_A1RBZ5_ARTAT_Be      34  -ELRGPRRWLTRALGGATGWD-------------------------------------------------   53
sp_Q54468_CHB_SERMA_Chito     244  -LVKPAADVVSQRFALLGVPV-------------------------------------------------  263
tr_Q840G9_Q840G9_AGGAC_Ds          ---------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ---------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-      99  DLMEAAAKRFKEQVSLAIPRG-------------------------------------------------  119
sp_P07686_HEXB_HUMAN_Beta      85  STAGPSCTLLEEAFRRYHGYI-------------------------------------------------  105
sp_P06865_HEXA_HUMAN_Beta      59  SVLDEAFQRYRDLLFGSGSWP-------------------------------------------------   79
Consensus_ss:                       hhhhhhhhhhhhhhhhh

Conservation:                                                5
sp_P49610_STRH_STRPN_Beta     398  ANDLARIVKSHGLKPMAFND-GIYYNSDTSFGSFDKDIIVSMWTGGWGGYDVASSKLLAEKGHQILNTND  466
tr_Q8DRL6_Q8DRL6_STRR6_Be     398  ANDLARIVKSHGLKPMAFND-GIYYNSDTSFGSFDKDIIVSMWTGGWGGYDVASSKLLAEKGHQILNTND  466
tr_D2KW09_D2KW09_9BACL_He      72  -------ISV-------VKSN----------QPTAGSIYLTTVG-------------------------   91
tr_O85361_O85361_STRPL_B-      70  -------LPV-------TAHG----------HGGIRLRLAG-----------------------------   86
tr_B3TLD6_B3TLD6_BIFBI_La     102  -------IKA-------ATGS----------ATGKDISLTLDA-------------------------  120
tr_A1RBZ5_A1RBZ5_ARTAT_Be      54  -------LAP-------APAR----------EAGIRLLLDP-------------------------   70
sp_Q54468_CHB_SERMA_Chito     264  -------QTN-------GYP-----------IKTDIQPGK-------------------------  278
tr_Q840G9_Q840G9_AGGAC_Ds          ---------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      21  ----------------YLN----------LGEIAVSLVT-------------------------   33
tr_Q06GJ0_Q06GJ0_OSTFU_N-     120  -------STP-------KLTG----------KAVDVYLVNEN-----------------------  137
sp_P07686_HEXB_HUMAN_Beta     106  -------FGFYKWHHEPAEFQAKTQ--------VQQLLVSITLQS-------------------  135
sp_P06865_HEXA_HUMAN_Beta      80  -------RPY-------LTGKRHTL--------EKNVLVVSVVTP-------------------  102
Consensus_ss:                                                eeeee

Conservation:
sp_P49610_STRH_STRPN_Beta     467  AWYYVLGRNADGQGWYNLDQGLNGIKNTPITSVPKTEGADIPIIGGMVAAWADTPSARYSPSRLFKLMRH  536
tr_Q8DRL6_Q8DRL6_STRR6_Be     467  AWCYVLGRNADGQGWYNLDQGLNGIKNTPITSVPKTEGADIPIIGGMVAAWADTPSARYSPSHLFKLMRH  536
tr_D2KW09_D2KW09_9BACL_He          ---------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-          ---------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La          ---------------------------------------------------------------------
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ---------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ---------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ---------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ---------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ---------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ---------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ---------------------------------------------------------------------
Consensus_ss:

Conservation:                                                  6 5 6                  7
sp_P49610_STRH_STRPN_Beta     537  FANANAEYFAADYESAEQALNEVPKDLNRYTAESVTAVKEAEKAIRSLDS-----NLSRAQQDTIDQAIA  601
tr_Q8DRL6_Q8DRL6_STRR6_Be     537  FANANAEYFAADYESAEQALNEVPKDLNRYTAESVAAVKEAEKAIRSLDS-----NLSRAQQDTIDQAIA  601
tr_D2KW09_D2KW09_9BACL_He      92  ------------------------------------GNAALGNEGYDLITTS-----NQVTLTANKPEGVFR  122
tr_O85361_O85361_STRPL_B-      87  ------------------------------------GPYGDEGYRLDSGP-----AGVTITARKAAGLFH  115
tr_B3TLD6_B3TLD6_BIFBI_La     121  ------------------------------------SKKAELGDEGFKLNIGS-----KGLEVIGATDIGVFY  152
tr_A1RBZ5_A1RBZ5_ARTAT_Be      71  ---------------------------------------SLDAEAYRLEVS------DAVVISAGGAAGAFY   97
sp_Q54468_CHB_SERMA_Chito     279  ------------------------------------FKGAMAVSGAYELKIGK-----KEAQVIGFDQAGVFY  310
tr_Q840G9_Q840G9_AGGAC_Ds       1  ------------------------------------------------------NCCVKGN-------    7
tr_Q6ST21_Q6ST21_STRGN_N-      34  ---------------------------------------DSATSIKVEGRQ-----GYYQVSYKQPHQLYR   60
tr_Q06GJ0_Q06GJ0_OSTFU_N-     138  ------------------------------PNEKAFSLEMDESYGLRVSPSGADRVNATITANSFFGMRH  177
sp_P07686_HEXB_HUMAN_Beta     136  ------------------------------ECDAFPNISSDESYTLLVKE-----PVAVLKANRVWGALR  170
sp_P06865_HEXA_HUMAN_Beta     103  ------------------------------GCNQLPTLESVENYTLTIND-----DQCLLLSETVWGALR  137
Consensus_ss:                                                    eeeee        eeeeee  hhhhhh

Conservation:                      6     556                        5    55  68  8         55 56  7  6
sp_P49610_STRH_STRPN_Beta     602  KLQETVNNLTLTP---------EAQKEEEAKREVEKLAKNKVISIDAGRK-YFTLNQLKRIVDKASELGY  661
tr_Q8DRL6_Q8DRL6_STRR6_Be     602  KLQETVNNLTLTP---------EAQKEEEAKREVEKLAKNKVISIDAGRK-YFTLDQLKRIVDKASELGY  661
tr_D2KW09_D2KW09_9BACL_He     123  GNQTLLQLLPAGIEKNT-VVSGVQWVIPHSNISDKPEYEYRGLMLDVARH-FFTVDEVKRQIDLASQYKI  190
tr_O85361_O85361_STRPL_B-     116  GVQTLRQLLPPAVEKDS--AQPGPWLVAGGTIEDTPRYAWRSAMLDVSRH-FFGVDEVKRYIDRVARYKY  182
tr_B3TLD6_B3TLD6_BIFBI_La     153  GTRSVSQMLRQG-----------QLTLPAGTVATKPKYKERGATLCACQI-NISTDWIDRFLSDMADLRL  210
tr_A1RBZ5_A1RBZ5_ARTAT_Be      98  GAQTLLQLLGPAALRQAPVVAVEGWSVPRVSVEDKPRFGYRGTMLDVARH-FMPKDNVLRFIEVMAMHKL  166
sp_Q54468_CHB_SERMA_Chito     311  GLQSILSLVPSDG----------SGKIATLDASDAPRFPYRGIFLDVARN-FHKKDAVLRLLDQMAAYKL  369
tr_Q840G9_Q840G9_AGGAC_Ds       8  ----------------------------SIYPQKTSTKQTGLMLDIARH-FYSPEVIKSFIDTISLSGG   47
tr_Q6ST21_Q6ST21_STRGN_N-      61  ALALLSAALRSG-------------QDEVQIEEEAAYEDLAYMADCSRNAVLNLSSAKKMIEVLALMGY  116
tr_Q06GJ0_Q06GJ0_OSTFU_N-     178  GLETLSQLFVFDDIR------DHLLMVRDVNISDKPVYPYRGILLDTARN-YYSIESIKRTIEAMAAVKL  240
sp_P07686_HEXB_HUMAN_Beta     171  GLETFSQLVYQDSY--------GTFTINESTIIDSPRFSHRGILIDTSRH-YLPVKIILKTLDAMAFNKF  231
sp_P06865_HEXA_HUMAN_Beta     138  GLETFSQLVWKSAE--------GTFFINKTEIEDFPRFPHRGLLLDTSRH-YLPLSSILDTLDVMAYNKL  198
Consensus_ss:                      hhhhhhhhh               eee  eeeee       eeeee        hhhhhhhhhhhhh

Conservation:                      7 5655  6          6                  6                        67
sp_P49610_STRH_STRPN_Beta     662  SDVHLLLGNDG-LRFLLDDMTITANGKTYASDDVKKAIIEGTKAYY--------------DDPNGTALT  715
tr_Q8DRL6_Q8DRL6_STRR6_Be     662  SDVHLLLGNDG-LRFLLNDMTITANGKTYASDDVKKAIIEGTKAYY--------------DDPNGTALT  715
tr_D2KW09_D2KW09_9BACL_He     191  NKFHMHLSDDQGWRIEIKSWPDLIE--------------IGSKGQV--------------GGGPGGYYT  231
tr_O85361_O85361_STRPL_B-     183  NKLHLHLSDDQGWRIAIDSWPRLAT--------------YGGSTEV--------------GGGPGGYYT  223
tr_B3TLD6_B3TLD6_BIFBI_La     211  NYVLLEMKLKP----------EEDN--------------TKKAA--------------------TWSYYT  236
tr_A1RBZ5_A1RBZ5_ARTAT_Be     167  NVLHLHLTDDQGWRMQINRYPKLTE--------------TGAWRRESSLGSW---RAGVFDGRPHGGFYT  219
sp_Q54468_CHB_SERMA_Chito     370  NKFHFHLSDDEGWRIEIPGLPELTE--------------VGGQRCHDLSETTCLLPQYGQGPDVYGGFFS  425
tr_Q840G9_Q840G9_AGGAC_Ds      48  NFLHLHFSDHENYAIESHLLNQRAE--------------NAVQGKDGIY----------INPYTGKPFLS   93
tr_Q6ST21_Q6ST21_STRGN_N-     117  STFELYMEDTYEIE----NQPYF-----------------------------------------GYFRGRYT  143
```

```
tr_Q06GJ0_Q06GJ0_OSTFU_N-     241  NTFHWHITDSQSFPFVTTKRPNLYK--------------FGALS--------------------PQKVYT  276
sp_P07686_HEXB_HUMAN_Beta     232  NVLHWHIVDDQSFPYQSITFPELSN--------------KGSYS--------------------LSHVYT  267
sp_P06865_HEXA_HUMAN_Beta     199  NVFHWHLVDDPSFPYESFTFPELMR--------------KGSYN--------------------PVTHIYT  235
Consensus_ss:                       eeeeeee             hhhhh              h

Conservation:                            5   68  5 5 658666 869  5
sp_P49610_STRH_STRPN_Beta     716  QAEVTELIEYAKSKDIGLIPAINSPGHMDAMLVAMEKLG----IKNP-----------------QAHFD  763
tr_Q8DRL6_Q8DRL6_STRR6_Be     716  QAEVTELIEYAKSKDIGLIPAINSPGHMDAMLVAMEKLG----IKNP-----------------QAHFD  763
tr_D2KW09_D2KW09_9BACL_He     232  QEQFKDIVSYAAERYIEVIPEIDMPGHTNAALASYGELN----PDGKR---------------KAMRTDT  282
tr_O85361_O85361_STRPL_B-     224  KAEYKEIVRYAASRHLEVVPEIDMPGHTNAALASYAELN----CDGVA---------------PPLYTGT  274
tr_B3TLD6_B3TLD6_BIFBI_La     237  RDDVKKFVKKANNYGIDVIPEINSPGHMNVWLENYPEYQ----LADN---------------------SG  281
tr_A1RBZ5_A1RBZ5_ARTAT_Be     220  QDDLREIVAFAADRHITVIPEIDVPGHSQAAIAAYPELG----AGPADG-----------SSPVEVWTR  273
sp_Q54468_CHB_SERMA_Chito     426  RQDYIDIIKYAQARQIEVIPEIDMPAHARAAVVSMEARYKKLHAAGKEQEANEFRLVDPTDTSNTTSVQF  495
tr_Q840G9_Q840G9_AGGAC_Ds      94  YRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDR----GVKY-----------------LQGLKS  142
tr_Q6ST21_Q6ST21_STRGN_N-     144  VAELQEIEDYAADFDMSFVPCIQTLAHLSAFVKWGIKEV----------------------------QEL  185
tr_Q06GJ0_Q06GJ0_OSTFU_N-     277  KAAIREVVRFGLERGVRVLPEFDAPAHVGEGWQDTDLTV----CFKAE---------------PWKSYCV  327
sp_P07686_HEXB_HUMAN_Beta     268  PNDVRMVIEYARLRGIRVLPEFDTPGHTLSWGKGQKDLL----TPCY-----------------SRQNK  315
sp_P06865_HEXA_HUMAN_Beta     236  AQDVKEVIEYARLRGIRVLAEFDTPGHTLSWGPGIPGLL----TPCY-----------------SGSEP  283
Consensus_ss:                       hhhhhhhhhhhhh    eeee    hhhhhhhhhh hhh

Conservation:                                  6      5    5          6 9599
sp_P49610_STRH_STRPN_Beta     764  KVSKTTMDLKNEEAMNFVKALIGKYMDFFAG---KTKIFNFGTDEYAND---------------------  809
tr_Q8DRL6_Q8DRL6_STRR6_Be     764  KVSKTTMDLKNEEAMNFVKALIGKYMDFFAG---KTKIFNFGTDEYAND---------------------  809
tr_D2KW09_D2KW09_9BACL_He     283  AVGYSTLMPRAEITYQFVEDVISELAAISP-----SPYIHLGGDESNA---------------------  325
tr_O85361_O85361_STRPL_B-     275  KVGFSSLCVDKDVTYDFVDDVIGELAALTP-----GRYLHIGGDEAHS---------------------  317
tr_B3TLD6_B3TLD6_BIFBI_La     282  RKDPNKLDISNPEAVKFYKTLIDEYDGVFT-----TKYWHMGADEYMIG--------------------  325
tr_A1RBZ5_A1RBZ5_ARTAT_Be     274  WGINETVLEVSETSLEFYRNVLDEVVEIFP-----SPWISLGGDEVPLT---------------------  317
sp_Q54468_CHB_SERMA_Chito     496  FNRQSYLNPCLDSSQRFVDKVIGEIAQMHKEAGQPIKTWHFGGDEAKNIRLGAGYTDKAKPEPGKGIIDQ  565
tr_Q840G9_Q840G9_AGGAC_Ds     143  RQVDDEIDITNADSITFMQSLMSEVIDIFGD---TSQHFHIGGDEFGYS---------------------  188
tr_Q6ST21_Q6ST21_STRGN_N-     186  RDVEDILLIGEEKVYDLIEGMFQTMAHLHT------RKINIGMDEAHL---------------------  227
tr_Q06GJ0_Q06GJ0_OSTFU_N-     328  EPPCGQLNPTKDELYQYLEDIYSDMAEVFD----TTDIFHMGGDEVSEA---------------------  372
sp_P07686_HEXB_HUMAN_Beta     316  LDSFGPINPTLNTTYSFLTTFFKEISEVFP-----DQFIHLGGDEVEFK---------------------  359
sp_P06865_HEXA_HUMAN_Beta     284  SGTFGPVNPSLNNTYEFMSTFFLEVSSVFP-----DFYLHLGGDEVDFT---------------------  327
Consensus_ss:                             hhhhhhhhhhhhhhhhh          eeeee

Conservation:                                                  5            6      8
sp_P49610_STRH_STRPN_Beta     810  ---------ATSAQGWYYLKW-----YQLYGKFAEYANTLAAMAKE--RGL-QPMAFNDGFYYEDKD---  859
tr_Q8DRL6_Q8DRL6_STRR6_Be     810  ---------ATSAQGWYYLKW-----YQLYGKFAEYANTLAAMAKE--RGL-QPMAFNDGFYYEDKD---  859
tr_D2KW09_D2KW09_9BACL_He     326  --------------------------TSAADYDYFFGRVTAIANS--YGK-KVVGWDPSDTSS------  359
tr_O85361_O85361_STRPL_B-     318  --------------------------TPKADFVAFMKRVQPIVAK--YGK-TVVGWHQLAGAE------  351
tr_B3TLD6_B3TLD6_BIFBI_La     326  -----TSFDNYSKLKTFAEKQYGA-GATPNDAFTGFINDIDKYVKA--KGK-QLRIWNDGIVNTK-----  381
tr_A1RBZ5_A1RBZ5_ARTAT_Be     318  -----QWQASAQAQAKAAELGL----DDVSGLHSWFVGQLALHLKH--HGR-ATSVWDEIGDGG------  369
sp_Q54468_CHB_SERMA_Chito     566  SNEDKPWAKSQVCQTMIKEGKV----ADMEHLPSYFGQEVSKLVKA--HGIDRMQAWQDGLKDAESS---  626
tr_Q840G9_Q840G9_AGGAC_Ds     189  -----VES---------------------NHEFITYANKLSYFLEK--KGL-KTRMWNDGLIKNT-----  224
tr_Q6ST21_Q6ST21_STRGN_N-     228  ----------VGLGRYLIKHGF----QNRSLLMCQHLERVLDIADK--YGF-NCQMWSDMFFKLMSADGQ  280
tr_Q06GJ0_Q06GJ0_OSTFU_N-     373  -----CWNSSDSIQNFMMQNRWDLDKESFLKLWNYFQQKAQDKAYKAFGKKLPLILWTSTLTNYKHI---  434
sp_P07686_HEXB_HUMAN_Beta     360  -----CWESNPKIQDFMRQKGFG---TDFKKLESFYIQKVLDIIAT--INK-GSIVWQEVFDDK------  412
sp_P06865_HEXA_HUMAN_Beta     328  -----CWKSNPEIQDFMRKKGFG---EDFKQLESFYIQTLLDIVSS--YGK-GYVVWQEVFDNK------  380
Consensus_ss:                           hhhhhhhhh         hhhhhhhhhhhhhhhhhhh      eeeee  hh

