Biotechnological production of galactosides of pharmaceutical interest: enzyme screening, engineering and application

Estela Castilla Ypas

UNIVERSITAT RAMON LLULL

DOCTORAL THESIS

Title	Biotechnological production of galactosides of pharmaceutical interest: enzyme screening, engineering and application
Presented by	Estela Castilla Ypas
Centre	IQS School of Engineering
Department	Bioengineering
Directed by	Dr. Antoni Planas and Dra. Teresa Pellicer

C. Claravall, 1-3 | 08022 Barcelona | Tel. 93 602 22 00 | Fax 93 602 22 49 | info@url.edu | www.url.edu

Per tots aquells que ho han fet possible.
I especialment per als meus pares. Eloi i Rosa, va per vosaltres.

"…amar la trama más que el desenlace"

Acknowledgments

Un cop s'arriba al final d'una etapa és important mirar enrere, fer balanç i recalcar-ne allò més important. Després d'aquests 4 anys de doctorat he après infinitat de coses i puc dir que he crescut com a científica i com a persona. Aquest apartat intenta agrair i recalcar tot allò que he guanyat en aquesta época.

Primer m'agradaria agrair al Pla del Doctorat Industrial (Agaur i Generalitat de Catalunya) per atorgar-nos aquesta beca. Ha estat una fantàstica oportunitat per conèixer el món de la recerca a la universitat i compartir-ho amb la indústria. També vull agrair a IQS i a Interquim S.A haver-me permès moure'm entre els dos mons amb tanta facilitat i comoditat.

Agrair al Dr. Antoni Planas, director d'aquesta tesi, confiar en mi i donar-me la possibilitat de realitzar aquest doctorat en el seu laboratori. Gràcies per les discussions i crítiques, és un plaer aprendre tanta ciència d'algú com tu. Gràcies per fer-me descobrir que el món dels carbohidrats pot donar més energia del que sembla!

Agrair a la Dra. Teresa Pellicer, directora d'aquesta tesi, primer de tot per confiar en mi: tan en l'inici com en el llarg de tota la tesi. Gràcies per tota la ciència que has compartit i per la teva manera tant instructiva d'ensenyar. També m'has ensenyat a veure que més enllà de la ciència, hi ha *punts i comes* a tindre en compte. Moltes gràcies per fer-me veure molt més que números i enzims!

També vull agrair al Dr. Xavier Camps i al Dr. Francisco Marquillas permetre'm conèixer Interquim i el projecte gaxilosa des de ben aprop. Gràcies per confiar en mi i en la biocatàlisi. Gràcies per tant gran oportunitat. També vull donar les gràcies al grup del projecte gaxilosa amb els que he tingut la sort de treballar: ha estat un plaer treballar amb vosaltres i veure la investigació des dels ulls de la indústria. Part de lo aconseguit en aquest treball és vostre! Fede, gracias por estar siempre 100% disponible, sé que al final le has cogido cariño a las enzimas. Santos, gracias por tu ayuda, por ensenyarme la parte más industrial del proceso. He aprendido muchísimo en las visitas a los laboratorios de Interquim. Cristina, gràcies per la teva visió analítica: els teus anàlisis han fet palpable els èxits del projecte. També m'agradaria agrair a la Paqui totes les gestions. Moltes gràcies per fer-ho tot TAN fàcil!

Thanks to Dr. Charles Tellier to allow me entering to your laboratory and work with you and your team. It was a wonderful opportunity and a great experience. Merci beaucoup pour tout!

Gracias, también, al Dr. Alberto Marina por darme la posibilidad de acercarme al complicado, pero apasionante, mundo de la cristalografía. Fue todo un reto para mí! Gracias por todos los resultados compartidos!

També vull agrair al tribunal haver acceptat llegir i avaluar aquest treball. Espero que sigui una lectura agradable i en disfruteu tan com jo.// Thanks to the jury for accepting the evaluation of this work. I hope you a nice lecture and I hope you enjoy it as I much as I did.

Donar les gràcies a la Dra. Magda Faijes. Podria dir que el doctorat va començar amb tu durant el master en bioenginyeria. Gràcies pel teu entusiasme per la ciència i les ganes de saber més. Durant el master vas transmetre les ganes d'apendre i descrobir que sempre tens, i ho has mantingut durant aquest 4 anys on hem compartit laboratori, seminaris i reunions. Encara recordo aquell dinar on, crec, vaig començar a buscar doctorats. Moltes gràcies!

Gràcies al Dr. Xevi Biarnés. Agrair la teva entrega del 100% a bioquímica. Saps que això no seria el mateix sense tu i els teus ulls computacionals. Gràcies per compartir les teves idees i punts de vista. T'agraeixo les aportacions que has fet en aquest treball. Al teu costat sempre se n'aprèn moltíssim!! Moltes gràcies!

Agrair a diferents investigadors amb els que he pogut compartir hores al laboratori i aprendre, cada dia, de cada un de vosaltres: Amanda (gràcies per fer-nos començar el dia sempre amb un somriure), Dra. Marta Palomo (gràcies per demostrar que amb ganes, tot es pot), Dr. Xavi Pérez (gràcies per compartir tants coneixements d'anàlisi!), Dr. Edu Andrés (gracias por tus consejos en el mundo glicosintasa), Dr. Régis Fauré (thanks for you help during your stay, your desire to collaborate and share science. And thanks for the science you share with us, it was very helpful. Merci beaucoup!). Així com investigadors d'altres departaments com la Dr. Gemma Gotor i la Dra. Cristina Ribas, per ajudar sempre amb els problemes cromatogràfics!

Als diferents professors i col·laboradors de bioquímica: Dr. Pau Leiva, Dr. Marc Carnicer i Patri. Han estat 3-4 anys amb vosaltres compartint seminaris, pica-pica, reunions i dinars. Ha estat un plaer treballar al vostre costat. Gràcies per la vostra ajuda i idees! He après molt de vosaltres!

Que seria de quaselvol doctorat sense els cafés al bar de l'IQS... moltes gràcies a tot l'equip per fer d'aquests *breaks* una festa continua!! Per descomptat també agrair al personal d'IQS, recepció, reprografia, manteniment i magatzem, la disposició a solucionar imprevistos i resoldre embolics. Gràcies per fer el dia a dia més fàcil! I també més entretingut!

Durant aquests 4 anys he tingut la sort de compartir laboratori amb un munt gent, i de cada un d'ells me n'he quedat alguna cosa, amb o sense el seu permís. Cris (Dra. Val) fuiste un ejemplo a seguir, y demostraste que la maña siempre puede más que la fuerza. Aprendí infinidad de ciencia contigo, pero también me hiciste ver que no hay nada más importante que las ganas de seguir adelante pase lo que pase. (y siempre con un toque de humor). Sergi (Dr. Abad) ets un crack!! Ets treballador, persistent i mai deixes de tirar endavant. Sempre cuidaves dels "petits del lab" (que al teu costat ho erem!!) i per damunt de tot, sempre vas mantindre un somriure sense fi. I aquí ho vull deixar escrit: moltes gràcies! Per molts més berenars! Dr. Huguinho, te conocí con una glicosintasa un tanto peculiar (E134D) con la que aprendí muchísimo, y aunque escogiste el mundo nano y sus loops ha sido un placer compartir tiempo y ciencia contigo. Siempre recordaremos tus filloas y clases de swing. Pero que sepas que aún esperamos LA tarta de Santiago. Dra. Victoria!! Y lo digo así, con exclamaciones. Has sido, siempre, una inyección de energia y positivismo! Tu amor por la ciencia y por la vida es contagiosa! Mil gracias! Dr. Enea (Gibson), vam coincidir gaire bé un any al laboratori i cada dia tenies alguna sorpresa. Gràcies per la teva manera innocent de mirar les coses! És d'admirar! Quina sort de postdoc que vam tindre! També vull donar les gràcies a altres doctorands de la promoció anterior, amb els que vam coincidir menys però també vam compartir molts moments: Ellen moltes gràcies per la teva dolçor i amabilitat. Carles les tardes amb crispetes sempre seran fetes per tu! Marta Vilà, tot i la distància amb el Vall Hebron! Y Javi Romero, gracias por aguantar nuestros seminarios 100% experimentales y por ayudarnos siempre con el mundo bioinformático.

A l'Àngel Martínez, TFG, gràcies per triar el projecte gaxilosa (2 cops!) Gràcies per tot el que em vas ajudar i tot el que has aportat a aquest treball. També vull agrair-te els moments divertits i de *relax* al mig del lab (que sempre, sempre, són d'agrair).

També vull donar les gràcies a estudiants que han passat pel laboratori durant aquest temps i he pogut compartir discussions científiques i moments molt divertits. TFGs com l'Ignasi (gràcies per la teva calma i manera de mirar les coses tan pràctica), Sara (gràcies per portar sempre un somriure 100% sincer) o el Juan. I a TFMs com la Siwen (gracias por mostrarnos que todo es compatible con dormir y comer, hasta las simulaciones!), Oriol (gràcies per les millors truites de patates), Pau Jané (gràcies per la música DJ!!), Miquel Àngel (gràcies per demostrar-nos l'efecte de la cafeïna), Gerard (gràcies per les teves aturades, que sempre van ser molt inspiradores), Marcos, Klaus (gràcies pels millors postres en els pica-picas!), Olivia (tot i que vam coincidir molt poc, va ser un plaer compartir poyata!). Estudiants de més llarga estada com la Nuria Abajo, gracias por estar siempre de buen rollo. Gracias por tus ganas de ayudar y aprender

siempre. O a estudiants de més curta estada com l'Stefan i la Chiara, thanks for share your expierience with us!. També m'agradaria donar gràcies a la Xisca. Vas estar molt poc temps, però ens vas donar aire i alegria quan bioquímica ho necessitava més!

Vull donar les gràcies també als veïns de biomaterials i teixits: Sejin, Cris Castells, Pere, Pau, Alba, Mireia Alemany, Caterina, Lourdes, Biel, Anna Reyes, Pol; els moments compartits al voltant del laboratori també són molt importants! Moltes gràcies!

També vull agrair al ja Dr. Jordi Donderis la teva paciència amb el monstre que és la galactosidasa amb la que hem treballat. Gràcies per deixar-me entrar al món de la cristal·lografia i, sobretot, gràcies per la feina que vas fer amb la galactosidasa: ara ja sé que és traçar l'estructura d'un monstre de proteïna. I would like to thank the students from Nantes for all expieriences and science shared: Ennys (thanks for french and english translation, merci!!), Florian (thanks for so many coffee breaks), Maruthy (thanks for share your love for India and indian food), Lydia (thanks for so many many PCRs!!), Benoit, Iyanar, Johann, Johann Dion, Dr. Charles Dion, Dr. Emily, Dr. Amelie, Dr. Cristoph. Thank you and merci beaucoup! I hope to see you soon!