Conservation:                                     5  9                                  8          5
sp_P49610_STRH_STRPN_Beta     860  --------------DVQFDKDVLISYWSKGWWG----------YNLASPQYLASKGYKFLNTNGDWYYIL  905
tr_Q8DRL6_Q8DRL6_STRR6_Be     860  --------------DVQFDKDVLISYWSKGWWG----------YNLASPQYLASKGYKFLNTNGDWYYVI  905
tr_D2KW09_D2KW09_9BACL_He     360  ----------------GATSDSVLQNWTCS--------------ASTGTAAKAKGMKVIVSP-ANAYLD  397
tr_O85361_O85361_STRPL_B-     352  ----------------PVEGALVQYWGLDRTG---------DAEKAEVAEAARNGTGLILSPADRTYLD  395
tr_B3TLD6_B3TLD6_BIFBI_La     382  --------------NVSLNKDIVIEYWYGA---------------GRKPQELVQDGYTLMNAT-QALYWS  421
tr_A1RBZ5_A1RBZ5_ARTAT_Be     370  ----------------LPDGALVASWRG---------------YEGGIDALRKGYDVVMCPEHKLYLD  406
sp_Q54468_CHB_SERMA_Chito     627  --------------KAFATSRVGVNFWDTLYW----------GGFDSVNDWANKGYEVVVSNPDYVYMD  671
tr_Q840G9_Q840G9_AGGAC_Ds     225  --------------FEQINPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLAKGFTVLNYNSYYLYIV  280
tr_Q6ST21_Q6ST21_STRGN_N-     281  YDRDVEIPEETRVYLDRLKERVTLVYWDYY---------------QDSEEKYNRNFQ-NHHKISQDIAFA  334
tr_Q06GJ0_Q06GJ0_OSTFU_N-     435  -------------DDYLNKDDYIIQVWTTGV-------------DPQIKGLLEKGYRLIMSNYDALYFD  477
sp_P07686_HEXB_HUMAN_Beta     413  ---------------AKLAPGTIVEVWKDSA------------YPEELSRVTASGFPVILSAPWYLDLI  454
sp_P06865_HEXA_HUMAN_Beta     381  ---------------VKIQPDTIIQVWREDIPV----------NYMKELELVTKAGFRALLSAPWYLNRI  425
Consensus_ss:                              eeeeee                 hhhhhhhhh    eee    h eee

Conservation:                                                                    8     9 7
sp_P49610_STRH_STRPN_Beta     906  GQKPEDGGGFL---KKAI-ENTGKTPFNQLAS----------------TKYPEVDLPTVGSMLSIWADRP  955
tr_Q8DRL6_Q8DRL6_STRR6_Be     906  GNHKQDEAYPL---SKAV-ENSGKVPFNQLAS----------------TKYPEVDLPTVGSMLSIWADRP  955
tr_D2KW09_D2KW09_9BACL_He     398  MKYYSDSPIGL--QWRGFVNTNRAYNWDPTD-------------------CIKGANIYGVESTLWTETF  445
tr_O85361_O85361_STRPL_B-     396  MKYTKDTPLGL--SWAGYVEVQRSYDWDPAG----------------YLPGAPADAVRGVEAPLWTETL  446
tr_B3TLD6_B3TLD6_BIFBI_La     422  RSAQVYKVNAA-------RLYNNNWNVGTFDGG---------------RQIDKNYDKLTGAKVSIWPDSS  469
tr_A1RBZ5_A1RBZ5_ARTAT_Be     407  HRQADGDDEPV--PVGFVTTLQAVYEFEPLPG----------------VEGTDFPGRLLGAQANIWSEHL  458
sp_Q54468_CHB_SERMA_Chito     672  FPYEVNPDERGYYWGTRFSDERKVFSFAPDNMPQNAETSVDRDGNHFNAKSDKPWPGAYGLSAQLWSETQ  741
tr_Q840G9_Q840G9_AGGAC_Ds     281  PKASPTFSQDAAFAAKDVIKNWDLGVWDGRNT----------------KNRVQNTHEIAGAALSIWGEDA  334
tr_Q6ST21_Q6ST21_STRGN_N-     335  GGA---------------WKWIGFTPHNHFSRLVA-----------IEANKACRKNQVKEVIVTGWGDNG  378
tr_Q06GJ0_Q06GJ0_OSTFU_N-     478  CGYGAWVGAGNN-WCSPYIGWQKVYDNSPAV-----------------IALEHRDQVLGGEAALWSEQS  528
sp_P07686_HEXB_HUMAN_Beta     455  SYG---------------QDWRKYYKVEPLDF----------------GGTQKQKQLFIGGEACLWGEYV  493
sp_P06865_HEXA_HUMAN_Beta     426  SYG---------------PDWKDFYIVEPLAF----------------EGTPEQKALVIGGEACMWGEYV  464
Consensus_ss:                                hhh                                hhh  eeeeeeeee
```

```
Conservation:                          5         5     56  5
sp_P49610_STRH_STRPN_Beta      956  ---SAEYKEEEIFELMTAFADHNKDYFRA----------------------NYNALREELAKIPT-NLEG  999
tr_Q8DRL6_Q8DRL6_STRR6_Be      956  ---SAEYKEEEIFELMTAFADHNKDYFRA----------------------NYNALREELAKIPT-NLEG  999
tr_D2KW09_D2KW09_9BACL_He      446  V--TQDHLDYMLYPKLLSNAEVGWTARG-----------------------DRNWDDFKERLIE-HTPR  488
tr_O85361_O85361_STRPL_B-      447  S--DPDQLDYMAFPRLPGVAELGWSPAS-----------------------THDWDTYKVRLAA-QAPY  489
tr_B3TLD6_B3TLD6_BIFBI_La      470  YFQTENEVEKEIFDGMRFISQMTWSDSR-----------------------PWATWNDMKADIDK-IGYP  515
tr_A1RBZ5_A1RBZ5_ARTAT_Be      459  D--SPRRVQFAAFPRLSAISEVFWSNPA-----------------------GRDYDEFLTRLTGAHLAR  502
sp_Q54468_CHB_SERMA_Chito      742  R--TDPQMEYMIFPRALSVAERSWHRAGWEQDYRAGREYKGGETHFVDTQALEKDWLRFANILGQRELAK  809
tr_Q840G9_Q840G9_AGGAC_Ds      335  ----KALKDETIQKNTKSLLEAVIHKTN-----------------------GDE---------------  361
tr_Q6ST21_Q6ST21_STRGN_N-      379  ----GETSQFSVLPALQIWAELAYRNDL-----------------------KKVSEHFLVSTGL-DFDD  419
tr_Q06GJ0_Q06GJ0_OSTFU_N-      529  ---DTSTLDGRLWPRAAALAERLWAEPA-----------------------TSWQDAEYRMLH-IRER  569
sp_P07686_HEXB_HUMAN_Beta      494  ---DATNLTPRLWPRASAVGERLWSSKD-----------------------VRDMDDAYDRLTR-HRCR  535
sp_P06865_HEXA_HUMAN_Beta      465  ---DNTNLVPRLWPRAGAVAERLWSNKL-----------------------TSDLTFAYERLSH-FRCE  506
Consensus_ss:                          hhhhhhhhhhhhhhhhhhh                              hhhhhhhhhh hhhh

Conservation:                          5
sp_P49610_STRH_STRPN_Beta     1000  YSKESLEALDAAKTALNYNLNRNKQAELDTLVANLKAALQGLKPAVTHSGSLDENEVAANVETRPELITR 1069
tr_Q8DRL6_Q8DRL6_STRR6_Be     1000  YSKESLEALDAAKTALNYNLNRNKQAELDTLVANLKAALQGLKPAATHSGSLDENEVAANVETRPELITR 1069
tr_D2KW09_D2KW09_9BACL_He      489  LQNKGIKFFADPIVWELP-------------------------------------------------------  506
tr_O85361_O85361_STRPL_B-      490  WEAAGIDFYRSPQVPWT--------------------------------------------------------  506
tr_B3TLD6_B3TLD6_BIFBI_La      516  LDIREYDYTPVDAGIYDI-------------------------------------------------------  533
tr_A1RBZ5_A1RBZ5_ARTAT_Be      503  LEAMGVEYRPLSGPAPWQ-------------------------------------------------------  520
sp_Q54468_CHB_SERMA_Chito      810  LDKGGVAYRLPVPGARVAA------------------------------------------------------  828
tr_Q840G9_Q840G9_AGGAC_Ds           ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      420  FMKIDLANLLPD-------------------------------------------------------------  431
tr_Q06GJ0_Q06GJ0_OSTFU_N-      570  LVRMGIQAESLQPEWCYQ-------------------------------------------------------  587
sp_P07686_HEXB_HUMAN_Beta      536  MVERGIAAQPLYAGYCNH-------------------------------------------------------  553
sp_P06865_HEXA_HUMAN_Beta      507  LLRRGVQAQPLNVGFCEQ-------------------------------------------------------  524
Consensus_ss:                       hhh

Conservation:
sp_P49610_STRH_STRPN_Beta     1070  TEEIPFEVIKKENPNLPAGQENIITAGVKGERTHYISVLTENGKTTETV-------------LDS----- 1121
tr_Q8DRL6_Q8DRL6_STRR6_Be     1070  TEEIPFEVIKKENPNLPAGQENIITAGVKGERTHYISVLTENGKTTETV-------------LDS----- 1121
tr_D2KW09_D2KW09_9BACL_He      507  ----------------------------------IVQIN---------SEWKMDEGTGTVVKDTSGYLN  532
tr_O85361_O85361_STRPL_B-           ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      534  ----------------------------------PQLKS---------ISKGPWELITTPDGYYQMKDT  559
tr_A1RBZ5_A1RBZ5_ARTAT_Be      521  --------------------------------QRPGVEGWKRDYDAEQLVRN----------------  540
sp_Q54468_CHB_SERMA_Chito      829  ------------------------------GKLEANIALPGLGIEYSTDGGK------------------  850
tr_Q840G9_Q840G9_AGGAC_Ds           ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      432  ----------------------------------LPDNL---------SGINPNRYVLYQDVL------  451
tr_Q06GJ0_Q06GJ0_OSTFU_N-      588  ----------------------------------NEGYCYS----------------------------  594
sp_P07686_HEXB_HUMAN_Beta      554  ----------------------------------ENM-------------------------------  556
sp_P06865_HEXA_HUMAN_Beta      525  ----------------------------------EFEQT------------------------------  529
Consensus_ss:

Conservation:
sp_P49610_STRH_STRPN_Beta           ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be           ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He      533  GTLVGGA-----------------------------------------------------------------  539
tr_O85361_O85361_STRPL_B-           ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      560  VSGKCLALFTGSKHLDVVTQVGARPELRNCADVSVGQDQRNTANERNTQKWQIRADKDGKYTISPALTQQ  629
tr_A1RBZ5_A1RBZ5_ARTAT_Be           ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito           ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds           ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-           ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-           ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta           ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta           ------------------------------------------------------------------------
Consensus_ss:

Conservation:
sp_P49610_STRH_STRPN_Beta           ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be           ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He           ------------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-           ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      630  RLAIATGNEQNIDLETHRPAAGTVAQFPADLVSDNALFTLTGHMGMSATVDSKTVNPASPSKITVKVRAA  699
tr_A1RBZ5_A1RBZ5_ARTAT_Be           ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito           ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds           ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-           ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-           ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta           ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta           ------------------------------------------------------------------------
Consensus_ss:

Conservation:
sp_P49610_STRH_STRPN_Beta     1122  ------------------------------------------------------------QVTK---- 1125
tr_Q8DRL6_Q8DRL6_STRR6_Be     1122  ------------------------------------------------------------QVTK---- 1125
tr_D2KW09_D2KW09_9BACL_He      540  -------------------------------------------------KWTAGKQGNGVSFDGSSGYIN  560
```

```
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      700 SNANTGDVTVTPVVPEGWEIKPGSVSLKSIPAGKAAIAYFNVVNTTGTGDATVQFKLTNTKTGEELGTTS  769
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito      851 ------------------------------------------------------------QWQRYDAK  858
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1126 EVINQVVEVGAPVTHKGDESGLAPTTEVKPRLDIQEEEIPFTTVTCENPLLLK--------------GKT 1181
tr_Q8DRL6_Q8DRL6_STRR6_Be     1126 EVINQVVEVGSPVTHKGDESGLAPTTEVKPRLDIQEEEIPFTTVTRENPLLLK--------------GKT 1181
tr_D2KW09_D2KW09_9BACL_He      561 LGGQDITGNWTAAVWVYGQPNTTNNETLLSGTTSAIKINQYNKTGKVGITIYGTKDYTYNYSIPSNKWTH  630
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      770 VALTGSLTKDVEASDYAASSQETTGEHAPVGNAFDKNANTFWHSKYSNPSANL--------------PHW  825
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1182 QVITK-------------------------------------GVNGHRSNFYSVSTSADGKEV------ 1207
tr_Q8DRL6_Q8DRL6_STRR6_Be     1182 QVITK-------------------------------------GVNGHRSNFYSVSTSADGKEV------ 1207
tr_D2KW09_D2KW09_9BACL_He      631 LTFVGTSTGTALYENGVLKETIAAKMNGPMALVGAEKTGGSGDLTSYFRGSLDELKIFNRALS-------  693
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      826 LAFKASPGEGNK----------------------IAAITHLYRQDKLNGPAKNVAVYVVAASDANSVAD  872
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      452 ------------------------------------------------------CPLLEQHIR-------  460
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta          ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be          ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He      694 --ASEVVELAKS-------------------------------------------------------  703
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      873 VTNWGEPVATAEFPYTKELQTIALPNTIPSGDVYVKFQINDAWGLTETSAGVTWAAVAELAATAKATPVE  942
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      461 --PEKDKQHFAS-------------------------------------------------------  470
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta          ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be          ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He          ------------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La      943 LTEPEQPKDNPEVTETPEATGVTVSGDGVANGALSLKKGTTAQLTAKVAPDDADQAVTWASSDDKVVTVD 1012
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta          ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be          ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He      704 ----PAPKASLTGPQSANPGQSFDVKMGLSDVSPSEFGQMYAQDWTINYDSAKLQLDSITSLQDKFQVID  769
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La     1013 KTGKVTAVAKGVAKVTATTANGKSASVTVTV----------------------------------- 1043
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito      859 ----AKPAVSGEVQVRSVSPDGKRYSRAEKV-----------------------------------  885
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
```

```
tr_Q6ST21_Q6ST21_STRGN_N-     471  ----SAQQLGEISKRAGEYAYIFETQAQL----------------------------------------  495
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1208  KTLVNSVVAQEAVTQIVEVGTMVTHVGDENGQAAIAEEKPKLEIPSQPAPSTAPAEESKVLPQDPAPVVT 1277
tr_Q8DRL6_Q8DRL6_STRR6_Be     1208  KTLVNSVVAQEAVTQIVEVGTMVTHVGDENGQAAIAEEKPKLEIPSQPAPSTAPAEESKALPQDPAPVVT 1277
tr_D2KW09_D2KW09_9BACL_He      770  QKELAPGQIRIVAANAAANQGVTPQGDLFAFKFTVKAGTDVKTTISADHIVIGNAQGKELEIAGATHEIQ  839
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La     1044  TEDSEVPGPTGPTEPTKPGTEKPTTKPTTKPNDGKLSAT------------------------------ 1082
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-          ------------------------------------------------------------------------
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1278  EKKLPETGTHDSAGLV------------------------------------------------------ 1293
tr_Q8DRL6_Q8DRL6_STRR6_Be     1278  EKKLPETGTHDSAELV------------------------------------------------------ 1293
tr_D2KW09_D2KW09_9BACL_He      840  VSIPVDKSQLNVLIANAQAKHDAAVEGNEDGLYAAGSKAQLQTAIHTAKAVADNSNASQQQVDSAKSALE  909
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La     1083  ----------GADTAV------------------------------------------------------ 1088
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      496  -------------NAL----------------------------------------------------------  498
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1294  ------------------------------------------------------VAGLMSTLAAY----- 1304
tr_Q8DRL6_Q8DRL6_STRR6_Be     1294  ------------------------------------------------------VAGLMSTLAAY----- 1304
tr_D2KW09_D2KW09_9BACL_He      910  EAVQVFESKKISADVNGDGQVSIGDLAIIAGAYGKEEGQAGWNKKADVNHDGKVDIIDLTIVAKA-----  974
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La     1089  ------------------------------------------------------LATIAALFALA----- 1099
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      499  ------------------------------------------------------LALKISITSGIQKAYR  514
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta          ------------------------------------------------------------------------
tr_Q8DRL6_Q8DRL6_STRR6_Be          ------------------------------------------------------------------------
tr_D2KW09_D2KW09_9BACL_He          ------------------------------------------------------------------------
tr_O85361_O85361_STRPL_B-          ------------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La          ------------------------------------------------------------------------
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      515  NGDKEHLSALAEKDFPQLYQMVEDFSDQFSRQWQQENKIFGLDTIDIRFGGLLKRIKRAQERLEQFISGQ  584
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------------------------------------
Consensus_ss:


Conservation:
sp_P49610_STRH_STRPN_Beta     1305  ----------GLTK---------RKED---------------- 1312
tr_Q8DRL6_Q8DRL6_STRR6_Be     1305  ----------GLTK---------RKED---------------- 1312
tr_D2KW09_D2KW09_9BACL_He      975  ----------ILQI-----------------------------  978
tr_O85361_O85361_STRPL_B-          ------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La     1100  ----------GGAVV------AVRRRSVR-------------- 1112
tr_A1RBZ5_A1RBZ5_ARTAT_Be          ------------------------------------------
sp_Q54468_CHB_SERMA_Chito          ------------------------------------------
tr_Q840G9_Q840G9_AGGAC_Ds          ------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-      585  IDCVEELEQEILPFNDFYKDQGLTATTANQWHLIATASTIYTT  627
tr_Q06GJ0_Q06GJ0_OSTFU_N-          ------------------------------------------
sp_P07686_HEXB_HUMAN_Beta          ------------------------------------------
sp_P06865_HEXA_HUMAN_Beta          ------------------------------------------
```

Figure A3.2. Promals alignment of LnbB and crystallized GH20 family proteins.
B3TLD6: *B. bifidum* Lacto-N-biosidase (LnbB)
O85361: *S. plicatus* B-N-acetylhexosaminidase (HexSp)
Q5446Z: *Serratia marcenesis* Chitobiase
P0686: HEXA_HUMAN Beta-hexosaminidase subunit alpha
P0768: HEXB_HUMAN Beta-hexosaminidase subunit beta
Q6ST21: *Streptococcus gordonii* N-acetyl-beta-D-glucosaminidase
A1RBZ5: *Arthrobacter aurescens* Beta-N-acetylhexosaminidase
D2KW09: *Paenibacillus sp*. Hexosaminidase Hex1t
Q840G9: *Aggregatibacter actinomycetemcomitans* DspB (Fragment)
Q8DRL6: *Streptococcus pneumoniae* (strain R6)Beta-N-acetyl-hexosaminidase
P4961: *Streptococcus pneumoniae* serotype 4 (TIGR4)Beta-N-acetylhexosaminidase

```
Conservation:                          9
tr_Q840G9_Q840G9_AGGAC_Ds              ------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito        1     MN-----------------AFKLSALARLTATMGFLGGMGSAMADQQLVDQLSQLKLNVKMLDNRAGEN    52
tr_O85361_O85361_STRPL_B-        1     MT-------------------------------------------------------------------     2
tr_B3TLD6_B3TLD6_BIFBI_La        1     MEK--------SSNR-----RFGVRTVAAIVAGLMVGGMCTA-------------------------    29
sp_P07686_HEXB_HUMAN_Beta        1     MELCGLGLPRPPMLLALLLATLLAAMLALLTQVALVVQVA---------------------------    40
sp_P06865_HEXA_HUMAN_Beta        1     MT----------------------SSRLWFSLLLAA------------------------------    14
tr_Q6ST21_Q6ST21_STRGN_N-        1     MA-------------------------------------------------------------------     2
Consensus_ss:                                                  hhhh

Conservation:
tr_Q840G9_Q840G9_AGGAC_Ds              ------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito       53     GVDCAALGADWASCNRVLFTLSNDGQAIDGKDWVIYFHSPRQTLRVDNDQFKIAHLTGDLYKLEPTAKFS   122
tr_O85361_O85361_STRPL_B-              ------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La              ------------------------------------------------------------------
sp_P07686_HEXB_HUMAN_Beta              ------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta              ------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-              ------------------------------------------------------------------
Consensus_ss:

Conservation:
tr_Q840G9_Q840G9_AGGAC_Ds              ------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito      123     GFPAGKAVEIPVVAEYWQLFRNDFLPRWYATSGDAKPKMLANTDTENLDQFVAPFTGDQWKRTKDDKNIL   192
tr_O85361_O85361_STRPL_B-              ------------------------------------------------------------------
tr_B3TLD6_B3TLD6_BIFBI_La       30     ------------------------------------------------------------------M    30
sp_P07686_HEXB_HUMAN_Beta              ------------------------------------------------------------------
sp_P06865_HEXA_HUMAN_Beta              ------------------------------------------------------------------
tr_Q6ST21_Q6ST21_STRGN_N-              ------------------------------------------------------------------
Consensus_ss:

Conservation:                                                  5
tr_Q840G9_Q840G9_AGGAC_Ds              ------------------------------------------------------------------
sp_Q54468_CHB_SERMA_Chito      193     MTPASRFVSNADLQTLPAGALRGKIVPTPMQVKVHAQDADLRKGVALDLST----------LVKPAADVV   252
tr_O85361_O85361_STRPL_B-        3     ------TGAAPDRKAPVRPTPLDRVIPAPASVDPGGAPYRITRGTHIRVDDSR--------EARRVGDYL    58
tr_B3TLD6_B3TLD6_BIFBI_La       31     TASAADDSAAGYSATAPVNLTRPATVPSMDGWTDGTGAWTLGEGTRVVSSD----------ALAARAQSL    90
sp_P07686_HEXB_HUMAN_Beta       41     ---------EAARAPSVSAKPGPALWPLPLSVKMTPNLLHLAPENFYISHSPNS-------TAGPSCTLL    94
sp_P06865_HEXA_HUMAN_Beta       15     ----------------AFAGRATALWPWPQNFQTSDQRYVLYPNNFQFQYDVSSAAQPGCSVLDEAFQRY    68
tr_Q6ST21_Q6ST21_STRGN_N-        3     --------------------------------TFLGLSSKQEKALVRLDKYLN-----------------    23
Consensus_ss:                                                eeee    eeee    eeeee            hhhhhhhhh

Conservation:                                                                  5 5 7        5
tr_Q840G9_Q840G9_AGGAC_Ds        1     ------------------------------------------------------------NCCVKGN-     7
sp_Q54468_CHB_SERMA_Chito      253     SQRFALLGVP-------VQTNGYP--------IKTDIQPG---KFKGAMAVSGAYELKIGKKEAQVIGFD   304
tr_O85361_O85361_STRPL_B-       59     ADLLRPATGY-------RLPVTAHG------HGGIRLRLA------GGPYGDEGYRLDSGPAGVTITARK   109
tr_B3TLD6_B3TLD6_BIFBI_La       91     ASELTKFTDV-------DIKAATGS----ATGKDISLTLD---ASKKAELGDEGFKLNIGSKGLEVIGAT   146
sp_P07686_HEXB_HUMAN_Beta       95     EEAFRRYHGYIFGFYKWHHEPAEFQAKTQVQQLLVSITLQSECDAFPNISSDESYTLLVKEPVAVLKANR   164
sp_P06865_HEXA_HUMAN_Beta       69     RDLLFGSGSW-------PRPYLTGKRHTLEKNVLVVSVVTPGCNQLPTLESVENYTLTINDDQCLLLSET   131
tr_Q6ST21_Q6ST21_STRGN_N-       24     ----------------------------LGEIAVSLVTD----------SATSIKVEGRQGYYQVSYKQ    54
Consensus_ss:                          hhhhhhh                        eeeeeee              eeeee   eeeeee

Conservation:                            7   75   5556                      5    55  57  7           65  7
tr_Q840G9_Q840G9_AGGAC_Ds        8     --------------------------------SIYPQKTSTKQTGLMLDIARH-FYSPEVIKSFIDTIS    43
sp_Q54468_CHB_SERMA_Chito      305     QAGVFYGLQSILSLVPSDG--------SGKIATLDASDAPRFPYRGIFLDVARN-FHKKDAVLRLLDQMA   365
tr_O85361_O85361_STRPL_B-      110     AAGLFHGVQTLRQLLPPAVEKDSAQPGPWLVAGGTIEDTPRYAWRSAMLDVSRH-FFGVDEVKRYIDRVA   178
tr_B3TLD6_B3TLD6_BIFBI_La      147     DIGVFYGTRSVSQMLRQG---------QLTLPAGTVATKPKYKERGATLCACQI-NISTDWIDRFLSDMA   206
sp_P07686_HEXB_HUMAN_Beta      165     VWGALRGLETFSQLVYQDSY------GTFTINESTIIDSPRFSHRGILIDTSRH-YLPVKIILKTLDAMA   227
sp_P06865_HEXA_HUMAN_Beta      132     VWGALRGLETFSQLVWKSAE------GTFFINKTEIEDFPRFPHRGLLLDTSRH-YLPLSSILDTLDVMA   194
tr_Q6ST21_Q6ST21_STRGN_N-       55     PHQLYRALALLSAALRS-----------GQDEVQIEEEAAYEDLAYMADCSRNAVLNLSSAKKMIEVLA   112
Consensus_ss:                          hhhhhhhhhhhhhhh              eeeeeeeee       eeeee        hhhhhhhhhhhh

Conservation:                               7  555  7           55   5                        56         7
tr_Q840G9_Q840G9_AGGAC_Ds       44     LSGGNFLHLHFSDHENYAIESHLLNQRAENAVQGKDGIYI----------NPYTGKPFLSYRQLDDIKAY   103
sp_Q54468_CHB_SERMA_Chito      366     AYKLNKFHFHLSDDEGWRIEIPGLPELTEVGGQRCHDLSETTCLLPQYGQGPDVYGGFFSRQDYIDIIKY   435
tr_O85361_O85361_STRPL_B-      179     RYKYNKLHLHLSDDQGWRIAIDSWPRLATYGGSTEVG---------------GGPGGYYTKAEYKEIVRY   233
tr_B3TLD6_B3TLD6_BIFBI_La      207     DLRLNYVLLEMKLKP----------EEDNTKKAA--------------------TWSYYTRDDVKKFVKK   246
sp_P07686_HEXB_HUMAN_Beta      228     FNKFNVLHWHIVDDQSFPYQSITFPELSNKGSYS--------------------LSHVYTPNDVRMVIEY   277
sp_P06865_HEXA_HUMAN_Beta      195     YNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYNP------------------VTHIYTAQDVKEVIEY   245
tr_Q6ST21_Q6ST21_STRGN_N-      113     LMGYSTFELYMEDTYEIENQPY---------FGY--------------------FRGRYTVAELQEIEDY   153
Consensus_ss:                          h    eeeeeee             hhhhh                              hhhhhhhhhh

Conservation:                          9    5 5 77 5 7 9
tr_Q840G9_Q840G9_AGGAC_Ds      104     AKAKGIELIPELDSPNHMTAIFKLVQKDR----GVKYL-----------------QGLKSRQVDDEIDIT   152
sp_Q54468_CHB_SERMA_Chito      436     AQARQIEVIPEIDMPAHARAAVVSMEARYKKLHAAGKEQEANEFRLVDPTDTSNTTSVQFFNRQSYLNPC   505
tr_O85361_O85361_STRPL_B-      234     AASRHLEVVPEIDMPGHTNAALASYAELN----CDGVAP---------------PLYTGTKVGFSSLCVD   284
tr_B3TLD6_B3TLD6_BIFBI_La      247     ANNYGIDVIPEINSPGHMNVWLENYPEYQ----LADN--------------------SGRKDPNKLDIS   291
sp_P07686_HEXB_HUMAN_Beta      278     ARLRGIRVLPEFDTPGHTLSWGKGQKDLL----TPCYS------------------RQNKLDSFGPINPT   325
sp_P06865_HEXA_HUMAN_Beta      246     ARLRGIRVLAEFDTPGHTLSWGPGIPGLL----TPCYS------------------GSEPSGTFGPVNPS   293
```

```
tr_Q6ST21_Q6ST21_STRGN_N-     154  AADFDMSFVPCIQTLAHLSAFVKWGIKEV----------------------------QELRDVEDILLIG  195
Consensus_ss:                      hhh   eeee    hhhhhhhhhh hhh

Conservation:                            7     7              7 9599
tr_Q840G9_Q840G9_AGGAC_Ds     153  NADSITFMQSLMSEVIDIFGD---TSQHFHIGGDEFGYS-------------------------VES--  191
sp_Q54468_CHB_SERMA_Chito     506  LDSSQRFVDKVIGEIAQMHKEAGQPIKTWHFGGDEAKNIRLGAGYTDKAKPEPGKGIIDQSNEDKPWAKS  575
tr_O85361_O85361_STRPL_B-     285  KDVTYDFVDDVIGELAALTP-----GRYLHIGGDEAHS-------------------------------  317
tr_B3TLD6_B3TLD6_BIFBI_La     292  NPEAVKFYKTLIDEYDGVFT-----TKYWHMGADEYMIG-----------------------TSFDN  330
sp_P07686_HEXB_HUMAN_Beta     326  LNTTYSFLTTFFKEISEVFP-----DQFIHLGGDEVEFK-----------------------CWESN  364
sp_P06865_HEXA_HUMAN_Beta     294  LNNTYEFMSTFFLEVSSVFP-----DFYLHLGGDEVDFT-----------------------CWKSN  332
tr_Q6ST21_Q6ST21_STRGN_N-     196  EEKVYDLIEGMFQTMAHLHT------RKINIGMDEAHLV-------------------------------  228
Consensus_ss:                       hhhhhhhhhhhhhhhhh         eeeee

Conservation:                                                 7      9 5
tr_Q840G9_Q840G9_AGGAC_Ds     192  ------------------NHEFITYANKLSYFLEKKGL-KTRMWNDGLIKNT------------------  224
sp_Q54468_CHB_SERMA_Chito     576  QVCQTMIKEGKV---ADMEHLPSYFGQEVSKLVKAHGIDRMQAWQDGLKDAESS----------------  626
tr_O85361_O85361_STRPL_B-     318  ----------------TPKADFVAFMKRVQPIVAKYGK-TVVGWHQLAGA--------------------  350
tr_B3TLD6_B3TLD6_BIFBI_La     331  YSKLKTFAEKQYGAGATPNDAFTGFINDIDKYVKAKGK-QLRIWNDGIVNTK------------------  381
sp_P07686_HEXB_HUMAN_Beta     365  PKIQDFMRQKGFG--TDFKKLESFYIQKVLDIIATINK-GSIVWQEVFDDK------------------  412
sp_P06865_HEXA_HUMAN_Beta     333  PEIQDFMRKKGFG--EDFKQLESFYIQTLLDIVSSYGK-GYVVWQEVFDNK------------------  380
tr_Q6ST21_Q6ST21_STRGN_N-     229  -GLGRYLIKHGF---QNRSLLMCQHLERVLDIADKYGF-NCQMWSDMFFKLMSADGQYDRDVEIPEETRV  293
Consensus_ss:                       hhhhhhhhh        hhhhhhhhhhhhhhhhhh    eeeeee hh

Conservation:                                59                              7
tr_Q840G9_Q840G9_AGGAC_Ds     225  -FEQINPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLA---KGFTVLNYNSYYLYIVPKASPTFSQD  290
sp_Q54468_CHB_SERMA_Chito     627  -KAFATSRVGVNFWDTLY----------WGGFDSVNDWAN---KGYEVVVSNPDYVYMDFPYEVNPDER  681
tr_O85361_O85361_STRPL_B-     351  ---EPVEGALVQYWGLDRT---------GDAEKAEVAEAAR---NGTGLILSPADRTYLDMKYTKDTPLG  405
tr_B3TLD6_B3TLD6_BIFBI_La     382  -NVSLNKDIVIEYWYGA---------------GRKPQELVQ---DGYTLMNAT-QALYWSRSAQVYKVNA  431
sp_P07686_HEXB_HUMAN_Beta     413  --AKLAPGTIVEVWKDSA------------YPEELSRVTA---SGFPVILSAPWYLDLISYG-------  457
sp_P06865_HEXA_HUMAN_Beta     381  --VKIQPDTIIQVWREDI----------PVNYMKELELVTK---AGFRALLSAPWYLNRISYG-------  428
tr_Q6ST21_Q6ST21_STRGN_N-     294  YLDRLKERVTLVYWDYYQ----------DSEEKYNRNFQNHHKISQDIAFAGGAWKWIGFTP-------  345
Consensus_ss:                             eeeeee                  hhhhhhhhh       eee  hhh