L'última part del doctorat l'he viscuda amb vosaltres, i vull que sapigueu que les paraules se'm quedaran curtes: sou genials! Heu fet d'aquest doctorat una aventura molt més enriquidora! Gràcies per compartir tants bons moments, tants patiments i tantíssimes alegries. Mire, portes alegria i bon humor per allà on vas i converteixes qualsevol situació en un show: la guerra dels fags convertida en *Fag-busters*, els teus atacs de tos mortals o els esternuts sempre triples et fan única. Vas ser un fitxatge genial per bioquímica i pel que comenten per Barcelona!! NurNur, he aprés molt al teu costat (més del que et reconeixeré, =P). Ningú té les coses tan clares com les tens tu. Gràcies per ensenyar-me el teu cor, que és més gran del que dius!! Recorda que tot va començar sent companyes de taula! Sergi, gràcies per deixar-te conèixer. Gràcies per compartir ciència i les teves preocupacions, ha estat un plaer treballar al teu costat. Els atacs de riure amb tu són incotables i els gintònics també!!! Gràcies per cuidar tant a les *nenis* del lab, i per acabar sent una de nosaltres! Almu, vas ser una sorpresa per nosaltres. Vas arribar un dia a bioquímica i mai vas marxar. Ha estat un plaer treballar amb tu i conèixer-te: la teva alegria i manera d'escoltar i aconsellar és d'admirar. Espero que tornem a viure algun cop a la mateixa ciutat!! Gràcies per tot! Cris, la teva manera de treballar i ajudar als altres és espectacular. Gràcies per ajudar, sempre, a tothom! Gràcies per deixar-me conèixer-te i poder compartir moments tan especials per Barcelona, Sant Cugat, Nantes o Gante! I a partir d'ara on sigui! No deixis mai de ballar! Laia, gràcies per aparèixer a bioquímica i convertir-te en un gran pilar. Has sigut més que genial poder compartir tant amb tu: dins i fora del lab! Les experiències i

moments compartits són incontables (i no estic exagerant!), i sé que no pararan! Aventures, viatges, confessions, consells i riures. Gràcies per tot! Gràcies per cuidar tant a tota la gent que tens al teu voltant! Us trobaré molt a faltar! S'ha acabat una etapa genial al vostre costat, però... who runs the world??

Masters, el viatge va començar amb vosaltres, i tot i que estem bastant esparcits per Europa... sempre sabem com retrobar-nos i confio en que ho farem per molts anys més! I pel que sembla la familia va creixent! Benviguda Olivia! També vull dir gràcies a les meves doctores alemanyes preferides: Susanna i Anna Maria, tot i a la distància sempre esteu aquí, gràcies per compartir tantes experiències i per les ganes que sempre teniu d'aventures i viatges. Gràcies també per buscar la platja a Lleó i el sol a Alemanya!

També vull agrair als amics de sempre la paciència que heu tingut durant aquest 4 anys, per intentar entendre el món del doctorat i sobretot per donar-me sempre, sempre suport. Gràcies... Mil gràcies!! Esteu repartits per Horta, Carmel, Gracia, Sagrada Familia, Encants, Paral·lel o Pamplona, Sant Celoni i rodalies, Sant Quirze, Sant Cugat, Sants, London, Berlin, BS...!! Sense vosaltres tampoc hagués estat possible!

Gràcies a la Isa i Josep Maria, que són els avis postissos que més em cuiden! Així com a la Mercé, Manel i Maria. Gràcies per fer-me sentir una més i cuidar-me tant!

Ignasi i Cris, gràcies donar-me tantes alegries aquests últims anys! I sobretot, gràcies per fer-me tieta! Ignasi, ets el millor germà gran!!!

Gràcies a la música: "*thanks for all the joy you're bringing. Who can live without it, I ask in all honesty. What would life be? Without a song or a dance what are we? So I thank you for the music. For giving it to me*." Perquè, crec que aquest treball no hagués estat igual sense la música, capaç de canviar qualsevol situació: donar força quan cal, aire en el moments més enterbolits i vestir-ho tot de festa en les millors ocasions.

I amb més intensitat als meus pares, Eloi i Rosa. Gràcies per deixar-me viure aquesta experiència! Gràcies pel vostre recolzament sense fre, passi el que passi! Mil gràcies per tot el que em doneu, per confiar sempre amb mi i fer-me veure que la vida s'ha de gaudir, però també lluitar: sou un exemple a seguir. Només puc dir, moltes gràcies!!!

Manel, tot i la distància has viscut el doctorat de ben prop, i puc dir que una part és teva. Gràcies per entendre'm, cuidar-me i donar-me aire quan més ho necessitava. El teu recolzament, sobretot a la part final, ha estat essencial. Gràcies per ensenyar-me que cal lluitar pel que volem de veritat. I sobretot, gràcies perquè vas donar sentit a tot, quan res semblava que ho tenia. Mil gràcies de tot cor per donar-me tant i compartir tantíssims somnis i experiències! Ara comença una nova aventura!! (*"dejarse llevar..."*)

I gràcies a tu que, per un motiu o altre, estàs llegint aquest treball: espero que el gaudeixis.

The research leading to these results has received funding from Generalitat de Catalunya Doctorat Industrial Programme (2013 DI 034).

SUMMARY

Lactose is a milk disaccharide produced by nearly all mammalian species. For its digestion, it must be hydrolysed to galactose and glucose by lactase (EC 3.2.1.23), which is normally produced by the cells that line the small intestine. Deficiency or low levels of lactase (hypolactase) can cause common symptoms including bloating, abdominal pain or cramps, flatulence and diarrhea. Evaluation of enzyme deficiency is important in pediatrics and gastroenterology due to the high frequency of genetic predisposition (65%). Many of the standard diagnostic procedures are invasive, quite drastic and not applicable to infants and young childen. A non-invasive method was developed based on the use of 4-*O*-β-D-galactopyranosyl-D-xylose (gaxilose), a structural analogue of lactose (it only lacks the hydroxymethyl group at position 5). This compound is substrate of the lactase enzyme *in vivo*, yielding D-galactose and D-xylose. The latter is passively absorbed from the small intestine and is eliminated in the urine where it can be quantified by a colorimetric procedure.

Chemical synthesis of gaxilose requires the addition of protective groups and suffers from long, tedious reaction sequences with low overall productivity (9%). The biosynthetic route involving galactosyltransferase enzymes implies technical difficulties (unstable and expensive enzymes) and high costs. Glycosidases have been shown to catalyse the formation of glycosides by transglycosylation at low cost. Disadvantatges of this approach include yield limitations due to competing hydrolysis reactions, and purification problems because of the formation of other regioisomers and by-products.

Escherichia coli β-galactosidase was selected to produce gaxilose at industrial level. After an enzymatic reaction and several purification steps, a yield of 20-23% is reached.

The present work aims to increase gaxilose production yield by different strategies. First, enzymatic conditions were modified to increase enzymatic gaxilose production. On ther other hand, other enzymes were searched to be used as gaxilose biocatalysts. Bibliographic and experimental studies allowed to find one enzyme able to increase gaxilose yield up to 35% using 3.3-fold less enzyme, and thus diminishing the production costs). Despite the higher transglycosidase activity of the new enzyme, its hydrolase activity did not allow the increase of gaxilose prodution. Protein engineering (random and rational approaches) was used to modify the enzyme activity and increase gaxilose synthesis.

SUMARI

La lactosa és un dissacàrid de la llet produït per gaire bé tots els mamífers. Per a la seva digestió cal que sigui hidrolitzada per la lactasa (EC. 3.3.1.23), que és produïda per les cèl·lules epitelials de l'intestí prim, donant lloc a galactosa i glucosa. La seva deficiència o baixa quantitat (hipolactàsia) pot produir diversos símptomes incloent inflor, dolor abdominal, flatulències i diarrea. L'avaluació de la deficiència de la lactasa és principalment important en pediatria i en gastroenterologia a causa de l'elevada incidència de l'alteració genètica de la lactasa en aquests grups poblacionals (65%). La majoria dels mètodes de diagnòstic són invasius i dràstics i no són adequats per a ser aplicats a infants. Davant d'aquesta situació, un nou mètode no invasiu va ser desenvolupat. Aquest es basa en l'administració de 4-*O*-β-D-galactopiranosil-D-xilose (gaxilosa), un anàleg estructural de la lactosa. Aquest compost també es substrat per la lactasa i és hidrolitzat donant lloc a galactosa i xilosa. Aquest segon és absorbit passivament per l'intestí prim i és eliminat per orina on pot ser fàcilment detectat per un mètode colorimètric senzill.

La síntesi química de la gaxilosa necessita de l'ús de grups protectors i de llargs i complexes etapes sintètiques per arribar a rendiments de producció baixos (9%). La ruta biosintètica utilitzant galactosiltransferases implica diverses dificultats tècniques i costos elevats. Les glicosidases tenen la capacitat de formar enllaços *o*-glicosídics per transglicosidació amb costos baixos. Els desavantatges d'aquesta metodologia són els baixos rendiments limitats per la competència de l'activitat hidrolítica i els problemes de purificació per la formació de regioisòmers o altres productes.

La β–galactosidasa d'*Escherichia coli* va ser seleccioionada entre altres enzims per la producció de gaxilosa a nivell industrial. Després de diversos passos de purificació el rendiment obtingut és del 20-23%.

Aquest treball pretén augmentar el rendiment en la producció de gaxilosa. Primer, les condicions de reacció es van modificar per tal d'augmentar l'activitat enzimàtica de transglicosidació. D'altra banda, es va buscar un altre enzim per ser utilitzat en la producció de gaxilosa. Estudis bibliogràfics i experimentals van permetre la selecció d'un nou enzim capaç de sintetitzar més gaxilosa, arribant a un rendiment final del 35% utiiltzant 3,3 cops menys enzim (amb la corresponent disminució dels costos de producció). Tot i que el nou enzim presentava una activitat de transglicosidació superior a l'enzim d'*E. coli*, la seva activitat hidrolítica romanent no permetia augmentar-ne més el rendiment. Per modificar l'activitat enzimàtica i augmentar la síntesi de gaxilosa es va optar per utilitzar modificar l'enzim amb enginyeria de proteïnes (racional i aleatòria).

SUMARIO

La lactosa es un disacárido de la leche producido por casi todos los mamíferos. Para su digestión debe ser hidrolizado per la enzima lactasa (EC. 3.3.1.23), que es producida por las células epiteliales del intestino delgado, dando lugar a glucosa i galactosa. Su deficiencia o baja concentración (hipolactasia) puede producir diversos síntomas como son hinchazón, dolor abdominal, flatulencias y diarrea. La evaluación de la deficiencia de la lactasa es importante en pediatría y gastroenterología por la elevada frecuencia de esta alteración genética (65%). La mayoría de los métodos de diagnóstico son invasivos y drásticos y no son adecuados para su uso en población infantil. Frente a esta situación, se desarolló un nuevo método no invasivo. Éste se basa en la administración de 4-*O*-β-D-galactopiranosil-D-xilose (gaxilosa), un análogo estructural de la lactosa. Este compuesto también es sustrato de la lactasa y es hidrolizado dando lugar a galactosa y xilosa. El segundo se absorbe pasivamente a través del intetsino delgado y se elimina a través de la orina donde puede ser detectado por un método colorimétrico senzillo.

La síntesi química de la gaxilosa requiere del uso de grupos protectores y de largos y complejos pasos de síntesis para llegar a rendimientos de producción bajos (9%). La ruta biosintética utilizando galactosyltransferasas comporta dificultades técnicas y costes elevados. Por otro lado, las glicosidasas tienen la capacidad de síntesis de glicósidos por transglicosidación con costes bajos. La principal desventaja de esta metodologia son los bajos rendimientos por la actividad hidrolítica de la enzima y los problemas de purificación por la formación de regioisómeros y/u otros productos.

La β-galactosidasa de *Escherichia coli* fue seleccionada, entre otras enzimas, para la producción de gaxilosa a nivel industrial. Después de varios pasos de purificación el rendimiento es del 20-23%.

Este trabajo pretende aumentar el rendimiento de la producción de gaxilosa. Primero se modificaron las condiciones de reacción para aumentar la activitat enzimática de transglicosidación. Por otro lado, se buscó otra enzima para la producción industrial. Estudios bibliográficos y experimentales permitieron la selección de una nueva enzima capaz de sintetitzar más gaxilosa, llegando a un rendimiento del 35% utilizando 3,3 veces menos enzima (con la consecuente disminución de los costes de producción). Aunque la nueva enzima presentaba una actividad de transglicosidación mayor, la actividad hidrolítica remanente no permite aumentar el rendimento. Para modificar la actividad enzimática y aumentar la síntesis de gaxilosa se decidió modificar la enzima mediante ingeniería de proteínas (racional y aleatoria).