Conservation:                                  6                              7     9 6
tr_Q840G9_Q840G9_AGGAC_Ds     291  AAFAAKDVIKNWDLGVWDGRNT---------------KNRVQNTHEIAGAALSIWGEDA----KALKDE  340
sp_Q54468_CHB_SERMA_Chito     682  GYYWGTRFSDERKVFSFAPDNMPQNAETSVDRDGNHFNAKSDKPWPGAYGLSAQLWSETQR--TDPQMEY  749
tr_O85361_O85361_STRPL_B-     406  LSW--AGYVEVQRSYDWDPAG----------------YLPGAPADAVRGVEAPLWTETLS--DPDQLDY  454
tr_B3TLD6_B3TLD6_BIFBI_La     432  ARLY----NNNWNVGTFDGGR----------------QIDKNYDKLTGAKVSIWPDSSYFQTENEVEK  479
sp_P07686_HEXB_HUMAN_Beta     458  --------QDWRKYYKVEPLDF---------------GGTQKQKQLFIGGEACLWGEYV---DATNLTP  500
sp_P06865_HEXA_HUMAN_Beta     429  --------PDWKDFYIVEPLAF---------------EGTPEQKALVIGGEACMWGEYV---DNTNLVP  471
tr_Q6ST21_Q6ST21_STRGN_N-     346  -----HNHFSRLVAIE---------------------ANKACRKNQVKEVIVTGWGDNG----GETSQF  384
Consensus_ss:                             hhhh                          hhhh  eeeeeeeee        hhhhhh

Conservation:                               5      7  5
tr_Q840G9_Q840G9_AGGAC_Ds     341  TIQKNTKSLLEAVIHKTNG-DE-------------------------------------  361
sp_Q54468_CHB_SERMA_Chito     750  MIFPRALSVAERSWHRAGW-EQDYRAGREYKGG-------ETH----------------  784
tr_O85361_O85361_STRPL_B-     455  MAFPRLPGVAELGWSPAST-HDWDTYKVRLAAQAPYWEAAGIDFYRSPQVPWT-------  506
tr_B3TLD6_B3TLD6_BIFBI_La     480  EIFDGMRFISQMTWSDSRPWATWNDMKADIDKIGYPLDIREYDYTPVD-AGIYDI-----  533
sp_P07686_HEXB_HUMAN_Beta     501  RLWPRASAVGERLWSSKDV-RDMDDAYDRLTRHRCRMVERGIAAQPLY-AGYCNHENM--  556
sp_P06865_HEXA_HUMAN_Beta     472  RLWPRAGAVAERLWSNKLT-SDLTFAYERLSHFRCELLRRGVQAQPLN-VGFCEQEFEQT  529
tr_Q6ST21_Q6ST21_STRGN_N-     385  SVLPALQIWAELAYRNDLK-KVSEHFLVSTGLDFDDFMKIDLANLLPDLP----------  433
Consensus_ss:                       hhhhhhhhhhhhhh        hhhhhhhhhhhhhhhhhh
```

Figure A3.3. Promals alignment of the GH20 domains of LnbB and crystallized GH20 family proteins.
B3TLD6: *B. bifidum* Lacto-N-biosidase (LnbB)
O85361: *S. plicatus* B-N-acetylhexosaminidase (HexSp)
Q5446Z: *Serratia marcenesis* Chitobiase
P0686: HEXA_HUMAN Beta-hexosaminidase subunit alpha
P0768: HEXB_HUMAN Beta-hexosaminidase subunit beta
Q6ST21: *Streptococcus gordonii* N-acetyl-beta-D-glucosaminidase
Q840G9: *Aggregatibacter actinomycetemcomitans* DspB (Fragment)

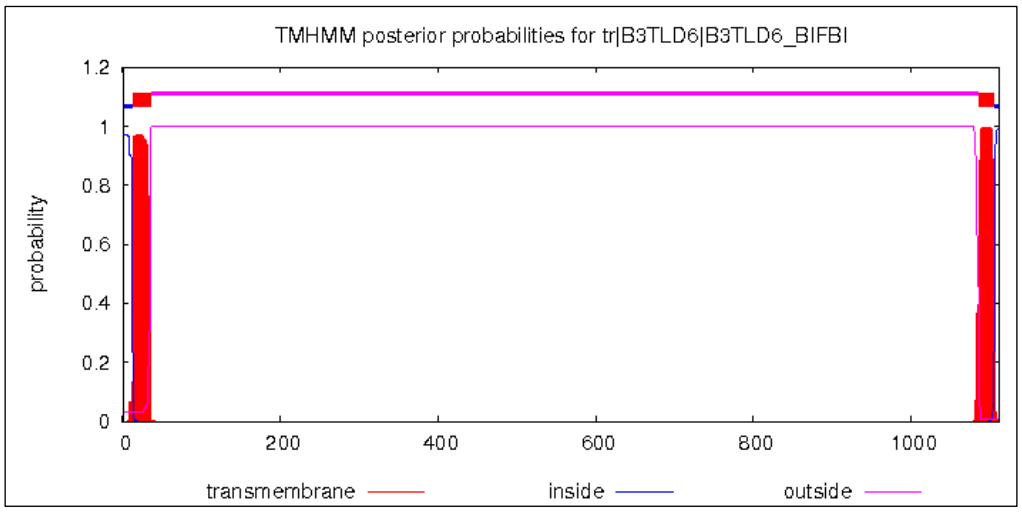


Figure A3.4. Prediction of transmembrane helices in proteins using TMHMM Server v. 2.0.

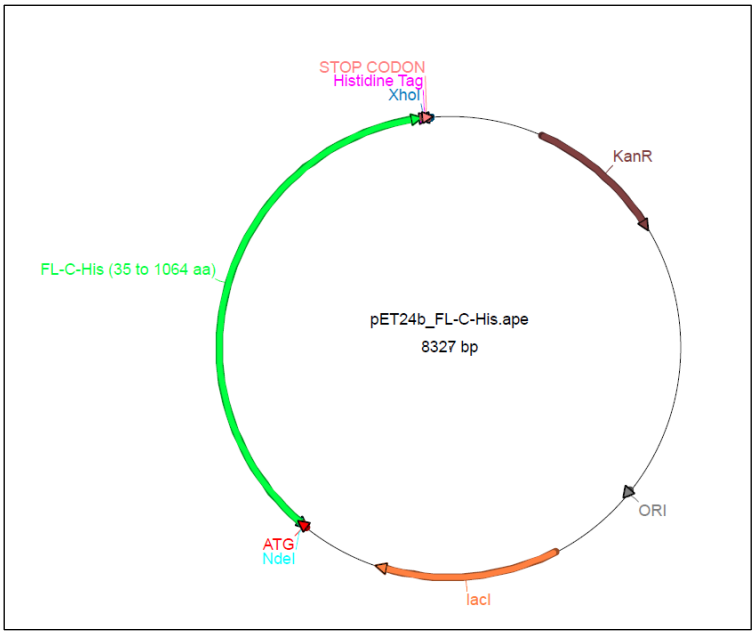


Figure A3.5. Graphic map of pET24b vector with LnbB C-terminal His-tagged protein.

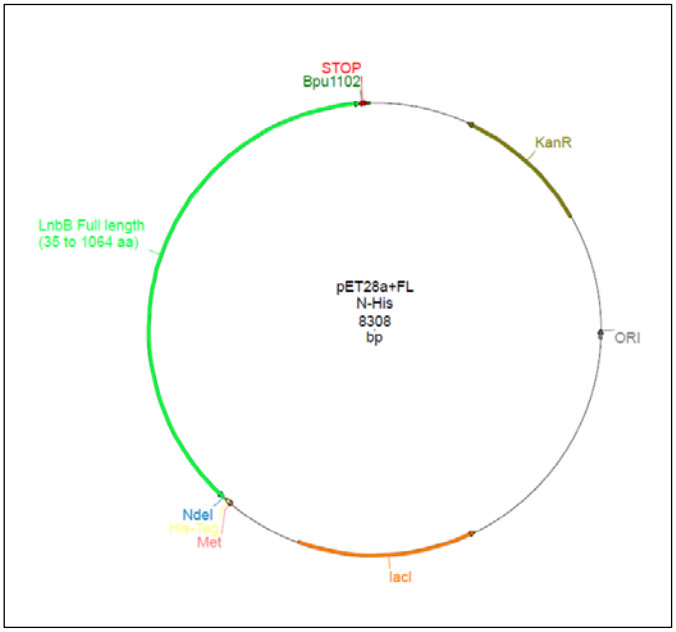


Figure A3.6. Graphic map of pET28a+ vector with LnbB N-terminal His-tagged protein.

APPENDIX 4

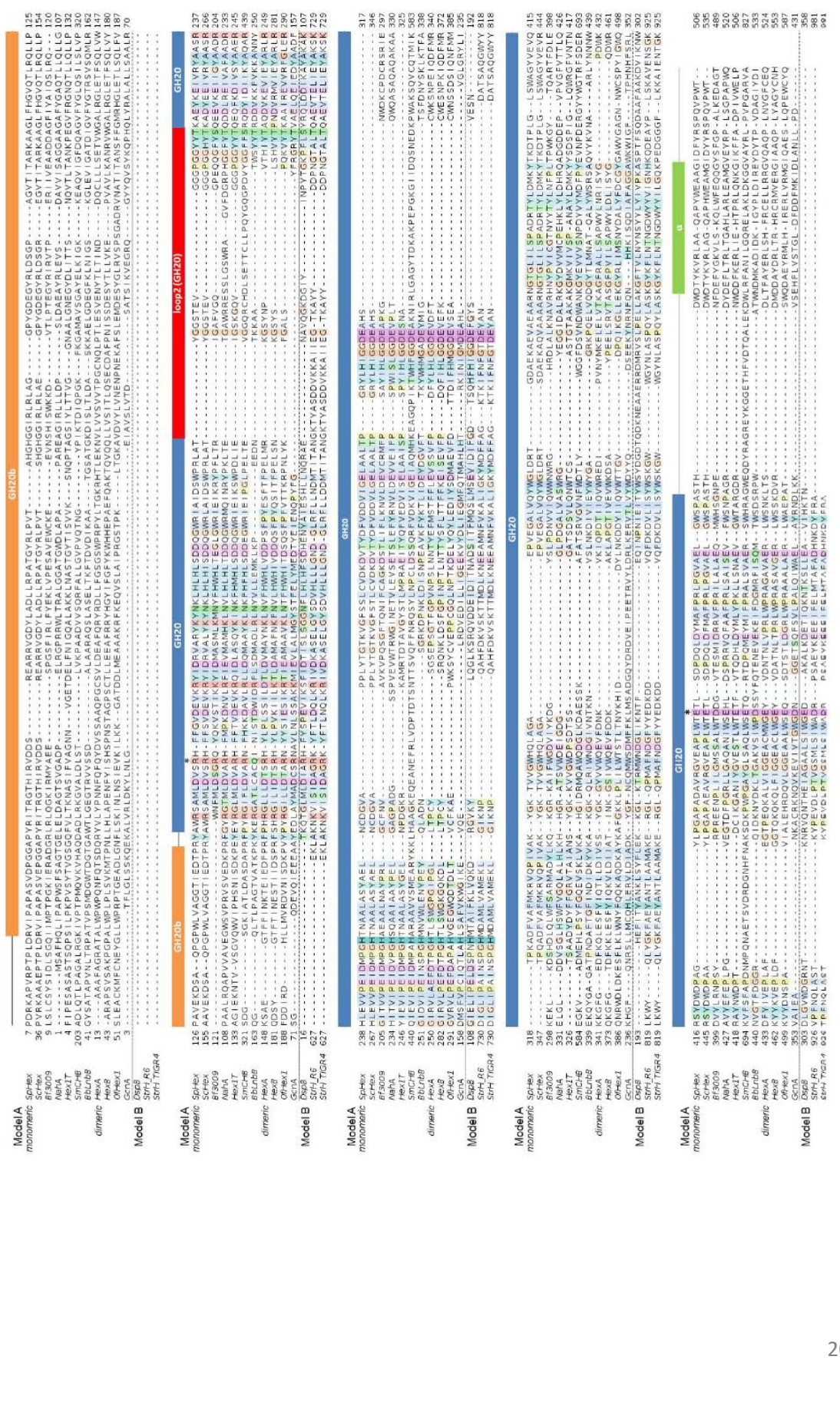


Figure A4.1. Multiple sequence alignment of β-*N*-acetylhexosaminidases of known structure from family GH20. Domains distribution is shown as coloured boxes on top. Accompanying domains are not shown. Conserved arginine and glutamate/aspartate sites are marked with a star. Sequences were downloaded from UniProt entry names: SpHex (O85361_STRPL), ScHex (Q9L068), Bf3009 (Q5LAT3) NahA (A1RBZ5_ARTAT), Hex1T (D2KW09_9BACL), SmCHB (CHB_SERMA), LnbB B3TLD6_BIFBI), HexA (HEXA_HUMAN), HexB (HEXB_HUMAN), OfHex1 (Q06GJ0_OSTFU), GcnA (Q6ST21_STRGN), DspB (Q840G9_AGGAC), StrH_R6 (Q8DRL6_STRR6), StrH_TIGR4 (STRH_STRPN). Alignment was performed with PROMALS. Conserved positions are coloured according to ClustalW colour scheme. Only the sequence fragments covering GH20b, GH20 domains and extra α-helix were considered in this alignment.

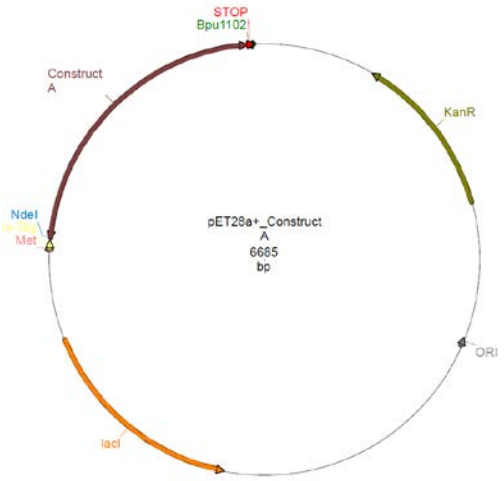


Figure A4.2. Graphic map of pET28a+ vector with Construct A N-terminal His-tagged protein.

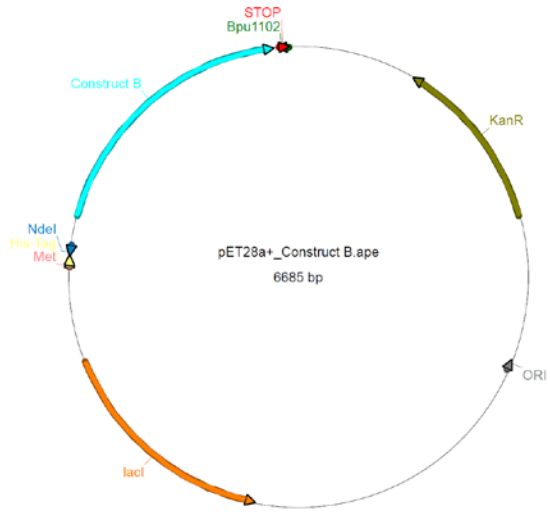


Figure A4.3. Graphic map of pET28a+ vector with Construct B N-terminal His-tagged protein.

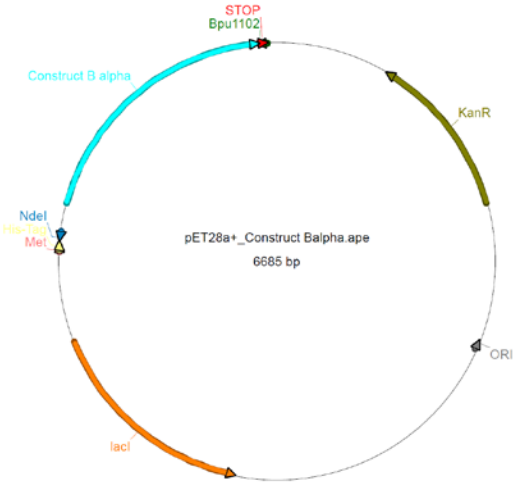


Figure A4.4 Graphic map of pET28a+ vector with Construct Bα N-terminal His-tagged protein.

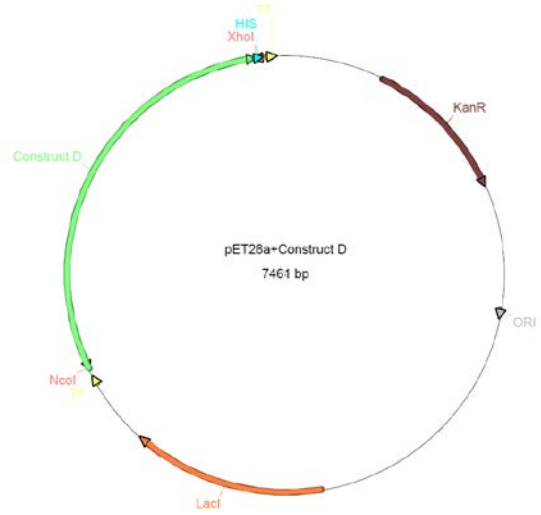


Figure A4.5 Graphic map of pET28a+ vector with Construct D C-terminal His-tagged protein.

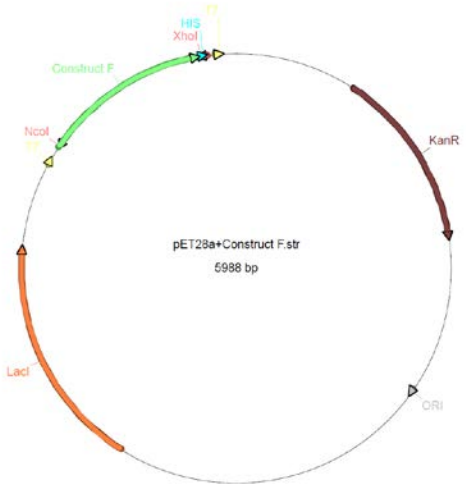


Figure A4.6 Graphic map of pET28a+ vector with Construct F C-terminal His-tagged protein.

PLOS ONE

RESEARCH ARTICLE

Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases

Cristina Val-Cid, Xevi Biarnés, Magda Faijes*, Antoni Planas

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain

* magda.faijes@iqs.edu

Citation: Val-Cid C, Biarnés X, Faijes M, Planas A (2015) Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases. PLoS ONE 10(5): e0128075. doi:10.1371/journal.pone.0128075

Academic Editor: Elena Papaleo, University of Copenhagen, DENMARK

Received: July 2, 2014

Accepted: April 23, 2015

Published: May 29, 2015

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported in part by grants NANO3BIO (FP7-KBBE-2013-7; # 613931) from the European Commission and GLYCOZYMES (BIO2013-49022-C2-1-R) from MEC, Spain. C. Val-Cid acknowledges financial support from the Fundación Instituto Danone (2012-2013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Hexosaminidases are involved in important biological processes catalyzing the hydrolysis of N-acetyl-hexosaminyl residues in glycosaminoglycans and glycoconjugates. The GH20 enzymes present diverse domain organizations for which we propose two minimal model architectures: Model A containing at least a non-catalytic GH20b domain and the catalytic one (GH20) always accompanied with an extra α-helix (GH20b-GH20-α), and Model B with only the catalytic GH20 domain. The large *Bifidobacterium bifidum* lacto-N-biosidase was used as a model protein to evaluate the minimal functional unit due to its interest and structural complexity. By expressing different truncated forms of this enzyme, we show that Model A architectures cannot be reduced to Model B. In particular, there are two structural requirements general to GH20 enzymes with Model A architecture. First, the non-catalytic domain GH20b at the N-terminus of the catalytic GH20 domain is required for expression and seems to stabilize it. Second, the substrate-binding cavity at the GH20 domain always involves a remote element provided by a long loop from the catalytic domain itself or, when this loop is short, by an element from another domain of the multidomain structure or from the dimeric partner. Particularly, the lacto-N-biosidase requires GH20b and the lectin-like domain at the N- and C-termini of the catalytic GH20 domain to be fully soluble and functional. The lectin domain provides this remote element to the active site. We demonstrate restoration of activity of the inactive GH20b-GH20-α construct (model A architecture) by a complementation assay with the lectin-like domain. The engineering of minimal functional units of multidomain GH20 enzymes must consider these structural requirements.

Introduction

N-acetyl-β-hexosaminidases are widely distributed in nature participating in many biological processes [1,2]. They catalyze the hydrolysis of N-acetyl-hexosaminyl residues in glycoproteins, glycosaminoglycans, chitin, and glycolipids, and are classified in CAZY families GH3, GH20 and GH84 (Carbohydrate Active Enzymes database, www.cazy.org [3]). GH20 hexosaminidases are found in crustaceans, insects and fungi, participating in the degradation of chitin, and in

Competing Interests: The authors have declared that no competing interests exist.

bacteria, where they are involved in the catabolism of oligosaccharides that serve as nitrogen and carbon sources. From the biotechnological point of view, hexosaminidases, together with chitinases, are important enzymes for industrial chitin degradation, the most abundant polysaccharide after cellulose [2]. They are also powerful insect and fungal control agents as biopesticides. In humans, the two lysosomal hexosaminidases HexA and HexB are responsible for the hydrolysis of GlcNAc and GalNAc residues of glycosphingolipids such as ganglioside GM2 and their dysfunction is linked to neurodegenerative disorders known as Tay-Sachs and Sandhoff diseases [4,5]. In addition, some bacterial hexosaminidases contribute to the virulence of some infections in humans as in the case of *Actinobacillus actinomycetemcomitans* that colonizes the human oral cavity and causes periodontitis and *Streptococcus pneumoniae* infections [2].

GH20 enzymes comprise not only β-N-acetylhexosaminidases (EC.3.2.1.52), catalyzing the removal of terminal non-reducing N-acetylhexosamine residues, but also lacto-N-biosidases (EC.3.2.1.140), hydrolyzing the terminal lacto-N-biosyl residues from the non-reducing end of oligosaccharides such as human milk oligosaccharides (HMOs). HMOs play an important role in the development of intestinal flora and in the modulation of the immune system of breast-fed infants [6–8]. HMOs are constituted by two core tetrasaccharide structures: the lacto-N-tetraose (type 1) characterized by lacto-N-biose (Galβ1,3GlcNAc) and lactose units linked by a β-1,3 glycosidic bond, and lacto-neo-N-tetraose (type 2) formed by β-1,3-linked N-acetyllactosamine (Galβ1,4GlcNAc) and lactose. They are elongated to different isomers of lacto-N-hexa-, octa- and decaose and then, decorated with fucose and sialic residues via α-1,2/3/4 linkages and α-2,3/6 respectively [9,10]. Bifidobacteria have a unique metabolism on HMO resulting in their predominant proliferation in the colon of breast-fed infants [7]. Specifically, *Bifidobacterium bifidum* metabolizes type 1 human milk oligosaccharides by the action of extracellular sialidases, fucosidases and lacto-N-biosidase.

To date, the structures of thirteen GH20 enzymes are known. Their domain organization is diverse and several accompanying domains are present apart from the catalytic GH20 domain. In addition, some of them exist as dimeric proteins. We aimed at defining the minimal functional domain architecture for enzymatic activity in these GH20 family enzymes. We chose the large *B. bifidum* lacto-N-biosidase (LnbB) as a model protein because of its complex domain organization among GH20 enzymes and its significance in the metabolism of human milk oligosaccharides. The LnbB gene encodes for a membrane-anchored extracellular enzyme of 1112 amino acid residues. Recently, it was cloned and the recombinant full length protein was characterized confirming that this enzyme catalyzes the hydrolysis of the tetrasaccharide Galβ1,3GlcNAcβ1,3Galβ1,4Glc to lacto-N-biose (Galβ1,3GlcNAc) and lactose (Fig 1) [11,12]. LnbB is a retaining exo-enzyme that acts by substrate-assisted catalysis, which is a common feature of GH20 enzymes.

By means of structural and sequence analysis of GH20 enzymes of known structure, we have established two levels of domains organization in this family, and further engineered LnbB in different truncated forms according to these models. We report the structural requirements for functionality, of this and the other GH20 enzymes, to obtain a minimal functional unit that retains the enzymatic activities. The importance of a remote element in the active site such as loop 2 seen in some GH20 enzymes is analyzed in all structures. It is shown that the isolated GH20 catalytic domain in LnbB is inactive and requires a remote element provided by

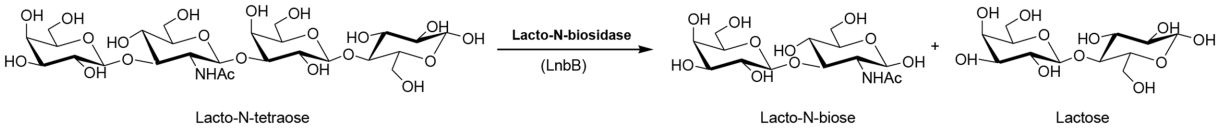


Fig 1. Hydrolytic reaction of lacto-N-tetraose catalyzed by *B. bifidum* lacto-N-biosidase.

doi:10.1371/journal.pone.0128075.g001

the adjacent lectin domain for activity, as observed in a complementation experiment. This structural and functional analysis provides new insights for further enzyme engineering for biotechnological and biomedical applications.

Materials and Methods

Cloning, expression and purification of Lacto-N-biosidase enzyme *from Escherichia coli*

Two synthetic genes of Lacto-N-biosidase from *Bifidobacterium bifidum* (LnbB), codon-optimized for the expression in *Escherichia coli* were produced by GenScript (GenScript, NJ, USA). The LnbB gene (37–1064) was cloned into the pET24b plasmid (Novagen, Madison, WI, USA) using the restriction sites NdeI and XhoI, and the construct A gene (40–528) into the pET28a + vector using NdeI and Bpu1102. Constructs B (176–497) and Bα (176–528) genes were obtained by PCR reaction using construct A as template, and constructs D (40–776) and F (546–776) genes using the full length LnbB gene. The following forward and reverse primers were used (restriction site are underlined): P1 (5′ -GGAATTCCATATGAAGCCGAAATATAAAGAACGT-3′) and P2 (5′ -TAATAGCTCAGCGGCGTCAAC-3′) for construct B; P1 and P3 (5′ -TAATAGCTCAGCACGGGAGTCACTCCA-3′) for construct Bα; P4 (5′ -GCGTATCCATGGCTGATGACTCCGCAGCAGGCTA-3′) and P5 (5′ -GCGTATCTCGAGCAGGCTGCCGGTCAGC-3′) for construct D, and P6 (5′ -GCGTATCCATGGTGGATGCGGGTATC-3′) and P5 for construct F. All constructs were cloned into the pET28a+ vector. *E. coli* DH5 cells were transformed and positive transformants were verified by DNA sequencing.

Expression and purification of the full length protein and constructs were carried out as follows: recombinant cultures of *E. coli* BL21 (DE3) *star* cells were grown following an autoinduction protocol [13] in LB medium with 30 $\mu g \cdot mL^{-1}$ kanamycin, 25mM Na_2HPO4, 25mM K_2HPO4, 50mM NH_4Cl, 5mM Na_2SO_4, 2mM MgSO4, 0.5% glycerol, 0.05% glucose and 0.2% lactose at 250 rpm and 30°C for 24 h. The full length protein and constructs A, B and Bα were also obtained using 1 mM IPTG induction after 16 h of growth at 37°C in LB medium with kanamycin achieving similar results. Cells were harvested by centrifugation (4000*g*, 15 min, 4°C), resuspended in buffer A (Na_2HPO_4 20mM, NaCl 150mM, pH 7.5) and disrupted in a cell disrupter (20000 psi). After centrifugation (20000*g*, 4°C, 2 h), the supernatant was loaded into a 1mL Ni^{2+}-charged HiTrap chelating column chromatography. Proteins were purified to homogeneity by a linear gradient of Buffer B (Na_2HPO_4 20 mM, NaCl 150 mM, imidazole 500mM, pH 7.5). Collected fractions were analyzed by SDS-PAGE and fractions containing the proteins were combined and dialyzed against buffer A with a 30 kDa cut-off membrane. All proteins were purified further using an XK16/60 Superdex 200 column (GE Healthcare), preequilibrated with buffer A. Fractions were pooled, concentrated with a Centricon YM-10 unit (Merk Millipore) and stored at 4°C. Protein concentrations were determined spectrophotometrically by the BCA method measuring A590 nm using BSA as standard [14].

Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK) with a laser light wavelength of 632.8 nm and a scattering angle of 173 degrees. Temperature was set at 25°C. Protein solutions in buffer A were measured without previous dilution. Full length protein showed a single peak for monodisperse particles of 10 nm and PDI 0.30, which is consistent with monomeric form.

Kinetics of the hydrolase activity

Lacto-N-biosidase activity assays with *p*-nitrophenyl β-lacto-N-bioside (Toronto Research Chemicals Inc.) were performed at 30°C in a volume of 0.1 mL using 96-wells microtiter plates

measuring p-nitrophenol concentration at 405 nm. Enzymatic reactions were performed in a reaction mixture containing 0.25 mM substrate and 50mM of citrate-phosphate buffer (pH 4.5). After pre-incubation for 5 min, reactions were initiated by addition of the enzyme (20 nM of full length protein and construct D, 20 nM-4 μM of construct A). At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 10) and p-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed.

Complementation assays

Enzymatic reactions with constructs A and F were performed using standard conditions in a Bravo liquid handling Robot (Agilent). After pre-incubation of construct A (50 nM) and F (25 nM-500 nM) for 5 minutes, addition of the substrate initiated the reaction. Triplicate measurements at each concentration were performed. The complex of constructs A and F was also characterized using gel filtration chromatography. The mixture of construct A (42 μM) with excess of construct F (245 μM) was loaded to the Superdex 200 column pre-equilibrated with buffer A and calibrated with protein standards. Fractions were further analyzed by SDS-PAGE and their hydrolase activity was determined. This purification was performed by duplicate.

Enzymatic activities of construct A (50 nM) in the presence of BSA at 1:1, 1:4 and 1:10 ratio were analyzed using standard conditions.

GH20 structures and sequences retrieval and analysis

The list of currently available 3D structures of GH20 enzymes was retrieved from CAZY database (http://www.cazy.org, [3]). Protein structures were downloaded from the Protein Data Bank (http://www.pdb.org). Protein sequences were retrieved from UniProt (http://www.uniprot.org). Assignment of protein domains were obtained from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY) and the consensus domain organizations are reported in Table 1. Protein structures were analysed and visualized with VMD [15]. Structure alignment was performed with MultiSeq [16] restricting the alignment region to GH20b and GH20 domains only. Sequence alignment was performed with PROMALS [17] with default options. Only the sequence fragments covering GH20b, GH20 domains and extra α-helix were considered.

Results and Discussion

1. Structural domains organization of GH20 proteins

According to CAZY classification [3], GH20 family is mainly composed of both prokaryotic and eukaryotic β-N-acetylhexosaminidases with different substrate specificities. Using data integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases), we report in Table 1 the consensus domains organization in GH20 proteins of known structure. The catalytic activity in this family of enzymes is assigned to the GH20 domain (PFAM code PF00728) which typically folds into a $(\beta/\alpha)_8$-barrel topology. Except for dispersin B from *A. actinomycetemcomitans* that presents only a single GH20 domain, the rest of structures show the catalytic domain accompanied by several domains with quite diverse functionalities: a non-catalytic domain, commonly named as GH20b, which is conserved in most GH20 enzymes although with unknown function, several lectin domains, carbohydrate binding domains, and other domains of unknown function.

Due to this diversity in domain organization, different questions come out: are all the accompanying domains essential for activity? Would the GH20 catalytic domain alone be

Table 1. Domain organization of GH20 β-N-acetylhexosaminidases.

Model A					
Enzyme	**Organism**	**Length (aa)**	**PDB**	**Domains[1]**	**GH20b-GH20-α location**
Monomeric					
β-N-acetylhexosaminidase (SpHex)	*Streptomyces plicatus*	506	1HP5[A]	GH20b-GH20-α	19–494
β-N-acetylhexosaminidase (ScHex)	*Streptomyces coelicolor* A3(2)	535	4C7G[A]	GH20b-GH20-α	11–511
β-N-acetylhexosaminidase (Hex1T)	*Paenibacillus* sp.	978	3GH5[A]	GH20b-GH20-α-Lectin-CBD	15–493
Chitobiase (SmCHB)	*Serratia marcescens*	885	1QBB[A]	CHB_HEX-GH20b-GH20-α-CHB_HEX_C	215–814
Lacto-N-biosidase (LnbB)	*Bifidobacterium bifidum*	1112	4JAW[A,B]	GH20b-GH20-α-Lectin-CBM32-Ig like	34–515
β-N-acetylhexosaminidase (NahA)*	*Arthrobacter aurescens*	540	3RCN[A]	GH20b-GH20-	5–506
β-N-acetylhexosaminidase (Bf3009*)	*Bacteroides fragilis*	518	4PYS[A,B]	GH20b-GH20-α	21–498
Heterodimeric					
β-N-acetylhexosaminidase (HexA)	*Homo sapiens*	529/556	2GK1 [A,B,C, D,E,F,G]	GH20b-GH20-α	α:23–510 β:56–540
Homodimeric					
β-N-acetylhexosaminidase (HexB)	*Homo sapiens*	556/556	1NP0[A,B]	GH20b-GH20-α	56–540
β-N-acetylhexosaminidase (OfHex1)	*Ostrinia furnacalis*	593/593	3OZO[A]	GH20b-GH20-α	65–574
β-N-acetylhexosaminidase (GcnA)	*Streptococcus gordonii*	627/627	2EPN[A,B]	GH20b-GH20-α-Domain III	2–414
Model B					
Enzyme	**Organism**	**Length (aa)**	**PDB**	**Domains[1]**	**GH20 domain location**
Monomeric					
β-1,6-N acetylglucosaminidase Dispersin B (DspB)	*Aggregatibacter actinomycetemcomitans*	361	1YHT[A]	GH20	20–342
β-N-acetylhexosaminidase (StrH)	*Streptococcus pneumoniae* R6	1312	3RPM[A,B]	GH20-GH20-G5-G5	190–538 635–972
	Streptococcus pneumoniae TIGR4	1312	2YL8[A]	GH20-GH20-G5-G5	190–538
			2YL9[A,B,C, D]	GH20-GH20-G5-G5	635–972

Data obtained using integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases). PDB structure accession code and chain identifier used for structural analysis are indicated and domains actually present in the PDB structure are underlined.

[1] **Domains**: domain nomenclature according to the consensus domain organization by integrating Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases. Preferentially CAZY annotation is used, if available. CBD: carbohydrate binding domain, CBM32: carbohydrate binding module family 32, CHB_HEX: chitobiose hexosaminidase accompanying domain in the N-terminus, CHB_HEX_C: chitobiose hexosaminidase C-terminal accompanying domain, Lectin: lectin-like domain according to PSI-BLAST, Ig like: Immunoglobulin-like fold domain, Domain III: C-terminal domain needed to dimerize according to authors, G5: accompanying domain of carbohydrate metabolism enzymes.

*: Oligomerization state in solution is unknown

doi:10.1371/journal.pone.0128075.t001

functional? Which is the importance of GH20b? Several studies on the structure-function relationships at the active site have been performed [18–22] but only few have addressed the structural organization of these enzymes [23–25].

Regardless of the number of accompanying domains, we propose two different models of organization (Table 1): Model A, comprising proteins with at least both GH20b and GH20 domains, which is characterized by a GH20b-GH20-α architecture, where GH20 is always accompanied by GH20b at the N-terminus and followed by an extra α-helix after the GH20 $(\beta/\alpha)_8$-barrel, and Model B with proteins that only present the GH20 domain without GH20b.

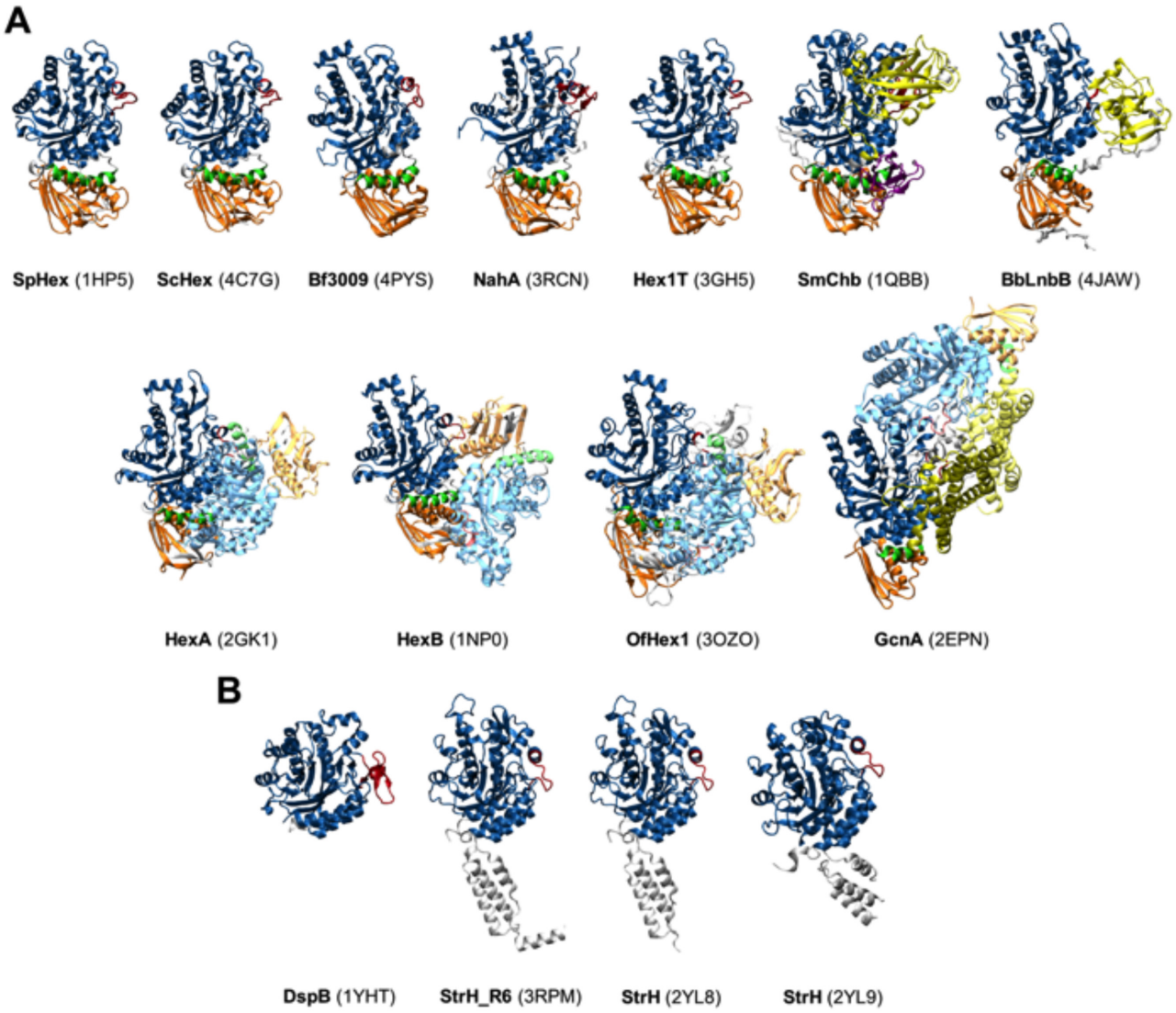


Fig 2. Crystal structures of GH20 β-N-acetylhexosaminidases. (A) Structures according to model A architecture. (B) Structures according to model B architecture. GH20b domains are colored in orange, GH20 domains in blue, α-helix in green, and the rest of accompanying domains in yellow and purple. Loop 2 from GH20 domain is in red. Dimer counterparts follow the same coloring scheme but with lower intensity.

doi:10.1371/journal.pone.0128075.g002

To date, eleven of the thirteen GH20 enzymes structures fit into Model A (Fig 2A). As summarized in Table 1, five of them are monomeric proteins in solution (SpHex, ScHex, Hex1T, SmCHB, and LnbB), whereas other four are dimeric proteins (HexA, Hex B, OfHex1, and GcnA). The oligomerization state of the *Arthrobacter aurescens* (NahA) and *Bacteroides fragilis* (Bf3009) enzymes remains unknown.

For the monomeric proteins, SpHex and ScHex structures, which share 94% sequence identity, comprise the GH20b domain followed by the catalytic GH20 domain [26,27]. Hex1T consists of four domains, GH20b and GH20, and domains III and IV which are described as concanavalin A-like lectin and cellulose-binding domains respectively, according to Superfamily database (Supfam). The Hex1T crystal structure is a truncated form (residues 1–502), with only the GH20b and GH20 domains [25]. SmCHB is also constituted by four domains, where the N-terminal domain (CHB_HEX) is annotated as a carbohydrate binding module although its

function remains unknown, and the C-terminal domain (CHB_HEX_C) as an immunoglobulin-like fold domain (Supfam). The central domains correspond to GH20b and GH20 domains [21]. LnbB is formed by five domains. Domains I and II are the GH20b and GH20 ones, Domain III (residues 546–700) is similar to a lectin domain according to a PSI-Blast search, Domain IV (784–932) has sequence homology to CBM32 (Supfam), and domain V (962–1041) is described as a Immunoglobulin-like domain in PFam. The recent crystal structure of this enzyme corresponds to a C-terminal truncated form (41–663) with the GH20b, GH20 and part of the lectin-like domain [23]. Finally, NahA and Bf3009, with unknown oligomerization states but included in the group of monomeric proteins in Table 1, present only the GH20b and GH20 domains.

For the dimeric proteins, each subunit of HexA, HexB [28,29] and OfHex1 [30] have only the two domains, GH20b and GH20, while the GcnA structure contains an additional domain III, which is predominantly α-helical and is reported to form a significant part of the dimer interface [31]. In general, these accompanying domains are non-catalytic domains involved in binding and substrate specificity but still the mechanism by which this occurs remains uncertain [32,33].

The remaining two GH20 enzymes present structures that are assigned to Model B (Fig 2B): dispersin B from *Aggregatibacter actinomycetemcomitans* (DspB) and the β-hexosaminidases from *Streptococcus pneumoniae* (StrH). Both are monomeric proteins in solution where DspB presents only one domain, the unique GH20 domain [34], and StrH consists of four domains, where the N-terminus is a tandem repeat of two GH20 domains, designated as GH20-1 and GH20-2, and the C-terminus is a tandem repeat of two domains tentatively designated as G5 with unknown function (Pfam). Truncated forms containing a single GH20-1 and GH20-2 domains of *S. pneumoniae* were crystallized [24]. The GH20-1 and GH20-2 domains are quite similar although specific features confer differential substrate specificities.

2. *B. bifidum* lacto-N-biosidase truncated forms

The full length protein was truncated according to the two different models of structure organization we propose, and three different truncated proteins were designed: construct A, B and Bα (Fig 3).

Construct A (residues 40–528) follows the architecture organization of Model A enzymes and is composed of the GH20b and GH20 domains always followed by the extra α-helix. This construct would be expected to be active since several enzymes of model A exhibit just these two domains alone in the wild-type protein (such as SpHex structure and HexA, HexB and OfHex1 subunits). Furthermore, a truncated form of Hex1T (encoding the GH20b-GH20-α domains alone) is reported to be active [25]. However, a previous work by Ito *et al.* (2013) in which different LnbB truncated forms were designed, revealed that a truncated protein (37–520) equivalent to construct A exhibits no catalytic activity. A longer truncated protein (37–663) composed by GH20b-GH20-α and a part of the lectin-like domain was reported to be as active as the full length. These data indicate that the GH20b-GH20-α architecture is not the minimal functional unit for LnbB, as it is for the rest of the model A GH20 enzymes. Further studies are presented to determine the details of the minimum requirement for LnbB to be catalytically active.