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa - Amino acids
Abs - Absorbance
ACN - Acetonitrile
Ala - Alanine
A_s – Peak symmetry
Ara – L-arabinose
Arg - Arginine
Asp - Aspartate
BCA - Bicinchoninic acid
CAZy - Carbohydrate Active enZymes
CLD – Congenital lactase deficiency
D - Domain
DNA - Deoxyribonucleic acid
DMF - Dimethylformamide
DMSO – Dimethyl sulfoxide
E – Enzyme
EK - Enterokinase
ESI – Electrospray ionization
Fuc - Fucose
Gal – Galactose
Gax or Gaxilose - 4-*O*-β-D-galactopyranosyl-D-xylose
GC- Gas chromatography
GH - Glycosidases
Glc - Glucose
Glu- glutamate
Gly - glycine
GOS – galacto-oligosaccharides
GT - glycosyltransferases
HILIC – Hydrophilic Interaction Chromatography
HPLC - High Performance Liquid Chromatography
HPSEC – High Performance Size Exclusion Chromatography
H – hydrolase activity
His - Histidine
HTS – Highthroughput Screening
IPTG – Isopropyl β–D-1-thiogalactopyranoside
K – Retention factor or time retention
LB – Luria Bertani
LOD – Limit of detection
LOQ – Limit of quantification
LPH – Lactose-phlorizin hydrolase
LTT – Lactose tolerance test
LTTE - Lactose tolerance test with ethanol
M – Mass
Man - Mannose
MCS – Multicloning site
Met - Methionine
MS – Mass spectrometry
M_w – Weight average molecular mass
OD_{600} - Optical Density at 600 nm
***o*NP –** *ortho*- nitrophenol
***o*NPG or *o*NPGal –** *ortho*-nitrophenyl-β-galactoside
P – Product
Phe - Phenylalanine
Pro - Proline
Protein Data Bank – PDB
PMSF – Phenylmethylsulfonyl fluoride
***p*NP –** *p*-nitrophenyl
RID – Refractive Index Detector
R_s - Resolution
RNA – Ribonucleic Acid
S – Substrate
Ser - Serine
SDS-PAGE – Sodium Dodecyl Sulfate-PoliAcrylamide Ge Electrophoresis
SCFA – Short Chain Fatty Acids

LIST OF ABBREVIATIONS

T – transglycosidase activity

T/H ratio – Transglycosidase to hydrolase ratio

TEA - Triethhylamine

THF - Tetrahydrofuran

TLC – Thin Layer Chromatography

Trp - Tryptophan

Tyr – Tyrosine

U - Units

UV – Ultravioler

VMD – Visual Molecular Dynamics

Wt - wild type

X-gal – 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Xyl - Xylose

INTRODUCTION

i.1 Lactose

i.1.1 Milk sugars and its commercial uses

Lactose is the main source of sugar in all mammals' milk (42-70 g/L) [1], except for the sea lion [2]. Human milk contains 7% of lactose, which is the most important energy source during the first year for human life [3]. Lactose concentration in milk varies widely between species (e.g. <1% in dolphins, 5% in domestic ruminants, 10% in green monkeys[4]). Moreover, the milk of a particular species varies with the breed, health, nutritional status, stage of lactation or age.

Lactose has no flavour and low sweetness (30% compared sucrose). Therefore, it is widely used to reduce the overall sweetness in confectionery and other products. In addition, lactose is a reducing sugar that enhances colour and flavour generation in many bakery products via Maillard and Strecker reactions (Box i. 1). In food and confectionery industries, lactose is commonly used as a filler or flavour carrier [5]. When properly crystallized, lactose has a low hygroscopicity, which makes it an attractive sugar for use in icing for the food industry [6].

Box i. 1 *Maillard reaction*

Maillard reaction involves reaction between an aldehyde (from reducing sugars) and an amino group (in foods, mainly NH_2 group of lysine in proteins) at high temperature. The Maillard reaction has desirable consequences in many foods (e.g. coffee, bread, crust, toast, french fried potato products) and it is responsible for many colours and flavours in those foods. However, its consequences in milk products are negative (e.g. brown colour, off-flavours, slight loss of nutritive value (lysine), loss of solubility) [395–397].

Nutritionally, lactose promotes the assimilation of calcium [7,8] and it has been used in numerous nutritional products. One of the main applications of lactose is in the preparation of infant formula. Moreover, it is also used in pharmaceutical industry: in the manufacture of pharmaceutical tablets and capsules as an excipient, a carrier for inhalation aerosols [9] or dry powder inhalers.

i.1.2 Lactose synthesis

Lactose is a disaccharide made up of glucose and galactose bonded by a β-1,4 linkage. It is formed within the Golgi apparatus in the epithelial cells of mammary glands [10].

It is synthesised from glucose, which is activated and isomerized to UDP-galactose via the four-enzyme Leloir pathway (Figure i. 1). UDP-Gal is then linked to another molecule of glucose in a reaction catalysed by the enzyme lactose synthase.

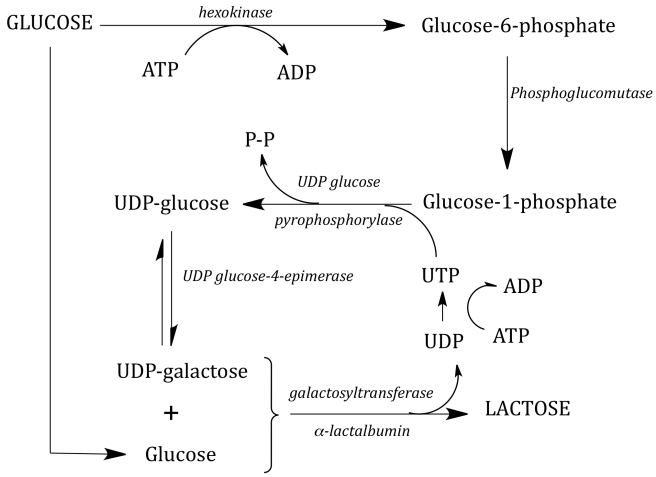


Figure i. 1. *Pathway for lactose biosynthesis*

Lactose is the major dietary source of galactose and it ensures that galactose levels do not become limiting during postnatal early development. Galactose, and its by-products, has a central role in eukaryotes in the biosynthesis of complex carbohydrates, glycoproteins and glycolipids. In humans, galactose is particularly important in early development and it has crucial structural function, for example, for the brain [11,12]. Additionally, the sugars present in milk are a determinant factor in neonatal host defense and inflammatory processes due to their prebiotic effect, and are an important source of energy in infants [13,14]. Genetic disorders that impair its metabolism inevitably cause disease, drawing attention to its key role. It has been studied that deficient activity of any of the enzyme of galactose metabolism (Leloir pathway) leads to galactosemia: a potentially lethal disorder during the neonatal period [14,15]. In addition to its broad role in human physiology, galactose has been recently reported as beneficial for the prevention of a number of diseases, particularly those affecting brain functions [16,17].

i.1.3 Lactose hydrolysis

Lactose hydrolysis into monosaccharides, needed for the correct absorption through intestinal mucosa, takes place in the first part of small intestine via the enzymatic reaction catalysed by the lactase enzyme (lactase-phlorizin hydrolase, LPH, EC 3.2.1.23/62) (Box i. 2). Lactase is an integral membrane glycoprotein of the intestinal brush border membrane, anchored into the membrane by its C-terminal end with the bulk of the enzyme projecting into the lumen of the gut. This enzyme is expressed by enterocytes or absorptive cells of the small intestine, being expressed at its highest level in the mid-jejunum (second part of the small intestine). The enzyme presents two

hydrolytic activities: phlorizin hydrolase, that digests β-glycosylceramides, and lactase, that hydrolyses lactose [18].

Box i. 2. *Gene and protein of Lactase*

Lactase is encoded by a single gene (*LCT*) of 50 kbp on chromosome 2 [398]. The gene has 17 exons and encodes a pre-protein of 1,927 amino acid residues. The protein is composed of five domains: 1) cleaved signal of 19 amino acids; 2) a large *pro* portion of 847 residues, those do not appear in mature LPH; 3) the mature LPH with two catalytic sites; 4) a membrane-spanning hydrophobic segment near C-terminus which serves as membrane anchor; and 5) a short hydrophilic segment which must be cytosolic [399].

i.2 Lactose intolerance

Most mammals cease to produce lactase after weaning. This phenomenon is known as lactose non-persistence (also referred as adult-type hypolactasia or lactase restriction). The reduction of lactase activity begins between 2 and 3 years and is complete by the age of 5 to 10 years, in most mammals [19,20]. In humans, reduction of lactase activity after infancy is genetically programmed. Approximately 65% of world population is lactose intolerant and the statistics vary from country to country (Figure i. 2).

Reduction of the expression or the activity of intestinal lactase allows lactose to reach the large intestine since intestinal epithelium does not have mechanisms to uptake disaccharides or higher carbohydrates. Lactose in the colon is fermented by gut microflora leading to the production of short chain fatty acids (SCFA) and gas (H_2, CO_2, CH_4). In addition, the presence of lactose in the colon has an osmotic effect, drawing water in from the blood. Products of bacterial digestion can cause discomfort with symptoms as diarrhoea, gas bloat, flatulence and abdominal pain [21]. These products can also cause systemic symptoms such as headaches, loss of concentration or heart arrhythmia [22]. Lactose intolerance depends not only on the expression of lactase but also on the intestinal flora, gastrointestinal motility, dose of lactose, sensitivity of the gastrointestinal tract to gas. Only when lactose malabsorption is associated with clinical symptoms (bloating, flatulence, abdominal pain and diarrhoea) lactose intolerance can be diagnosed

Reduction of lactase activity may be caused by different mechanisms:

- Congenital loss (congenital lactase deficiency, CLD). This is a rare autosomal recessive disease where enzymatic activity is absent or reduced [23]. The most common genotypes identified in CLD patients include premature stop codons and truncated protein as a result of frame shifts, missense mutations in the coding region or exon duplication [24–28].

- Secondary lactase deficiency. This lactase deficiency that results from small bowel injury, such as acute gastroenteritis, persistent diarrhoea, cancer chemotherapy, or other causes of injury to the small intestinal mucosa.
- Adult type lactase deficiency is common all over the world and it is based in the decline of lactase activity after weaning. The rate of adult type lactase deficiency varies among ethnic groups (Asia 80-100%, Africa 70-95%, USA 15-80%, Europe 15-70%). It is an autosomal recessive condition resulting from a developmentally regulated change of the lactase gene product, responsible for reduced synthesis of the precursor protein. The persistence or non-persistence of the expression of LPH is associated with polymorphism C/T 13910. It is one substitution in the sequence that regulates the gene: while genotype CC correlate with hypolactasia, genotype TT with lactase persistence [26,28–32].

Lactose intolerance is managed through the avoidance of lactose containing food. Additionally, oral administration of β-galactosidase represents a possible therapeutic approach (e.g. Nutira ®, http://nutira.es).

i.2.1 Lactase persistence

Despite the common cease of lactose production after weaning, about 35% of people in the world continue to produce lactase throughout adulthood and thus are able to digest lactose without discomfort [31]. Humans are the only species on the planet that drink milk from other species. Since lactase's only function is the digestion of lactose in milk, most mammalian species reduce that activity after weaning. Lactase persistence in humans has evolved as an adaptation to the consumption of non-human milk and dairy products consumed beyond infancy. During human evolution, some individuals experienced a mutation in the LTC gene that allowed to digest lactose when they became adults.

The global distribution of lactase persistence (LP) is now fairly well characterised (light blue in Figure i. 2) [31–33]. Continued production of lactase throughout adult life is found at moderate to high frequencies in north Europe, North-centre of Africa, Middle Eastern and Southern Asia.

Figure i. 2. *Interpolate map of phenotype frequencies. (Lactose intolerant in dark blue, lactose intolerance in grey)*

The correlation of lactase persistence phenotype with the cultural practise of milking has raised the hypothesis that this trait has been subject to strong positive selection [34,35]. In the 1970s it was established that LP has a genetic cause and it is inherited in an autosomal recessive manner [36,37]. In the early 1980s, it was shown that it is caused by a *cis*-acting element [38–40]. Further investigation of the molecular mechanism as well as the evolutionary forces is however needed to fully understand this variation, which is providing an important model for understanding gene/culture co-evolution and disease susceptibility. It has become clear that there are multiple, independently derived LP-associated alleles with different geographical distributions [41–43].

i.2.2. Diagnostic tests

Hypolactasia is a common condition, which causes unspecific symptoms. It cannot be diagnosed on the basis of symptoms alone, and patients do not always link their symptoms with the consumption of dairy and other lactose-containing products. It is therefore important to have simple and convenient methods to diagnose this disorder.

Several methods have been developed to evaluate lactose malabsorption [21,44–46]. They can be classified as invasive or non-invasive methods:

a) Invasive methods:

- MUCOSAL BIOPSY from the duodenum has been the reference standard test for years. It is based on direct determination of the lactase activity of the small intestinal mucosa [46]. The biopsy sample is divided in two

parts: one part for microscope and histological examination and the other for disaccharidase assay. The assay measures glucose formed after sample incubation with lactose. At the same time, maltase and sucrase activities are evaluated with maltose and sucrose as substrates [46–48]. Hypolactasia is diagnosed when lactase to sucrase ratio is less than 0.3 and lactase activity is less than 10 IU/g protein. A major advantage of disaccharidase assay is that it provides a direct answer to the question as to whether the patient has hypo- or normolactasia; moreover, it allows examination of the mucosa to discard its damage. However, it is an invasive method that requires special laboratory equipment as X-ray control and there is also a risk of serious complications during the procedure. Moreover, non-uniform distribution of lactase and possible damage caused by infection or drugs can lead to false positives or negatives. The exact site where the sample is taken is also an issue to take into account: different parts of the mid-jejunum express different enzyme amounts.

- GENETIC TESTS. The genetic analysis identified a single nucleotide polymorphism that shows complete association with the lactase non-persistence/persistence and is characterized by a C to T change (CC-associated with adult hypolactasia, TT with lactase persistence) [44,49–51]. Genotyping of lactose tolerance test has limitations, because a positive genetic test indicates decline in lactase activity but does not inform about clinical aspects [52]. Even in children, genotyping could only be used as a rule-out test since it does not release information about the age at which a child with CC genotype begins to decrease lactase expression.

b) Non- invasive methods:

- MEASUREMENT OF H_2 after lactose administration has been used for more than 30 years to diagnose lactase deficiency. Lactose (25-50 g) is administrated to patients and if it is not absorbed, it reaches the caecum where is fermented by bacteria and H_2 is produced. A portion of the gas diffuses into the circulation to be carried to the lungs and exhaled. The test is considered positive if the H_2 concentration in the exhaled air exceeds 20 ppm [53–56]. It is one of the most used methods due to its non-invasive methodology; nevertheless some factors can produce false negatives or false positives because of conditions affecting the intestinal flora, lack of hydrogen-producing bacteria, intestinal motility or slow

intestinal bacterial growth [45]. Its main drawback is the discomfort in lactose intolerance patients after the administration of a large dose of lactose.

- EVALUATION SYMPTOMS (nausea, blotting, diarrhoea, abdominal pain) developed after ingestion of lactose (i.e. 50 g) [54].
- LACTOSE TOLERANCE TEST (LTT) is based in the measurement of the increase in blood glucose after lactose ingestion. After a standard dose of 50 g lactose, blood samples are taken at intervals of 15 to 20 min. It has been recommended that samples are taken in those intervals, especially in children, because glucose is rapidly absorbed and utilized. It is commonly used because the test only needs blood glucose determination, which is available nearly everywhere. Nevertheless the method is not very reliable in diabetic patients and its results can be misleading due to hormonal influences on glucose levels. Also can lead to false positives in patients with delayed gastric emptying [46,57,58].
- LACTOSE TOLERANCE TESTS WITH ETHANOL (LTTE). Determination of urinary galactose after the intake of lactose with ethanol (ethanol inhibits galactose metabolism) [59–61].

i.3. A new approach: *Lactest®*

A non-invasive evaluation procedure was developed based on the administration of an analogue of lactose *(*Figure i. 3). The presented method allows to determine lactose intolereance with high sensitivity (94%) and specificity (92%), and using simple colourimetric technic (available in most of the hospitals).

The first version of the test was attempted using 3-methyllactose, *center*-Box i. 3 [62–64], that is hydrolysed by intestinal lactase to release 3-methylglucose. This compound is absorbed from the intestine without demethylation, transported across the intestinal mucosa [65], and detected in urine by gas chromatography analysis (Box i. 3). The main drawback for the application of this method is the need of specialised analytical facilities (gas or liquid chromatography) and the toxicity of the compound's reported.

Later on, an improved method based on the use of 4-galactosylxylose was developed [64,66,67], *right*-Figure i. 3.

Box i. 3. *Analysis of 3-o-methylglucose*

Analysis of 3-*O*-methylglucose include urine collection before and after administration. Then, samples are lyophilised and the residues trimethylsilated for chromatographic analysis.

Lactose **3-methyllactose** **4-*O*-β-D-galactopyranosyl-D-xylose (gaxilose)**

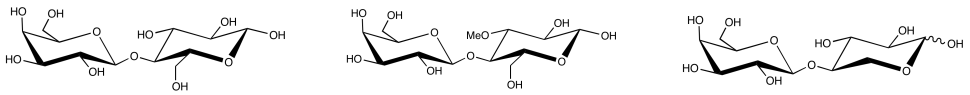


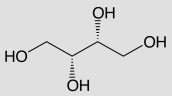
Figure i. 3. *Lactose structure and structure analogues used in the development of the diagnostic method*

This disaccharide, a structural analogue of lactose, is a substrate of intestinal lactase yielding galactose and xylose (Box i. 5)

While galactose is phosphorylated and metabolized, xylose is passively absorbed from the small intestine and, although a small portion may be metabolized (not phosphorylated) (Box i. 4), it is mainly eliminated through the urine where it can be measured using a simple colorimetric procedure [68–70]. Moreover it also can be measured in blood. The amount of xylose is a measure of total lactase activity *in vivo* (Figure i. 4).

Box i. 4. *Xylose: transport and metabolism*

The transport of xylose across the mucosal border is completely diffusional [401,402]; therefore it does not involve phosphorylation step [403]. It is described that some amount of xylose is catalysed in the kidney and liver and D-threitol is the main end product of xylose metabolism [404].



Box i. 5. *Analysis of galactoxylose regioisomers*

4-, 3-, and 2-galactosylxylose were analysed to find the best correlation with lactase activity *in vivo*. The 4-*O*-β-D-galactopyranosyl-D-xylose was the isomer with the best diagnostic and clinical correlation [400].

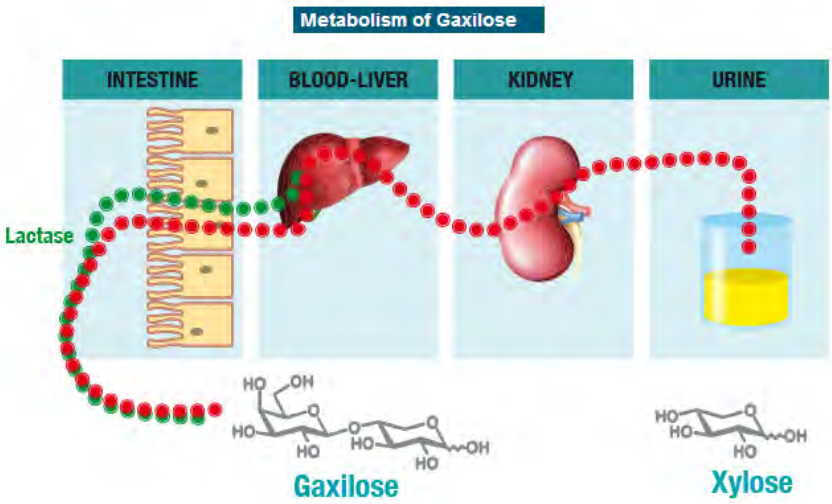


Figure i. 4. *Metabolism of Gaxilose (image from http://www.ferrerincode.com/en/dealers/products/lactest)*

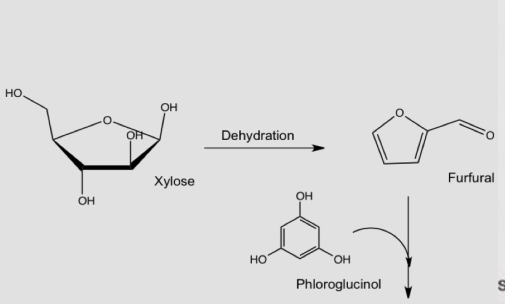
Several methods have been described in the literature for the determination of D-xylose in urine, plasma and serum: including phloroglucinol colorimetric method [70–72],

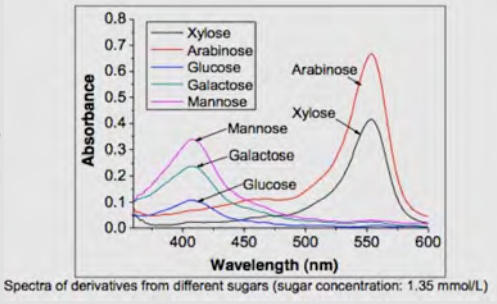
o-toloudine reactions [73,74], gas chromatography (GC) with previous chemical modification [69,75], high performance liquid chromatography (HPLC), high performance anion exchange chromatography (HPAEC) or even with enzymatic reactions [76]. Although HPLC and HPAEC are regarded as the best methods for sugar analysis, qualitatively and quantitatively, they require high-cost analytical columns, effluent reagents, and instrumentation maintenance.

The most widely used is the phloroglucinol colorimetric method, due to its simplicity and reliability (Box i. 6). In addition, the equipment required is available in almost all clinical laboratories.

Box i. 6 *A method for determination of xylose*

Xylose determination is based on the following reaction: in a solution of hydrochloric acid, phloroglucinol gives color reaction with sugars or their degradation products, showing maximum absorbance at 553 nm for xylose [405]. Among different colorimetric approaches, this method has the advantage of not only having higher meaurement accuracy, but also low toxicity of its color reagent, phologlucinol.





Spectra of derivatives from different sugars (sugar concentration: 1.35 mmol/L)

Non-xylose interfering constituents in the determination of xylose have been described [70,75]. In order to avoid intereferences, proteins are precipitated without affecting xylose reproducibility and recovery. However, glucose still constitutes the major source of background signal. Therefore, basal blank samples need to be included in the analysis.