Construct B (residues 176–497) mimics Model B enzymes and is formed by the single catalytic GH20 domain alone. Lastly, construct Bα (residues 176–528) was designed to tackle the open question whether the accompanying GH20b domain can be removed in Model A enzymes. This construct encodes the catalytic GH20 domain with the extra α-helix at the C-terminus. Both constructs, B and Bα, were designed for the study of the behaviour of the GH20 catalytic domain in model A proteins.

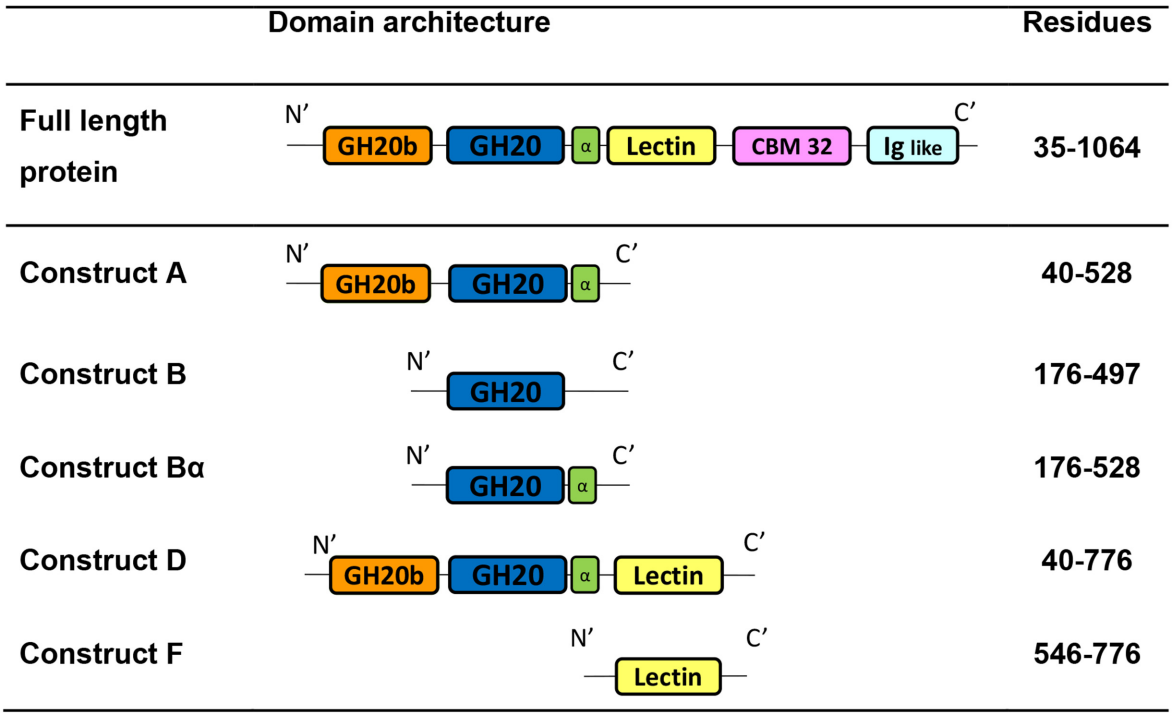


Fig 3. Truncated forms of *B.bifidum* Lacto-N-biosidase.

doi:10.1371/journal.pone.0128075.g003

Engineered proteins of LnbB enzyme were cloned and expressed in *E. coli*. The full length protein (37–1064), without the signal peptide and membrane anchor regions, was expressed as C-terminal His_6-Tagged protein. Constructs A, B and Bα were expressed as N-terminal His_6-Tagged proteins. Cultures were grown at 30°C for 24 h with autoinduction medium. The full length protein and construct A were obtained as soluble intracellular proteins with yields of 16 and 27 mg/L of culture respectively. SDS-PAGE analysis and MALDI-TOF mass spectrometry confirmed the theoretical 112 kDa of the full length protein and 57 kDa of Construct A. Further characterization by gel filtration chromatography showed that the full length protein is monomeric with particle size of about 10 nm by dynamic light scattering, whereas Construct A is mainly dimeric in solution (Fig 4). The full length protein was also expressed as N-terminal His_6-Tagged protein to evaluate the effect of the N-terminal tag. Both, C-terminal and N-terminal His_6-Tagged full length proteins, were soluble and had the same specific activity (data not shown).

On the other hand, constructs B and Bα were only present in the insoluble pellet and no soluble protein was achieved after different expression conditions such as autoinduction or IPTG induction at different temperatures (data not shown). This seems to indicate that the single GH20 catalytic domain of LnbB, with or without the extra α-helix, is not stable enough and yields insoluble aggregates. Thus, GH20b domain at the N-terminus of GH20 is required for LnbB expression and stability. This domain is present in all model A enzymes being the relative orientation of GH20b and GH20 catalytic domain the same in all structures (Fig 2). The number of β-strands of the GH20b domain varies (for example five in GcnA, six in HexA and HexB and seven in SpHEx and LnbB) but the interface between both domains is well conserved: two loops (residues 54–60 of loop 1 and residues 127–130 of loop 4 in LnbB) and three helices (residues 82–99, 146–163 and the extra α-helix of the GH20 domain which is oriented toward an hydrophobic pocket). GH20b domain is not only found in GH20 enzymes but also in some GH84 hexosaminidases and GH67 glucoronidases annotated in Pfam as GH20b and GH67N

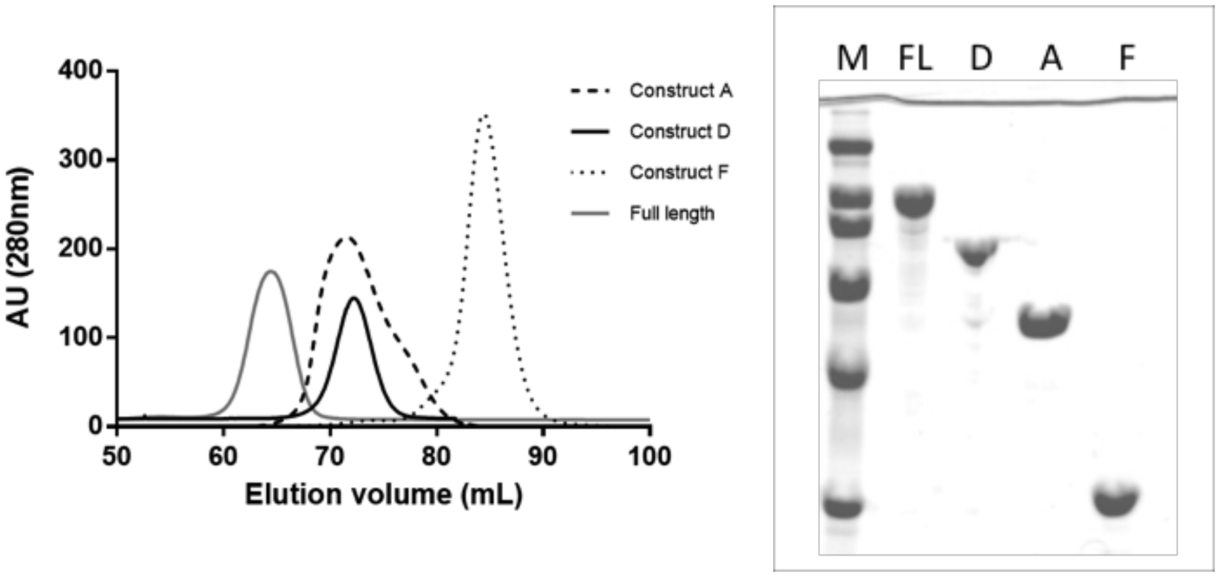


Fig 4. Size exclusion chromatography of the full length protein and different constructs of Lacto-N-biosidase from *B.bifidum*. (A) Chromatogram of the full length protein (FL) and constructs A, D and F. (B) SDS-PAGE of the eluted proteins. M: broad range protein molecular weight marker (200, 116.2, 97.4, 66.2, 45, 31 kDa).

doi:10.1371/journal.pone.0128075.g004

[35], respectively, always located at the N-terminus of the catalytic domain. Considering this scenario, we conclude that GH20b-GH20-α architecture of model A enzymes cannot be reduced to that of model B.

The hydrolytic activity of the soluble constructs was assayed using p-nitrophenyl β-lacto-N-bioside as substrate at pH 4.5 and 30°C. Full length protein showed a specific activity of $27.2 \pm 3.5\ s^{-1}$, but construct A was inactive, as expected according to previous studies [23]. The GH20b-GH20-α architecture of LnbB is enough to guarantee protein expression, but this architecture does not correspond to the minimal functional unit maybe requiring other accompanying domains from the C-terminus for the enzyme to be active.

3. Structural requirements in the GH20 domain for functionality

While this minimal GH20b-GH20-α architecture assures the activity of SpHex and Hex1T [25,36] and also of dimeric proteins such as HexA, HexB and OfHex1 [30,37,38], it is not enough for LnbB. The crystal structure of LnbB (residues 41–663) in complex with the cyclic intermediate analogue lacto-N-biose-thiazoline [23] consisting of three out of the five domains (GH20b, GH20 and part of the lectin-like domain) was active. The accompanying lectin domain resulted to be the missing element required in our construct A to be fully active. We have carefully analyzed the interaction of the catalytic GH20 domain with other C-terminus domains in all crystallized protein structures of model A (Fig 5). Exploring the structure, important structural features can be assigned. The lectin domain interacts with the catalytic GH20 domain, providing a remote element that folds on the active site (Fig 5A). This element (residues 570–578) consists of a beta-turn, in which Leu574 completes the substrate binding cavity favouring a hydrophobic interaction with the hydroxymethyl group of the galactose moiety of the substrate. In addition, Leu574 is positioned between two conserved amino acids Gln190 and Asp467 of subsites -1 and -2 forming further interactions between this turn and the active site (Fig 5A). For instance, through a water mediated hydrogen bond, the backbone nitrogen atom of Leu574 favours the orientation of Gln190 towards the O4 of the galactose unit of the

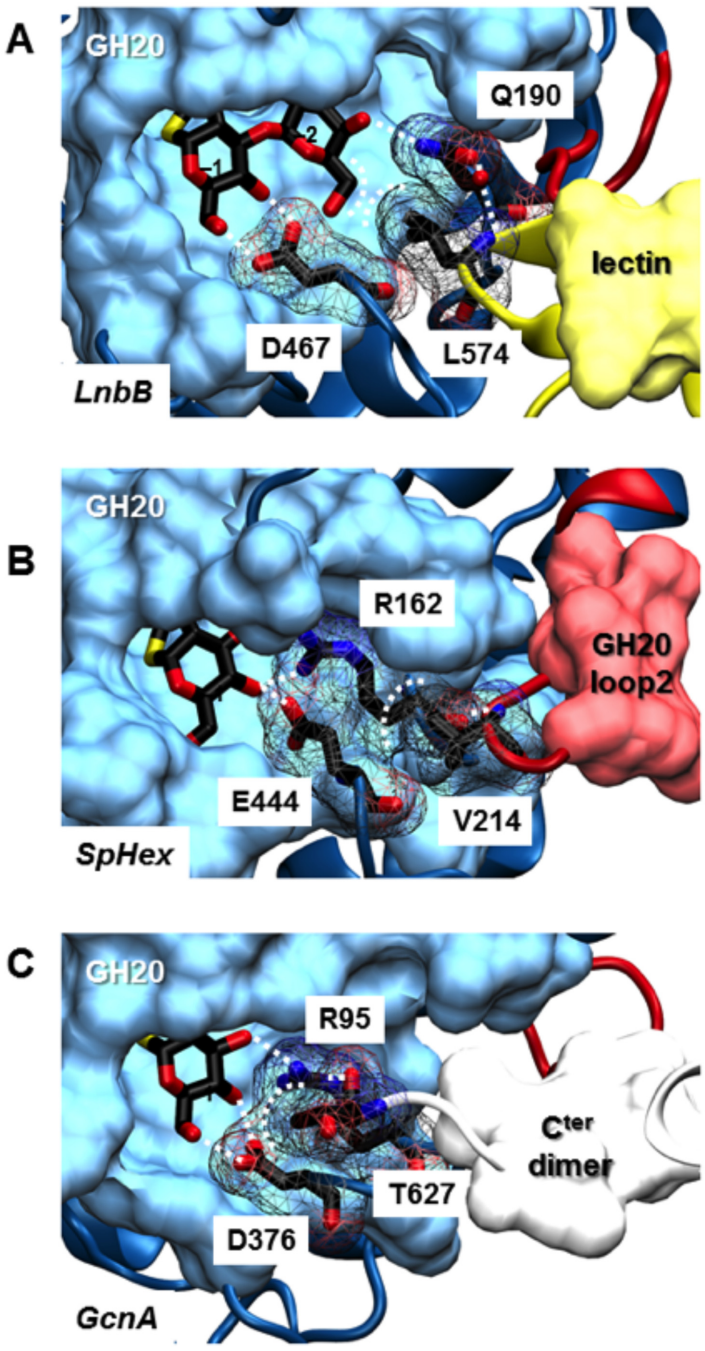


Fig 5. Remote elements and their key residues that complement the active site of GH20 β-N-acetylhexosaminidases. (A) *B. bifidum* lacto-N-biosidase. (B) *S. plicatus*. (C) *S. gordonii* N-acetylhexosaminidases. Dot lines represent interactions between residues of the remote element and the active site as indicated in Table 2.

doi:10.1371/journal.pone.0128075.g005

disaccharide ligand. Engineered forms of LnbB in which part of the C-terminus of the lectin domain were even shorter [23] or completely removed as in our construct A, showed an important decrease in activity of the enzyme, indicating that the interactions we observe in the structure between the remote lectin domain and the substrate binding site are essential for catalysis.

Table 2. Remote element and conserved interactions with the active site of model A GH20 β-N-acteylhexosaminidases.

Enzyme	Loop 2 length	Remote element	GH20 conserved Arg/Asp or Glu at subsite –1	Remote element interactions with conserved Arg/Asp	
				non-polar interactions	polar interactions
SpHex	Long G209-T223	Loop 2	R162	V214	V214
			E444	V214	-
ScHex	Long G238-T252	Loop 2	R191	V243	V243
			E473	V243	-
NahA	Long G193-T219	Loop 2	R146	-	R196
			E456	-	-
Bf3009	Long G200-S214	Loop 2	R153	-	Q209
			D450	-	
Hex1T	Long G217-T231	Loop 2	R170	V222	V222
			E443	V222	-
SmCHB	Long G396-S425	Loop 2	R349	Y413	R399
			E739	Y413	-
LnbB	Short T231-T236	Lectin G570-T578	Q190*	L574	S233
			D467	L574	-
HexA	Short αG225-αT235	C-terminus subunit β A543-N552	αR178	βA548	βG549
			αE462	βY547	-
HexA	Short βG258-βT267	C-terminus subunit α A514-E523	βR211	αV519	αG520
			βE491	αV519	-
HexB	Short G258-T267	C-terminus dimer A514-E523	R211	A548	G549
			E491	Y547	-
OfHex1	Short G267-T276	C-terminus dimer A577-S594	R220	P582	E583
			E526	P582	-
GcnA	Short G136-T143	C-terminus dimer S622-T627	R95	-	T627
			D376	T627	-

*Residue in LnbB substituting the function of the conserved arginine but in subsite -2.

doi:10.1371/journal.pone.0128075.t002

Interestingly, we have detected that this pattern is reproduced in all proteins of model A: there is always a remote element assisting in the definition of the substrate binding cavity by means of direct interactions with conserved amino acids directly involved in protein—substrate interactions. Table 2 summarizes the exact localization of these remote elements for each member of model A and the interactions it forms with conserved active site residues. In general, this remote element is provided by a long loop of the GH20 domain itself (loop 2). In the cases where this loop is too short, the remote element comes from an accompanying domain (such as the lectin domain in the case of LnbB) or from the C-terminus of the other monomer in the case of dimeric proteins. In detail, structures of SpHex, ScHex, NahA, Bf3009, SmCHB and Hex1T present this remote element coming from the extended loop 2 of the GH20 domain itself (Fig 5B). This loop positions a hydrophobic residue, such as Val214 in SpHex (equivalent to Leu574 in LnbB), into the second shell of the subsite -1 cavity just behind the conserved arginine and aspartate in all known GH20 hexosaminidases, Arg 162 and Asp444 (equivalent to Gln190 and Asp467 in LnbB, see alignment in Fig 6). These amino acid residues can form hydrophobic interactions between their side chains and also, a hydrogen bond between the

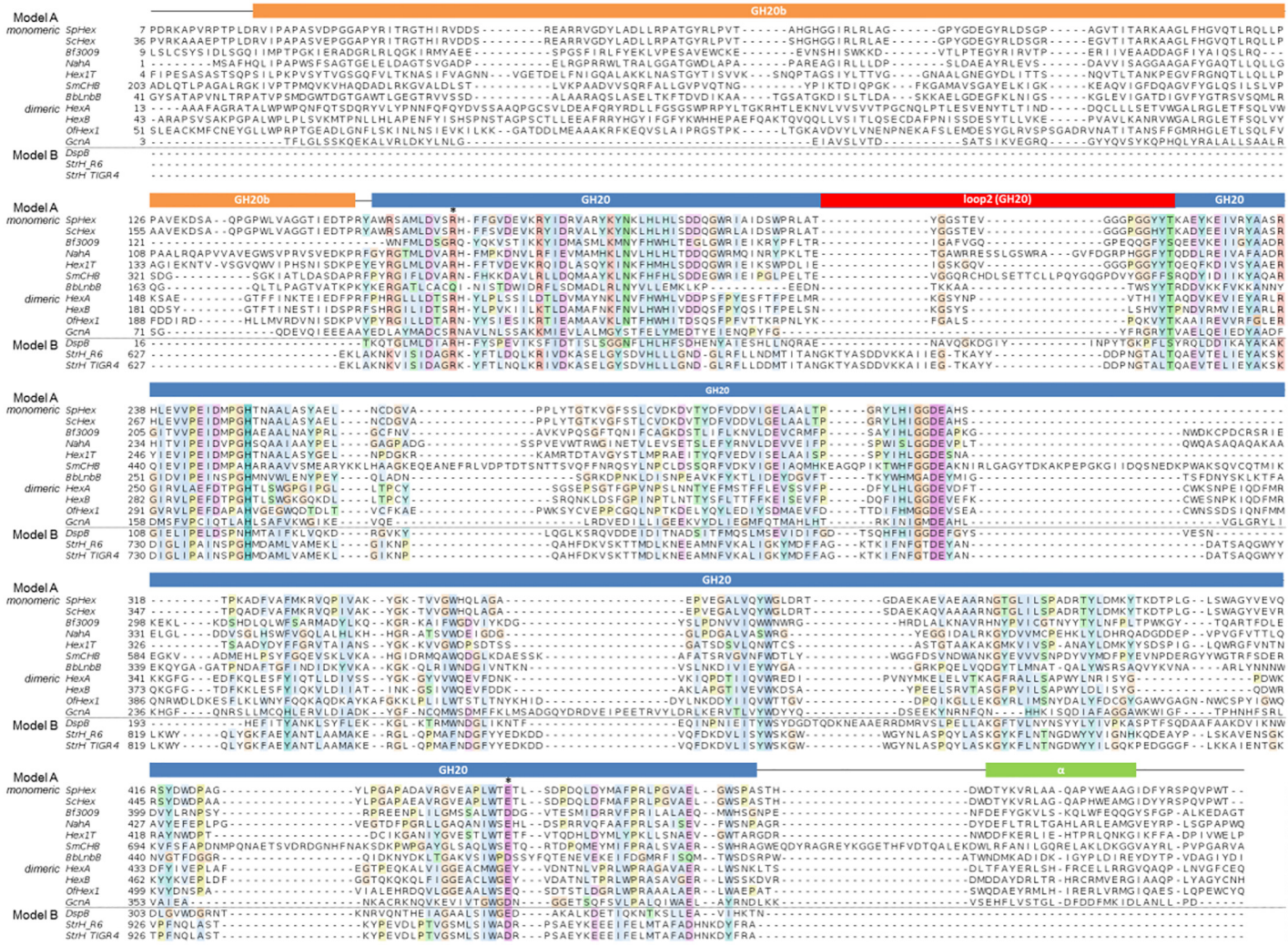


Fig 6. Multiple sequence alignment of β-N-acetylhexosaminidases of known structure from family GH20. Domains distribution is shown as colored boxes on top. Accompanying domains are not shown. Conserved arginine and glutamate/aspartate sites are marked with a star. Sequences were downloaded from UniProt entry names: SpHex (O85361_STRPL), ScHex (Q9L068), Bf3009 (Q5LAT3) NahA (A1RBZ5_ARTAT), Hex1T (D2KW09_9BACL), SmCHB (CHB_SERMA), LnbB (B3TLD6_BIFBI), HexA (HEXA_HUMAN), HexB (HEXB_HUMAN), OfHex1 (Q06GJ0_OSTFU), GcnA (Q6ST21_STRGN), DspB (Q840G9_AGGAC), StrH_R6 (Q8DRL6_STRR6), StrH_TIGR4 (STRH_STRPN). Alignment was performed with PROMALS [17]. Conserved positions are coloured according to ClustalW colour scheme.

doi:10.1371/journal.pone.0128075.g006

backbone NH group of Val214 and the carbonyl group of the backbone of Arg162 (Fig 5B). These interactions between residues of the remote element and the conserved arginine and aspartate of subsite -1, are present in all the structures with long loop 2 (Table 2). ScHex presents Val243 of loop2 next to Arg191 and Asp473 in the same way. NahA and SmCHB show the longest loops 2 as seen in the alignment. They provide the remote elements next to the corresponding arginines, Arg146 and Arg349, and aspartates, Asp456 and Asp739 respectively. One arginine of the remote element, Arg196 in NahA and Arg399 in SmCHB, is hydrogen-bonded to the backbone of the active site arginine. Similarly, Bf3009 presents Gln209 of loop2 next to Arg153 and Asp450. In Hex1T, Val222 from the remote element is positioned behind Arg170 and Asp443 and the backbone NH group of this valine donates a hydrogen bond to the arginine carbonyl oxygen atom. Interestingly, a truncated form of Hex1T with only the GH20b and GH20 domains has been reported to be active [25]. Consistently with our explanations,

this truncated form does not affect the definition of the active site, because this construct keeps the long loop 2 in the GH20 domain.

By contrast, when loop 2 is short, this remote element comes from the accompanying domains, as in LnbB, or from the C-terminus domains of the other monomer in dimeric structures, such as, GcnA, HexA, HexB and OfHex1. For example, in the homodimeric structure of GcnA, domain III forms a significant part of the dimer interface [31]. The C-terminus of domain III of chain B (622–627) winds back into the structure and contacts the catalytic GH20 of chain A (Fig 5C). Specifically, the C-terminus amino acid residue Thr627 projects into subsite -1 and it is adjacent to the conserved arginine and aspartate (Arg95, Asp376) and is part of the active-site pocket of the dimer partner. Besides, the carboxyl group of this C-terminal Thr627 and the Arg95 side chain are at less than 3 Å distance, establishing electrostatic interactions. As in LnbB, any truncated form of this C-terminus domain III is believed to be inactive. In the dimeric structures of HexA, HexB and OfHex1 [28–30], dimerization occurs exclusively between the GH20 catalytic domains. Dimerization is essential for catalysis since residues of one subunit structurally complete the active site of the other one. For example, Tyr456 of subunit β is part of the active site of subunit α in HexA. In addition, the C-terminus of subunit β provides this remote element very close to the subsite-binding cavity of subunit α and vice versa (Table 2). Particularly, the hydrophobic residue Ala548 (subunit β) of this element in HexA forms part of the second shell of the cavity behind the conserved Arg178 (subunit α) and Asp462 (subunit α), and the backbone nitrogen atom of Gly549 (subunit β) interacts by hydrogen bond with the carbonyl oxygen atom of Arg178 (subunit α). The equivalent interactions between subunits are also seen in HexB and OfHex1 crystal structures (Table 2).

Therefore, all GH20 structures of model A clearly show the presence of a remote element folding on the substrate-binding cavity providing amino acid residues that either complete the active site, as in the extended subsite -1 of the *S. gordonii* hexosaminidase or in subsite -2 of the LnbB, or participate in the second shell of this cavity behind the conserved arginine and aspartate residues in subsite -1 of hexosaminidases.