Nowadays, gaxilose is the API of the diagnosis test named *LacTEST*® and it is used mainly in hospitals [77,78] (administration of 0.45 g of gaxilose and urine collection during 5 hours).

i.4. Synthesis of Gaxilose

i.4.1 Chemical synthesis

Chemical synthesis of 4-*O*-β-D-galactopyranosyl-D-xylose comprises synthesis of benzyl β-D-xylopyranoside and that follows a sequence of operations that implies selective protection, glycosylation and deprotection reactions. The high number of steps of the reaction and the use of expensive reagents (e.g. silver triflate in the

glycosylation reaction) produce costs and have difficulties to carry out this process on an industrial scale [67] (patent number ES-P-90016580). Therefore, alternative synthetic routes have been sought.

i.4.2 Enzymatic synthesis

In this context, enzymes offer an attractive solution with mild conditions, no need for group protection and in a regio- and stereo-selective approach.

There are two main classes of carbohydrate active enzymes able to synthesise glycosdic bonds: **glycosyltransferases (GTs)** and **glycosylhydrolases** (GHs or glycosidases). Enzymatic gaxilose synthesis approach is based on the used of β-galactosidase from GHs family.

i.4.2.1 Glycosyltransferases

Glycosyltransferases (GTs; EC 2.4.x.y) are the enzymes responsible for creating the diverse and complex collection of oligosaccharides and glycoconjugates found in nature. GTs catalyse glycosidic bond formation using activated sugar donors and acceptors. Donor sugars are most commonly activated in the form of nucleoside diphosphate sugars (e.g. UDP-Gal, GDP-Man). However, nucleoside monophosphate sugars (e.g. CMP NeuAc), lipid phosphates (e.g. dolichol phosphate oligosaccharides) and unsubstituted phosphate can also be donor moieties [79]. The acceptors can be a range of biomolecules including sugars, proteins, lipids or small molecules. Furthermore, although glycosyl transfer mostly occurs at the nucleophile oxygen of a hydroxyl substituent of the acceptor, it can occur at nitrogen (N-linked glycoproteins), sulphur (thioglycosides in plants) and carbon (C-glycoside antibiotics) nucleophiles too.

GTs have been classified based on sequence similarities into 105 families (as of January, 2018) in the Carbohydrate Active enZymes data base (CAZy) [80]. Structural information is only available for over 100 GTs and mainly two different three dimensional folds are observed: GT-A and GT-B. Both folds are related to the nucleotide-binding domain of Rossmann-like fold type. GT-A topology consists of α/β/α sandwich that resembles a Rossman fold. The central β–sheet is flanked by a smaller one, and the association of both creates the active site. On the other side, the GT-B fold is based on two seprate Rossmann domains facing each other through a flexible link and a catalytic site located between the domains [81,82].

When glycosyltransferases use sugar nucleotides as glycoside donor, enzymes are defined as Leloir glycosyltransferases, whereas if they use non-nucleotide donors (sugar phosphates, polyprenol phosphate) then are defined as non-Leloir glycosyltransferases (Figure i. 5).

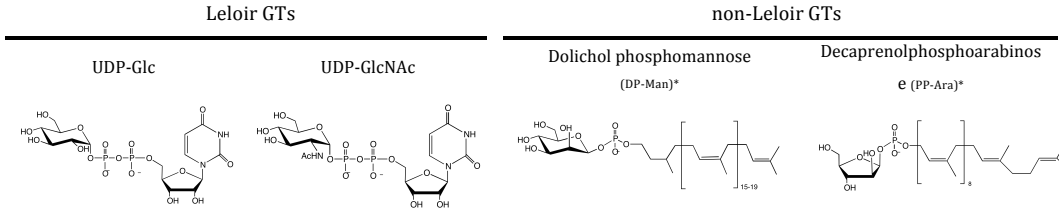
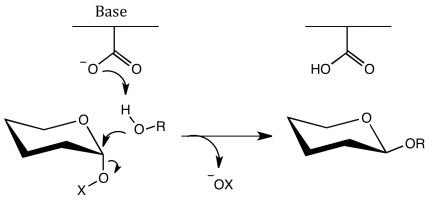
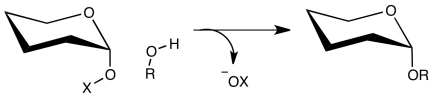


Figure i. 5. *Examples of common sugar donors used by glycosyltransferase: classified in Leloir and non-Leloir GTs. *DP-Man play an important role in the transfer of the first mannosyl residue to the hydroxyaminoacids* [83]*. PP-Ara is a donor used by arabinosyltransferases to assemble polysaccharides for mycobacteria cell wall* [84]

There are two stereochemical outcomes for the glycosyltransferase reactions, (Figure i. 6): *inverting* or *retaining* [85].



Inverting GTs proceed via a single displacement S_N2 mechanism with coordinate nucleophilic attack by the acceptor at the anomeric carbon, facilitated by proton transfer to the catalyic base, and leaving group release with an inversion of the anomeric configuration [86].



Retaining GTs if the anomeric configuration is retained. Its mechanism is still a matter of debate. The two proposed mechanisms are: i) a double displacement mechanism via a covalent intermediate (for GTs with a catalytic nucleophile) analogous to the mechanism of glycoside hydrolases (Figure i. 9); and ii) a SNi-like front attack mechanism in a single step.

Figure i. 6. *Inverting and retaining mechanisms of GTs*

Glycosyltransferases are attractive tools for the synthesis of oligosaccharides because they show a high degree of specificity and a large number of these enzymes can catalyse the formation of unique linkages.

They can be produced by recombinant technology and some commercial ones are available at high prices. Moreover, different technologies have been developed for large-scale synthesis (production of hyaluronic acid using hyaluran synthase enzyme and UDP-ClcNAc and UDP-Glc as substrates linked with sugar regeneration [87]).

However, GTs are frequently membrane-bound proteins expressed in low concentrations and quite difficult to express and purify. Additionally, they are unstable once isolated.

i.4.2.2 Glycosidases

Glycosyl hydrolases or glycosidases (GHs) are enzymes that catalyse the stereospecific hydrolysis of glycosidic bonds in oligo- and polysaccahrides or glycoconjugates. Moreover, GHs are also able to catalyse bond formation through transglycosylation [88–91].

Glycosyl hydrolases are classified in 149 families (as of January, 2018) based on amino acid sequence similarities. In this classification, enzymes with different substrate specificities are found in the same family (divergent evolution) and enzymes that act on the same substrate are sometimes found in different families (convergent evolution) [88]. GHs-genes are abundant and present in the vast majority of genomes corresponding to almost half of the enzymes classified in CAZy (47%) [92].

Glycosidases can be classified as *exo*- or *endo*-enzymes depending on the position of hydrolysed glycosidic bond in the substrate chain. Whereas *exo*-glycosidases release mono- or disaccharides from one of the chain ends, usually from the non-reducing end, *endo*-glycosidases act on internal glycosidic bonds within the oligosaccharide chain.

This specificity for cleavage is a consequence of the active site topology. Although many protein folds are represented in the GHs families, the overall of the active sites topologies fall into three general classes, linked to endo-/exo- classification.

1) **Pocket** topology encountered in *exo*-glycosidases and optimal for recognition of a saccharide on the chain end (*1*, Figure i. 7*)*.
2) **Cleft** is an "open" structure that allows binding of polymeric substrates and is commonly found in *endo*-enzymes such as lysozymes, chitinases or xylanases. (*2*, Figure i. 7*)*.
3) **Tunnel** topology derives from cleft structure, but with long loops that cover part of the cleft (3, Figure i. 7*)*. This topology allows that the enzymes release the products while polysaccharides remain firmly bound, thereby originating processive conditions [88,93,94].

Figure i. 7. *The three different active sites topologies in glycosidases. 1) The pocket: β-galactosidase from* E. coli (PDB: *1JZ7) 2) The cleft: 1,3-1,4-β-D-glucanase from* Bacillus licheniformis [95] *(*PDB*: 1GBG) 3) The tunnel: cellobiohydrolase from* Trichoderma reesei *(PDB: 1CB2,*[88]*). VMD representation: QuickSurface*

Glycosidases work by general acid-base catalysis involving glutamic acid (Glu), aspartic acid (Asp) or tyrosine (Tyr) as catalytic residues, but differ in their mechanisms as a consequence of their active site topology. Depending on the catalytic mechanism, GHs can also act with retention or inversion of configuration.

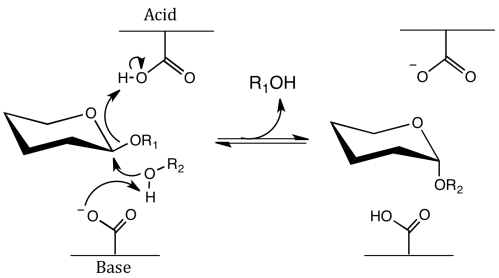


Figure i. 8. *Inverting glycosidase mechanism*

Inverting glycosidases work by a single step mechanism: protonation by the general acid of the glycosidic oxygen and aglycon release are coupled with the attack of a water molecule (or another sugar molecule), which is activated by the base residue (Figure i. 8). Catalytic residues are located ~10 Å apart from each other allowing the binding of the substrate(s).

Retaining glycosidases operate by a double-displacement reaction thanks to two amino acid residues: a general acid/base residue which acts as a proton donor in the first step (and as a base in the second) and a catalytic nucleophile. Sometimes, a third catalytic residue is involved in the modulation of the pK_a of the acid/base or nucleophile residues [96]. In the first step (glycosylation), the general acid protonates the glycosidic oxygen while the nucleophile, a deprotonated carboxylate, attacks the anomeric carbon with simultaneous C-O breaking and leading to a covalent glycosyl-enzyme intermediate. The second step, deglycosylation, involves the attack by a molecule of water (or another acceptor such as a sugar) assisted by the conjugate base of the general acid residue which renders the free sugar with retention of

configuration, and the enzyme returns to its initial protonation state (Figure i. 9) [88,97,98]. In this mechanism catalytic residues are closer to each other (~5.5 Å).

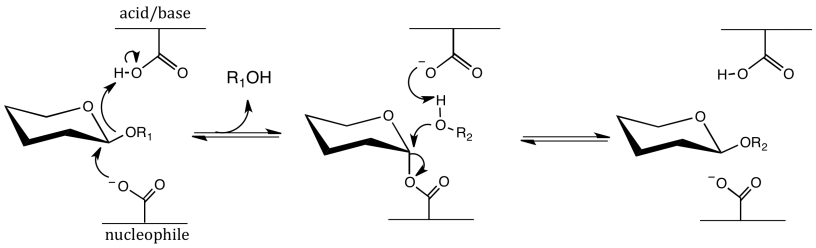


Figure i. 9. *Retaining glycosidase mechanism*

Although the two mechanisms differ significantly, they show some similarities: both involve oxocarbenium ion-like transition state and both involve a pair of carboxylic acids with different functions [99]: acid/base and nucleophile roles.

Among the retaining glycosidases group, some GHs catalyse bond hydrolysis/formation on 2-acetyl-hexosaminyl units of a glycan (GH18 chitinases [100,101], GH20 hexosaminidases, GH85 endo-β-N-acetylglucosaminidase). These enzymes have only a single carboxylic group acting as general acid/base, lacking the catalytic nucleophile and, therefore, operating by **substrate-assisted** catalysis, where the *N*-acetyl group of the substrate acts as internal nucleophile. In the second step, the oxazolinium intermediate is attacked by the hydroxyl group of a water molecule or of a sugar molecule, assisted by the conjugate base of the general acid residue. The product is then released with net retention of the anomeric configuration (Figure i. 10). There is an assistant residue in the mechanism that forms a hydrogen bond with de NH- of the acetamido group and helps the formation of the oxazolinium intermediate [102].

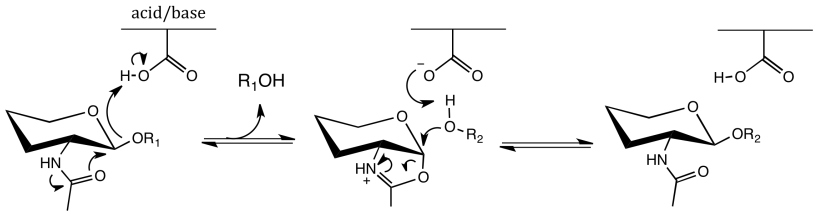


Figure i. 10. *Substrate assisted mechanism*

Retaining glycosidases have been employed in enzymatic synthesis of oligo- and polysaccharides and glycoconjugates. This can be achieved under conditions that favour the reversal of their hydrolytic reaction. For this purpose, reaction equilibrium could be displaced towards glycoside bond formation by thermodynamic control of the reaction or by the kinetic control using activated glycosyl donors.

Thermodynamic controlled synthesis involves the shift of the equilibrium towards products by altering the reaction conditions. This can be performed with different strategies: a) minimizing water activity in the reaction with the addition of water-miscible organic co-solvents, often limited by low organic contents due to enzyme instability or sugar solubility [103–106]; b) using high substrate concentrations [107], including the use of the acceptor as the reaction solvent (alcohol as acceptor [104,108]); c) increasing temperature [109].

Kinetic control is based in a faster trapping of the glycosyl-enzyme intermediate by a glycosyl acceptor than water. This strategy relies on the fast formation of the intermediate from an activated donor, with a good leaving group (such an aryl glycoside), which reacts with the acceptor faster than with water. Through this methodology a new glycosidic bond is formed with the same anomeric configuration in the product than in the donor (retaining reaction). The most commonly used strategies are: a) using high acceptor concentration to occupy the acceptor subsites; b) employing a highly reactive glycosyl donor.

Although glycoside bond formation is kinetically favoured, hydrolysis is still thermodynamically preferred because the product is a substrate for the enzyme itself, and thus the equilibrium is shifted towards hydrolysis.

Hydrolysis competes with transglycosylation by means of two processes: primary hydrolysis (donor hydrolysis) and secondary hydrolysis (hydrolysis of the product formed by transglycosylation). In order to avoid secondary hydrolysis several strategies have been developed [110]; such as removal of synthesis products from the reaction (crystallization or insolubilization of the product [111]), absorption of the product [112], transformation of the product into another compound that is non-hydrolysable (by coupling the reaction to another enzymatic process) or immobilization of the enzyme or the use of lipid-coated glycosidases which are stable and active in organic solvents and insoluble in buffer solutions.

GHs are potentially very attractive for chemo-enzymatic synthesis, because they are abundant, easy to produce and very stable, in contrast to GTs. Furthermore, requirement of glycosyl donors is relatively cheap and easy to produce at large industrial scale.

i.4.2.2.1 Engineered GHs

i.4.2.2.1.1 Glycosynthase

In 1998, Whiters and co-workers [113] engineered the first exo-glycosynthase, mutant of an exo-glycosidase, while Planas and colleagues [114] introduced the glycosynthase

concept for endo-glycosidases. Glycosynthases are engineered retaining glycoside hydrolases in which the catalytic nucleophile has been replaced by a non-nucleophilic residue. They are inactive hydrolases but able to catalyse glycosyl transfer to an acceptor when using activated glycosyl fluoride donors with opposite anomeric configuration than the original wild type enzyme. Donor mimics the covalent glycosyl-enzyme intermediate and, in the presence of an acceptor, deglycosylation occurs.

In a one-step inverting mechanism, the catalytic residue acts as a base to deprotonate the acceptor, activating the acceptor nucleophile that will attack the glycosyl donor in the active site and will create a new glycosidic bond (Figure i. 11).

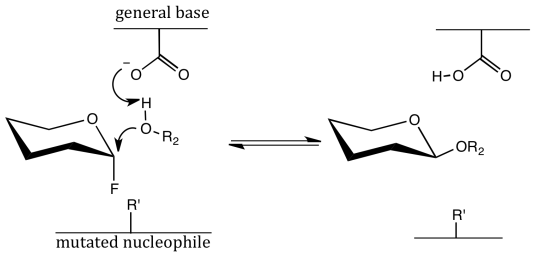


Figure i. 11. *Glycosynthase mechanism: nucleophile mutant (R¹) of a retaining glycosidase with glycosyl fluoride donor*

Mutations of the catalytic nucleophile disable the enzyme as a hydrolase because no glycosyl-enzyme intermediate can be formed; therefore, once the transglycosylation product is synthesised, it cannot be hydrolysed by the enzyme and thus higher product yields may be achieved.

Most glycosynthases are derived from retaining glycosidases. Glycosynthases originated by protein engineering from exo-glycosidases have moderate substrate specificity and regioselectivity and can synthesise up to tetra-oligosaccharides. On the other hand, endo-glycosynthases generally present higher regioselectivity and more specificity. Additionally, they can synthesise longer oligosaccharides because they have long glycan-binding sites to accommodate longer glycosyl donor substrates [115–119]. More recently, endo-glycosynthases are being developed for the enzymatic synthesis of artificial polysaccharides and functionalised polysaccharides [89,91,120,121].

Simple and general screening methods to select glycosynthase mutants have been developed: most of them using glycosyl fluorides as donors. Some methods rely on the pH drop caused by the release of hydrofluoric acid when glycosidic linkage is formed [122]. Other methods are based on the use of a specific chemical sensors (silyl ether of a fluorogenic methylumbelliferone) to transduce fluoride concentration into a fluorescence signal [123].

i.4.2.2.1.2 Engineered transglycosylases

Natural transglycosylases have only been identified in some GH families (xyloglucan endo-transglycosylases of GH16, sucrose enzymes of GH13 and GH70, sialidases of GH33)[124]. For others GHs, modification of the transglycosidase/hydrolyse (T/H) ratio has been carried out using different strategies of protein engineering: a random approach [125,126] or rational design such as Iterative Saturation Mutagenesis (ISM) [127] or combinatorial active site test (CAST) [128].

To increase T/H ratio by protein engineering, different authors agree to focus on a reduced number of sites within the enzymes [124,129]. Mutations of **subsite -1** to modify interactions with the donor (first or second shell of interactions) to, consequently, develop new transition state structure that lower the efficiency of water-mediated deglycosylation [130–133]. **Subsite +1** modifications are focused in the second step of the reaction, the deglycosylation, usually increasing acceptor interaction and/or decreasing water entrance.

Water has a key role in T/H ratio. Therefore, water interactions have been analysed by computational tools to diminish water entrance in the active site and further reduce hydrolysis [134–136].

i.4.2.3 β-galactosidases

β–Galactosidases (EC 3.2.1.23) are widespread in many organism including plants, animals and microorganism such as fungi, bacteria and yeast.

These enzymes hydrolyse galactosyl residues from polysaccharides, oligosaccharides or secondary metabolites. β–Galactosidases are mainly found in four GH families described in CAZy: GH1, GH2, GH35 and GH42 [*], all of them belonging to clan GH-A (clans are groups of families sharing a fold and catalytic machinery).

It has been previously described that GHs present double activity: hydrolysis and transglycosylation. Therefore, galactosidases can catalyse both reactions and thanks to this double activity, these enzymes are used for different applications [137,138]:

a) REMOVAL OF LACTOSE FROM DAIRY PRODUCTS (hydrolytic activity).

The removal of lactose from dairy products contributes to the digestibility, taste and organoleptic properties and improved processing of dairy products [139]. Furthermore, the fact that more than 70% of the world's population suffers from the inability of digesting lactose increases the interest of lactose hydrolysis.

[*] Families GH59 and GH147 also include one β-galactosidase (one in each family), however these β-galactosidases had not been expressed neither characterised, only identified [394].

Hydrolysis of lactose increases the sweetness of milk and dairy products (Glc and Gal are four-fold sweeter than lactose); in addition, lactose is a hygroscopic sugar and has a strong tendency to absorb flavours and odours and causes many defects in refrigerated products such as crystallization in dairy foods, development of sandy and gritty texture and deposit formation [140,141].

The main drawbacks of these enzymes are their slow rate of hydrolysis and their high cost [142]. To overcome this problem, enzymes can be immobilized. The enzyme immobilization has proven to increase enzyme thermostability and enzymatic activity [143]. The immobilization protocol needs to be simple and preferably use inexpensive inert insoluble materials. There are several methods to immobilize enzymes such as entrapment (e.g. with calcium alginate [144,145]), covalent binding (e.g. on glutaraldehyde-activated chitosan macropaticles[146],) or physical adsorption (e.g. with sephadex or chitosan beads [147]). Moreover, for food or pharma applications, nontoxicity and biocompatibility are also required [148].

b) PRODUCTION OF GALACTOSYLATED PRODUTCS (transgalactosylation activity).

Galacto-oligosaccharides (GOS) are used as non-digestible, carbohydrate-based food ingredients in human and animal nutrition. They have been labelled as prebiotics. They are not digested by humans or other animals and selectively increase the beneficial microflora of the intestine, leading to health benefits that are extensively recognised [149]. GOS are defined as a mixture of those substances produced from lactose, which comprise between two and eight saccharide units, with one of these units being a terminal glucose and the remaining saccharides units being galactose [150].

Currently, GOS are produced using lactose as substrate and a GH as catalyst to produce high-GOS-content mixtures. GHs need to have good ability to better perform transgalactosylation relative to hydrolysis, and additionally have low affinity for the formed GOS [138,151,152]. Using high initial lactose concentration (30-40%, weight/volume) yields can increase. High temperatures significantly increase GOS yield due to higher lactose solubility. This latter approach can only be attained when using thermostable enzymes (e.g. *S. solfataricus, P. furiosus* [153–155]*, T. maritima* [156,157]).

c) ANALYTICAL APPLICATIONS

Biosensors are important tools in a variety of fields including immunoassays, toxicology analysis, forensics, drug screening, gene expression analysis, gene identification, agro diagnostics and pharmacogenetics. It combines the selectivity of molecular recognition of the biomolecules and the sensitivity of the signal traducers. A wide variety of biosensors have been developed using different bio-recognition elements such as enzymes, antibodies, peptides, whole cells or nucleic acids. Quality control of milk and its derivates is a very demanding field. β–Galactosidase can be covalently immobilized onto gold-coated magnetoelastic film via a self-assembled monoloyer of ω –carboxylic acid alkylthiol. The application of magneoelastic transduction allowed for wireless *monitoring of enzymatic activity* through the associated changes in the frequency and amplitude of magnetic fields [158].

The enzyme has many analytical uses, being a favourite label in various affinity recognition techniques such as enzyme linked immunosorbent assays (ELISA). Luckacheva et al. [159] developed a procedure fot the determination of glucose and lactose in food products with the use of biosensors based on Berlin Blue. Signal transducer linked to hydrolytic activity (Berlin blue [159]) is used to determine the remaining amount of lactose in food. Briefly, lactose-fermenting microorganisms, are mixed with potential remaining lactose. If there is undigested lactose in the mixture, then glucose and galactose are generated and their metabolisme releases CO_2 that is easily measutable [160].

i.4.2.3.1 β-galactosidase from *E. coli*

i.4.2.3.1.1 The enzyme

β-Galactosidase, the product of the *lacZ* gene of the *lac* operon of *E. coli*, has a rich and large history in molecular biology and biochemistry. It played a central role in Jacob and Monod's development of the operon model for the regulation of gene expression [161]. It

Box i. 7. *Components of lac operon*

The *lacZ* gene codes for the cytoplasmaic enzyme β-galactosidase. The *lacY* gene encodes the Lac permease, a cytoplasmic membrane protein that transports lactose into the cytoplasm. The *lacA* gene encodes a cytoplasmic galactoside transacetylase, Figure i. 13.

was the first genetic clearly understood regulatory mechanism, where the key idea is that protein expression could be auto regulated.