To provide experimental evidence of the requirement of this remote element in LnbB, a complementation assay was designed. If Construct A (composed of the GH20b and GH20-α domains) is inactive because of the lack of the remote element, would addition of the lectin domain (as a single protein) reconstitute the active site and restore activity? Toward this end, Construct F (the single lectin domain, Fig 3) to assay complementation with Construct A, as well as Construct D (Fig 3), which contains the non-catalytic GH20b domain, the catalytic GH20 domain and the lectin domain, were expressed. The C-terminal His_6-Tagged proteins were purified and analyzed by SDS-PAGE and MALDI-TOF, confirming the corresponding theoretical molecular masses of 83 and 28 kDa for constructs D and F, respectively. Gel filtration chromatography showed that both proteins were monomers in solution (Fig 4). Construct D was active with the same specific activity than the full length enzyme on p-nitrophenyl β-lactobioside (27 s^{-1}). Therefore, both C-terminal domains in LnbB, the CBM32 and Ig like domains are not important for the catalytic activity, and construct D is the minimal functional unit of LnbB. Next, a complementation experiment was assayed by adding Construct F (the single lectin domain) on the inactive Construct A. Activity was restored in a concentration-dependent manner when increasing the concentrations of Construct F (Fig 7). At a molar ratio of 1:6 activity levels off at 11% of the hydrolytic activity of the full length enzyme. The interaction between constructs A and F at a 1:6 molar ratio was analyzed by gel filtration chromatography (Fig 8). Fractions L1–L2 correspond to the excess of Construct F, the lectin domain with molecular weight of 28 kDa, and fractions C1–C6 to proteins with molecular weight of 74 kDa that could be the dimer of construct A (as in Fig 4) and the complex between A and F with theoretical molecular mass of 85 kDa. The hydrolase activity of Fractions C1–C6 and the presence of

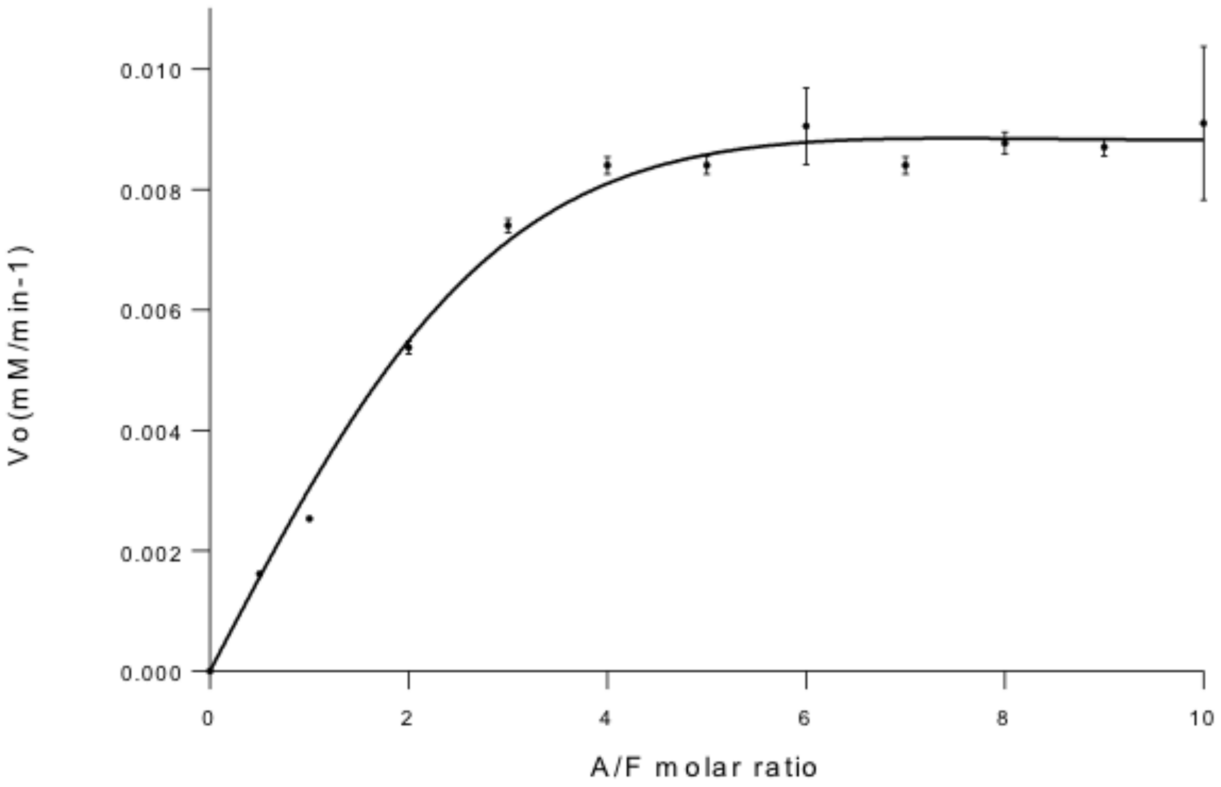


Fig 7. Complementation assay of construct A with construct F at different molar ratios. Conditions: 50 nM construct A, 25 nM-500 nM construct F, 0.25 mM p-nitrophenyl β-lacto-N-bioside, 25 mM citrate-25 mM phosphate buffer, pH 4.5, 30°C. The relative standard deviation did not overcome 2% except at 6 and 10 molar ratio that was lower than 15%.

doi:10.1371/journal.pone.0128075.g007

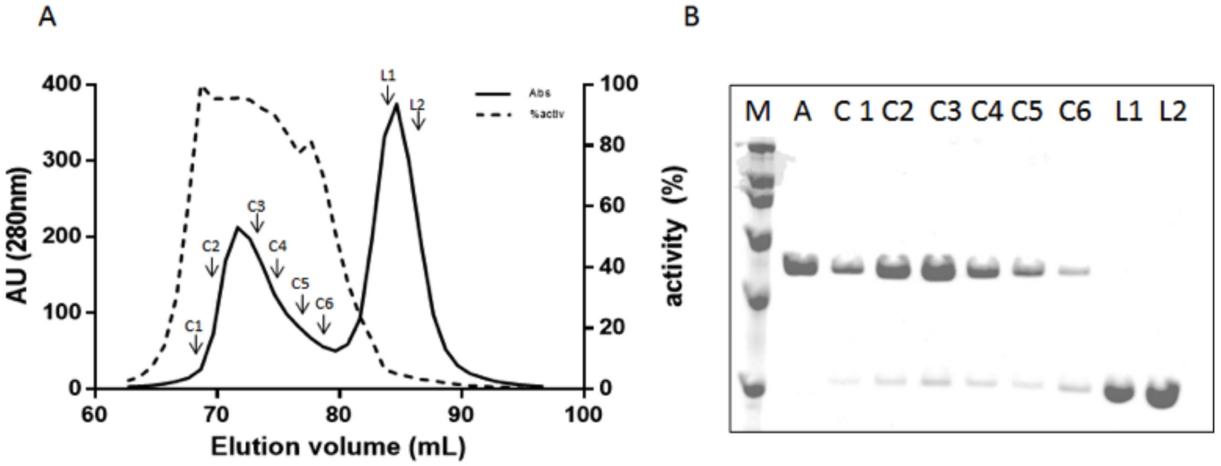


Fig 8. Size exclusion chromatography of the complementation assay in a ratio 1:6 of constructs A and F. (A) Chromatogram of the eluted fractions: absorbance at 280nm (-), hydrolase activity (•••). (B) SDS-PAGE of different fractions. M: broad range protein molecular weight marker (200, 116.2, 97.4, 66.2, 45, 31 kDa); A: construct A Standard, fractions C1–C6, fractions L1–L2. The relative standard deviations for the molecular weights and activities were less than 2% and 15% respectively.

doi:10.1371/journal.pone.0128075.g008

lectin in these fractions seen in SDS-PAGE (Fig 8) demonstrate that the complex between A and F is formed and active. As it is observed in Fig 8, not all construct A protein is complexed with construct F. By densitometry analysis we determined that the lectin represents 10% of total protein in the mixture. Assuming a 1:1 interaction, 10% of Construct A is in complex with Construct F and therefore the specific activity of the complex is similar to that of the full length protein. To exclude non-specific interactions, complementation assay was performed with Construct A and BSA at different concentration ratios and no activity recovery was found (data not shown).

Conclusions

This structural-functional analysis allows us to come to different conclusions. First, we identified two levels of domain organization in GH20 enzymes: model A with GH20b-GH20-α architecture and model B with the single catalytic GH20 domain alone. As LnbB constructs show, model A enzymes cannot be reduced to model B, since GH20b domain is a structural requirement in the N-terminus of GH20 to assure protein stability. Secondly, GH20b-GH20-α architecture is not always the minimal functional unit. For the LnbB enzyme here studied, the isolated GH20b-GH20-α domain is inactive and requires the lectin domain which provides an essential loop to shape the active site. Based on these results and analysis of the structures of GH20 enzymes, we propose a broader mechanism by which important interactions in the substrate binding cavity of the catalytic GH20 domain are provided by a remote element in all structures, which must be preserved to ensure a proper definition of the active site. The remote element can directly be provided by a long loop 2 of the GH20 domain itself. In this case, GH20b-GH20-α can act as a minimal functional unit, and large multidomain enzymes can be engineered to reduced forms with a minimal model A architecture. In other circumstances, when loop 2 is short, proteins dimerize or present an accompanying domain in the C-terminus of GH20 that provide this remote element. In the latter cases, protein engineering of these enzymes must be designed maintaining the accompanying domains. The proposed model can stimulate further studies on other members of the superfamily.

Author Contributions

Conceived and designed the experiments: CVC MF XB AP. Performed the experiments: CVC MF XB. Analyzed the data: CVC MF XB AP. Contributed reagents/materials/analysis tools: CVC MF XB AP. Wrote the paper: CVC MF XB AP.

References

1. Liu T, Yan J, Yang Q. Comparative biochemistry of GH3, GH20 and GH84 beta-N-acetyl-D-hexosaminidases and recent progress in selective inhibitor discovery. Curr Drug Targets. 2012; 13: 512–525. PMID: 22280348
2. Slamova K, Bojarova P, Petraskova L, Kren V. beta-N-acetylhexosaminidase: what's in a name. . .? Biotechnol Adv. 2010; 28: 682–693. doi: 10.1016/j.biotechadv.2010.04.004 PMID: 20438826
3. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucl Acids Res. 2008; 37: D233–238. doi: 10.1093/nar/gkn663 PMID: 18838391
4. Liu T, Chen L, Ma Q, Shen X, Yang Q. Structural Insights into Chitinolytic Enzymes and Inhibition Mechanisms of Selective Inhibitors. Curr Pharm Des. 2013.
5. Rountree JS, Butters TD, Wormald MR, Boomkamp SD, Dwek RA, Asano N et al. Design, synthesis, and biological evaluation of enantiomeric beta-N-acetylhexosaminidase inhibitors LABNAc and DABNAc as potential agents against Tay-Sachs and Sandhoff disease. ChemMedChem. 2009; 4: 378–392. doi: 10.1002/cmdc.200800350 PMID: 19145603
6. Bode L. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology. 2012; 22: 1147–1162. doi: 10.1093/glycob/cws074 PMID: 22513036

7. Garrido D, Kim JH, German JB, Raybould HE, Mills DA. Oligosaccharide binding proteins from Bifidobacterium longum subsp. infantis reveal a preference for host glycans. PLoS One. 2011; 6: e17315. doi: 10.1371/journal.pone.0017315 PMID: 21423604
8. Sela DA, Mills DA. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. Trends Microbiol. 2010; 18: 298–307. doi: 10.1016/j.tim.2010.03.008 PMID: 20409714
9. Kobata A. Structures and application of oligosaccharides in human milk. Proc Jpn Acad Ser B Phys Biol Sci. 2010; 86: 731–747. PMID: 20689231
10. Kunz C, Rudloff S, Baier W, Klein N, Strobel S. Oligosaccharides in human milk: structural, functional, and metabolic aspects. Annu Rev Nutr. 2000; 20: 699–722. PMID: 10940350
11. Hattie M, Debowski AW, Stubbs KA. Development of tools to study lacto-N-biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. Chembiochem. 2012; 13: 1128–1131. doi: 10.1002/cbic.201200135 PMID: 22514018
12. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M et al. Bifidobacterium bifidum lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. Appl Environ Microbiol. 2008; 74: 3996–4004. doi: 10.1128/AEM.00149-08 PMID: 18469123
13. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005; 41: 207–234. PMID: 15915565
14. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD et al. Measurement of protein using bicinchoninic acid. Anal Biochem. 1985; 150: 76–85. PMID: 3843705
15. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996; 14: 33–38. PMID: 8744570
16. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics. 2006; 7: 382. PMID: 16914055
17. Pei J, Grishin NV. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. Bioinformatics. 2007; 23: 802–808. PMID: 17267437
18. Pluvinage B, Higgins MA, Abbott DW, Robb C, Dalia AB, Deng L et al. Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. Structure. 2011; 19: 1603–1614. doi: 10.1016/j.str.2011.08.011 PMID: 22078560
19. Sumida T, Stubbs KA, Ito M, Yokoyama S. Gaining insight into the inhibition of glycoside hydrolase family 20 exo-beta-N-acetylhexosaminidases using a structural approach. Org Biomol Chem. 2012; 10: 2607–2612. doi: 10.1039/c2ob06636j PMID: 22367352
20. Williams SJ, Mark BL, Vocadlo D, James MNG, Withers SG. Aspartate 313 in the Streptomyces plicatus Hexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. Biol Chem. 2002; 277 (42): 40055–40065. PMID: 12171933
21. Prag G, Papanikolau Y, Tavlas G, Vorgias CE, Petratos K, Oppenheim AB. Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. J Mol Biol. 2000; 300: 611–617. PMID: 10884356
22. Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG, James MN. Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. J Biol Chem. 2001; 276: 10330–10337. PMID: 11124970
23. Ito T, Katayama T, Hattie M, Sakurama H, Wada J, Suzuki R et al. Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-N-biosidase from Bifidobacterium bifidum. J Biol Chem. 2013; 288: 11795–11806. doi: 10.1074/jbc.M112.420109 PMID: 23479733
24. Jiang YL, Yu WL, Zhang JW, Frolet C, Di Guilmi AM, Zhou CZ et al. Structural basis for the substrate specificity of a novel beta-N-acetylhexosaminidase StrH protein from Streptococcus pneumoniae R6. J Biol Chem. 2011; 286: 43004–43012. doi: 10.1074/jbc.M111.256578 PMID: 22013074
25. Sumida T, Ishii R, Yanagisawa T, Yokoyama S, Ito M. Molecular cloning and crystal structural analysis of a novel beta-N-acetylhexosaminidase from Paenibacillus sp. TS12 capable of degrading glycosphingolipids. J Mol Biol. 2009; 392: 87–99. doi: 10.1016/j.jmb.2009.06.025 PMID: 19524595
26. Mark BL, Vocadlo DJ, Zhao D, Knapp S, Withers SG, James MN. Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor. J Biol Chem. 2001; 276: 42131–42137. PMID: 11522797
27. Thi NN, Offen WA, Shareck F, Davies GJ, Doucet N. Structure and activity of the Streptomyces coelicolor A3(2) beta-N-acetylhexosaminidase provides further insight into GH20 family catalysis and inhibition. Biochemistry. 2014; 53: 1789–1800. doi: 10.1021/bi401697j PMID: 24559145
28. Lemieux MJ, Mark BL, Cherney MM, Withers SG, Mahuran DJ, James MN. Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. J Mol Biol. 2006; 359: 913–929. PMID: 16698036

29. Mark BL, Mahuran DJ, Cherney MM, Zhao D, Knapp S, James MN. Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. J Mol Biol. 2003; 327: 1093–1109. PMID: 12662933

30. Liu T, Zhang H, Liu F, Wu Q, Shen X, Yang Q. Structural determinants of an insect beta-N-Acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. J Biol Chem. 2011; 286: 4049–4058. doi: 10.1074/jbc.M110.184796 PMID: 21106526

31. Langley DB, Harty DW, Jacques NA, Hunter N, Guss JM, Collyer CA. Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen Streptococcus gordonii and its complex with the mechanism-based inhibitor NAG-thiazoline. J Mol Biol. 2008; 377: 104–116. doi: 10.1016/j.jmb.2007.09.028 PMID: 18237743

32. Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B et al. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc Natl Acad Sci U S A. 2012; 109: 20889–20894. doi: 10.1073/pnas.1212034109 PMID: 23213210

33. Drickamer K, Fadden AJ. Genomic analysis of C-type lectins. Biochem Soc Symp. 2002;59–72. PMID: 12655774

34. Ramasubbu N, Thomas LM, Ragunath C, Kaplan JB. Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen Actinobacillus actinomycetemcomitans. J Mol Biol. 2005; 349: 475–486. PMID: 15878175

35. Nurizzo D, Nagy T, Gilbert HJ, Davies GJ. The structural basis for catalysis and specificity of the Pseudomonas cellulosa alpha-glucuronidase, GlcA67A. Structure. 2002; 10: 547–556. PMID: 11937059

36. Mark BL, Wasney GA, Salo TJ, Khan AR, Cao Z, Robbins PW et al. Structural and functional characterization of Streptomyces plicatus beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. J Biol Chem. 1998; 273: 19618–19624. PMID: 9677388

37. Gravel RA, Clarke JTR, Kaback MM, Mahuran D, Sandhoff K, Suzuki K. In: The metabolic basis of inherited disease,(Eds. Scriver CV, Beaudet AL, Sly WS, and Valle D). New York: McGraw-Hill, 1995. 2839–79.

38. Mahuran D, Novak A, Lowden JA. The lysosomal hexosaminidase isozymes. Isozymes Curr Top Biol Med Res. 1985; 12: 229–288. PMID: 3886595