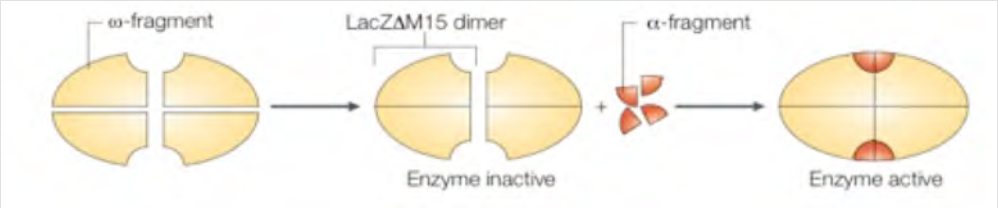
The *lac* operon contains three genes that are transcribed from a single promoter in the following order: *lacZ, lacY* and *lacA* (Box i. 7).

The *lacZ* gene is frequently used as a reporter gene in bacterial-based selection techniques due to the simplicity and sensitivity of the screening method. This is one of the most used screening method to determine whether a transformed bacterial colony bears an empty or a recombinant plasmid (Box i. 8).

The most sensitive indicator of LacZ activity is the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). This colourless molecule is hydrolysed by LacZ to produce a compound that dimerizes to form the insoluble dye indigo (bright blue colour).

Box i. 8. *α-complementation phenomenon*

The α-complementation is based on the fact that the β-galactosidase is a tetramer. Each of their monomers is made of two parts: *lacZ*-α *(*1-41 aa*)* and *lacZ*- ω (41-1023 aa). It has been determined that if the α fragment is deleted, the ω fragment is non-functional. However, the protein can return to its active tetrameric state in the presence of the $N_{terminal}$ fragment, α-peptide. This rescue is called α complementation [406–408].



*** Image from** [164]

The blue-white technique allows detecting if DNA of interest, the insert, is ligated into a vector or not. The vector carries within the lac*Z*α sequence an internal multiple cloning site. DNA ligated into the plasmid disrupts the α-peptide and therefore a non-functional galactosidase is formed (insertional inactivation). Once cells are transformed, the presence of an active or inactive galactosidase can be detected using X-gal in LB-plates. Cells transformed with vectors containing recombinant DNA will produce white colonies (insert breaks the gene); while cells transformed with non-recombinant plasmids (vector without insert) grow into blue colonies. This method requires a suitable bacterial strain with genotype LacZΔ15.

The *lacZ* gene has also been exploited as fusion reporter gene as it was shown that its $N_{terminus}$ can be replaced by other terminus of various sequences and lengths [162,163]. There are two types of *LacZ* fusions [164]: *transcriptional* mode, where the *lac* operon is placed downstream from a different promoter (*left*- Figure i. 12); or *translational*

mode where proteins are fused and result in the production of a hybrid protein, in most cases such fusions express an enzymatically active molecule (Figure i. 12). Both modes are used in different studies of regulation (transcription, translation) and stability.

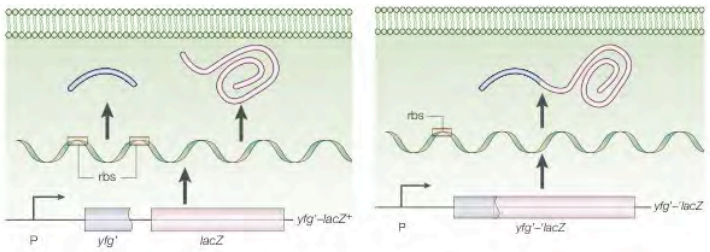
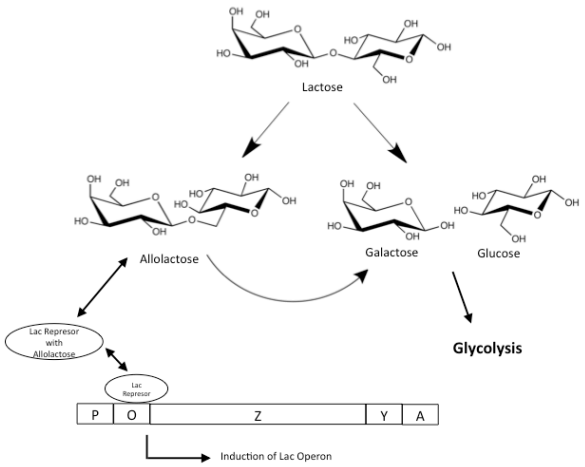


Figure i. 12. *Two types of LacZ fusion.* *yfg: your favourite gene Image from [164]

The β-galactosidase from *E. coli* has three enzymatic activities. Firstly, the enzyme can hydrolyse lactose to form galactose and glucose, the latter being then able to enter glycolysis. Secondly, it is also able to convert lactose, through transgalactosylation, to another disaccharide: allolactose, which is the natural inducer of the *lac* operon. Thirdly, the allolactose can be also hydrolysed to monosaccharides (Figure i. 13). These three activities allow the cell to self-control the transcription and subsequent protein expression of the *lac* operon (Box i. 9)).



Box i. 9. *Lac operon system*

Lactose is transform to allolactose, which binds to the *lac* repressor and reduces its affinity for the *lac* operon. The *lacZ* gene is activated and β–galactosidase is expressed.
Therefore, the *lac* operon allows the effective digestion of lactose when glucose is not available.

Figure i. 13. *β-galactosidase activity on lactose: hydrolysis and transglycosidase. Lac operon regulation by allolactose* [165]

β-Galactosidase from *E. coli* belongs to family GH2 of CAZy database. The GH2 family contains more than 9,000 entries, with nearly 150 characterized, most of them β-galactosidases [166].

E. coli β-galactosidase is a tetramer of four identical chains. Each monomer has 1,023 amino acids, resulting in a protein with a molecular weight of 116,253 Da (Box i. 10). Each monomer has one catalytic centre (4 catalytic centres in each tetramer) that

functions independently, and that shows no cooperativity or allosteric effects. There is a continuous system of channels running along the surface within the tetramer. These channels (5-20 Å) appear to be accessible to bulk solvent; the four active sites are 1 at the bottom of such surface channels. It has been described that β-galactosidases exposes more surface area to solvent than expected in comparison with other oligomeric proteins [167,168]. Furthermore, about 70% of the solvent atoms in the structure are in equivalent locations in the four monomers.

Box i. 10. *Extinction coefficient of β-galactosidase from E. coli*

The protein contains relatively high levels of tryptophan and arginine and low levels of lysine and tyrosine compared to most proteins. It is due the content of 39 residues of tryptophan explains the high molar extinction coefficient for *E. coli* β-galactosidase [409,410].

There are two main subunit interfaces: the "long" and the "activating". Domains 3, 4 and 5 form the *long* interface, whereas *activating* interface involves domain 1, 2 and 3. While the long interface is fairly planar, the activating is *S*-shaped. This interface is formed by two equivalent interactions between domain 3 and a loop that include residues 271-288 (Figure i. 14). The loop, from domain 2, extends across the interface and completes the active site of the neighbouring subunit with polar interactions (more polar than interactions of domain 3 interface part, 65% *vs* 45% [167]).

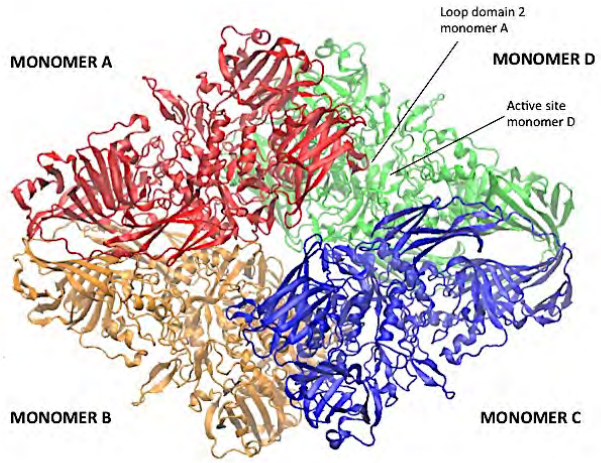


Figure i. 14. *Tetrameric structure of β-galactosidase from* E. coli *(*PDB*:1jz7): each monomer is differently coloured. Visualization with VMD software: representation in New Cartoon*

The tetramer is usually described as a dimer of dimers (A-D and B-C) (Figure i. 15). The formation of each active site requires the *activating* interface to be present. Dissociation of the galactoside tetramer into dimers removes the loop (272-288, from domain 2) from the active site. Therefore, dissociation of the tetramer to dimers means deactivation [167,169].

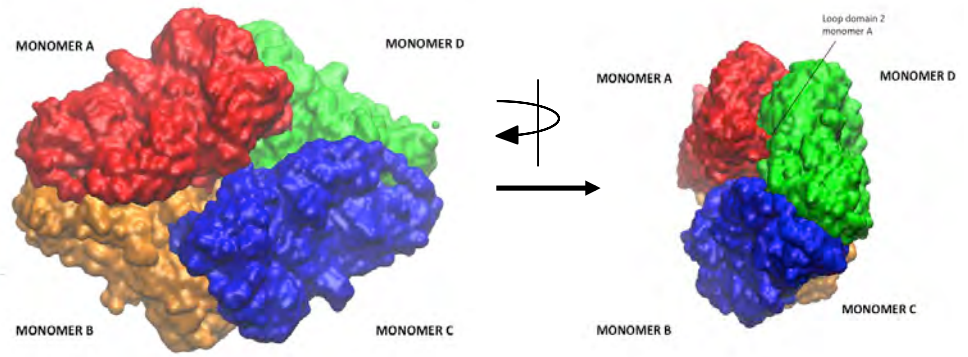


Figure i. 15. *Tetrameric structure of β-galactosidase from* E. coli *(*PDB*:1jz7) from two points of view. Visualization with VMD software: representation in Quick surface, each monomer is differently coloured*

Within each monomer, the 1023 amino-acid polypeptide chain folds into five structural domains [167,170,171] (Figure i. 16).

Figure i. 16. *Domain organization of β-galactosidase from* E. coli *by* Pfam

The **third/central domain (D3)** is the catalytic one. It is an eight-stranded α/β barrel that is called triose phosphate isomerase (TIM) or $\alpha_8\beta_8$ barrel fold with the active site forming a deep pocket at the C-terminal end of the barrel. Moreover, the active site is also formed by amino acids from other parts in the same polypeptide chain as well as from other chains within the tetramer [165,167], (green in Figure i. 17). This domain is common in β-galactosidases.The other domains have different roles in distinct β-galactosidases, which suggest that they have evolved as modules to provide stable scaffolds upon which binding residues can be presented to render specificity to the enzymes.

It has been described that domains 1, 2, 4 and 5 are independent folding modules that serve to supplement or to modify the central role of catalytic domain [172]. **Domain 1 (D1)** fills one end of the cleft in the catalytic domain with two loops (blue in Figure i. 17). Residues of these two loops interact with the non-reducing end of the galactosyl moiety, giving the enzyme specificity and defining the shape of the binding site. Domain 1 structure is similar to a cellulose binding domain (CBD), domains that usually are separated from their catalytic domains by flexible linkers (domain 2). **Domain 2 (D2)** (red in Figure i. 17*)* possesses residues that interact with the active site of a neighbouring subunit via a loop that reaches across a subunit interface, helping to convert the cleft to a pocket (grey in *down*-Figure i. 17*)*. In addition, D2 appears to

function as a linker between domains one and three. **Domain 4 (D4)** (yellow in Figure i. 17*)* is uncommon in β-galatosidases. Probably its function is to act as a linker between domains three and five. **Domain 5 (D5)** (orange in Figure i. 17) fills the other part of the cleft, opposite to D1, and its residues also interact with the substrate. Its role is uncertain, but it seems that this domain shapes the active site and determines the specialization among the hydrolysis of different galactosides or the product specificity in transglycosylation reaction [166].

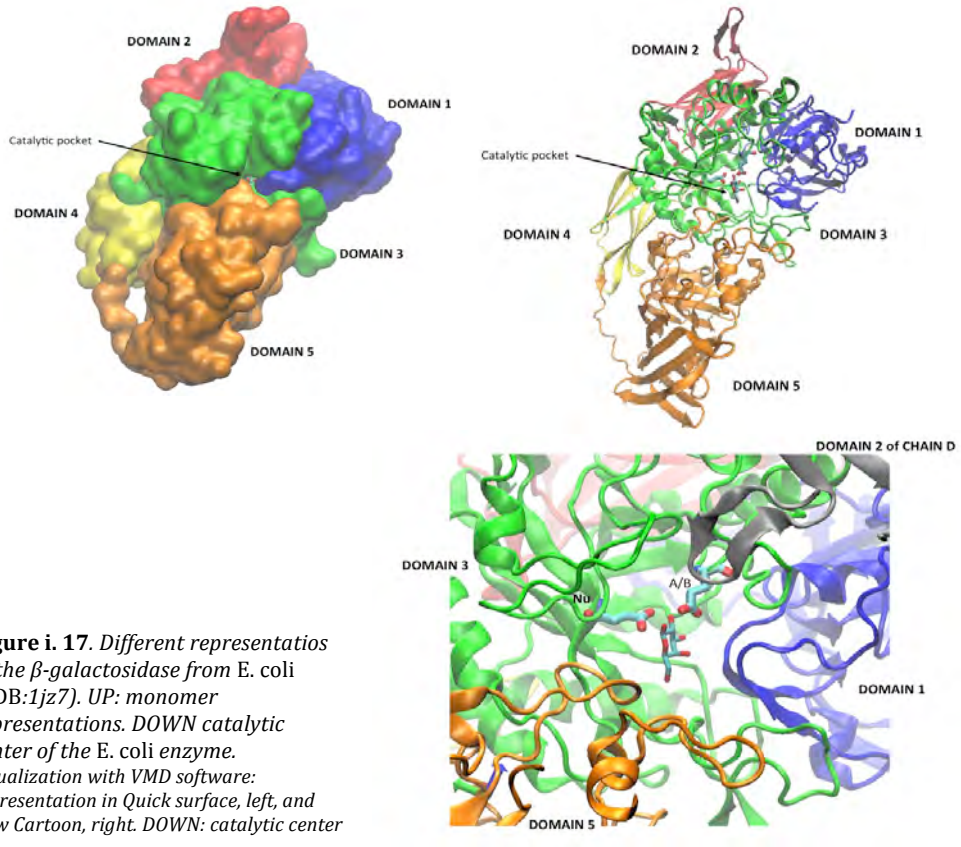


Figure i. 17. *Different representatios of the β-galactosidase from* E. coli *(*PDB*:1jz7). UP: monomer representations. DOWN catalytic center of the* E. coli *enzyme. Visualization with VMD software: representation in Quick surface, left, and New Cartoon, right. DOWN: catalytic center*

There are two substrate binding modes in β-galactosidase: a shallow and a deep mode. Substrate binds the free enzyme initially in a **shallow** mode and stacks with Trp999. It interacts with van der Waals forces and takes a position parallel to it with the galactosydic oxygen centred over the indole [173]. Moreover, there are specific bonds to each of the hydroxyls of the galactosyl component of the substrate in the shallow site. There are no specific bonds with the other part of the molecule (such as glucose in lactose, *o*-nitro group in *o*NPG). In this position, the substrate is bound pre-productively and is not in the catalytically correct position between Glu461 (nucleophile residue) and Glu537 (acid/base residue). The substrate has to move 3 Å deeper into the active site for catalysis to begin. Progression to the **deep** mode involves

changes in galactosyl interactions: in this mode galactose stacks another tryptophan, Trp568, while hydroxyls undergo substantial changes in environment.

Switch to the deep binding mode is accompanied by two conformational changes. The first one is related with a loop of domain 5, which swings toward the active site with residue shifts up to 9 Å. The loop is composed by ten residues, from Gly794 to Pro803 (orange-open and grey-closed in Figure i. 18).

It is open when the enzyme has no ligands bound and it closes when the substrate or some transition state analogues bind to the enzyme. It is partly closed when covalent intermediate is present. Gly794 has an important role: it is a "hinge" for opening and closing the loop. It has been studied that changes to another amino acids cause changes in loop conformation, such as Ala that favours loop closure [174].

Upon loop movement, Ser796 moves about 11 Å, more than any other residues in the loop. Serine appears to act like a two-way "clasp" that aids in keeping the loop either open or closed depending on reaction stage. Hydrophobic and H-bonding properties of Ser796 result in interactions that are strong enough to stabilize the open or the closed conformations of the loop, depending on the stage of the reaction, but weak enough to allow loop movement during the reaction [175] (Box i. 11).

Box i. 11. *Analysing deeply Ser796*

Ser796 substitutions were carried out to understand the role of this residue [175].

- The *alanine* substitution causes the loop to highly favour the closed position because of better hydrophobic interactions with Phe601. Additionally, interactions in the open form are lost due the hydroxyls are not present.
- *Threonine* causes the open form predominanate because probably its methyl group is within van der Waals distance of the side chain of Ile793, Ala805 nd Val807, ande gamma carbon of Glu808, creating a hydrophobic cluster.
- *Aspartic* results in decreases in stability of both the open and closed states.

The second conformation change is related to residue Phe601. The side chain of Phe601 swings approximately 60º while interacting with the hydroxymethyl moiety of galactose. It is most likely pushed by the C6 of the galactosyl part of the substrate whilst it moves into deep binding mode prior to the formation of the transition state (orange and grey in Figure i. 18) Three more residues are important for loop conformations, despite not conforming the core of the loop [176]: Arg599 which plays an important role in anchoring the open conformations of Phe601 and active-site loop [177]; Met542 that acts as a guide for the movement of the benzyl side chain of Phe601 [178]; and Glu 808 that interacts with Ser796 of the active loop in open conformation [175] (Figure i. 19)

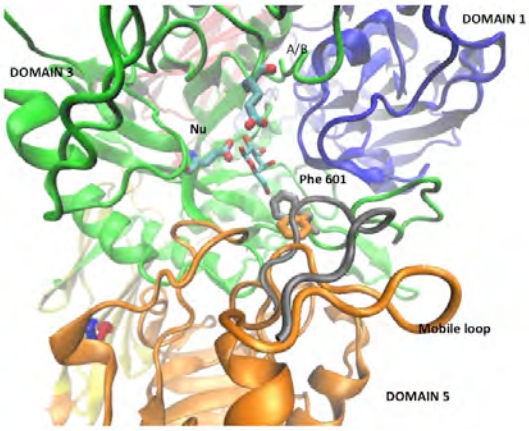


Figure i. 18. *Catalytic center of β-galactosidase from* E.coli. *Two loop (G704-S803) conformations are represented: closed loop in grey and opened loop in orange. Moreover Phe 601 shift is also represented near galactose moiety. Structures are visualized with VMD software and structures ar 1JZ7 for conformation of opened loop and 1JZ5 for closed loop; both downlanded from PDB database*

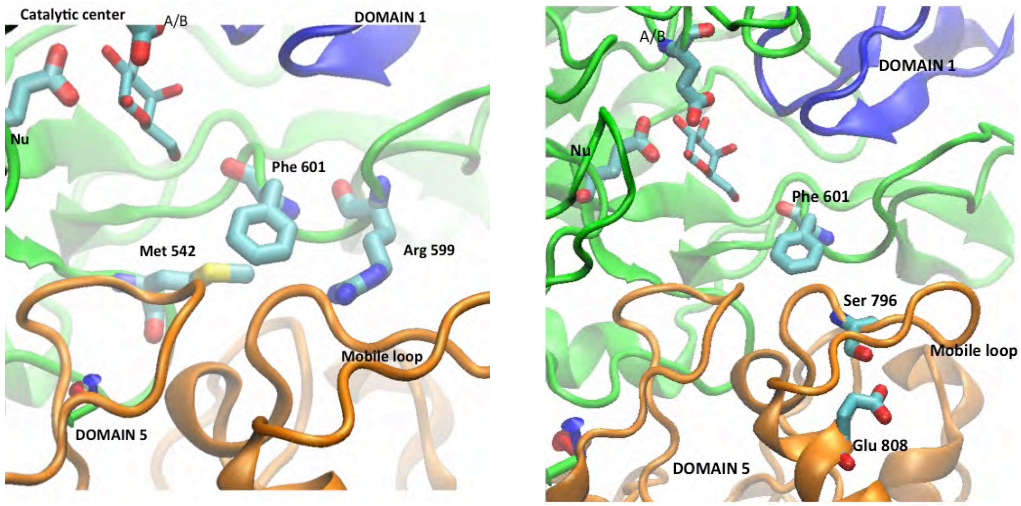


Figure i. 19. *Interactions among the mobile loop, Phe601 and other residues (the loop is open in both images)*

i.4.2.3.1.2 Gaxilose production

Gaxilose (4-*O*-β-D-galactopyranosyl-D-xylose) synthesis in a classical chemical reaches a maximum of 9% overall yield, which is makes it too inefficient for industrial production. In this scenario, enzymes represent a good alternative, allowing region- and stereo-selective synthesis under mild conditions.

In the early 1990's, a group from General Organic Chemistry Institute of CSIC (Madrid) in collaboration with a group from Biochemistry Department from Universidad Autónoma de Madrid (UAM) performed some studies attaining gaxilose synthesis by exploiting the transgalactosylation activity of β-galactosidase from *E. coli,* *o*NPG or lactose as donors and diverse substituted β-D-xylopyranosides. They reported different yields and regioselectivities depending on the substitution of the anomeric position of xylose [179–181].

In an attempt to increase yield production, various enzymes of diverse sources were selected by their transglycosylation properties. Therefore, six β-galactosidases were evaluated using an excess of xylose and *o*-nitrophenyl-β-D-galactopyranoside as donor:

bacterial galactosidased from *E. coli,* one fungal enzyme from *Aspergillus oryzae*, one from yeast *Saccharomyces fragilis* and three from mammalian β-galactosidases from bovine testes, from bovine liver and the intestinal lactase. Different yields and regioselectivities were achieved with those enzymes. In particular, β-galactosidase from *E. coli* and from bovine testes showed the highest yield and selectivity for gaxilose [182]. Enzymatic conditions, using β-galactosidase from *E. coli*, were optimized evaluating the effect of pH, temperature and the addition of cosolvents (DMF, DMSO, THF, diethylenglycodiethylether and acetonitrile) [183] (Table i. 1).

Cosolvent	T (°C)	pH	Yield (%)	Ratio β-1,4:(β-1,3+ β-1,2)
DMSO	37	7.0	35	70:30
DMF	37	7.0	38	72:28
-	37	7.0	50	71:29
-	37	5.0	22	81:19
-	37	8.5	36	68:32
-	45	7.0	48	68:32
-	25	7.0	45	79:21
-	5	7.0	48	80:22
-	-5	7.0	46	83:17

Table i. 1. *Yield and regioselctivity of disaccharides from reactions using oNPG and xylose as donor and acceptor and β-galactosidase from* E. coli [183]

It was observed that pH changes had a negative effect on yield, although an increase of selectivity was observed at pH 5.0. Temperature decreases regioselectivity. On the other hand, to work at lower temperatures required a larger amount of enzyme to complete the reactions (even 12-fold more). The best results in terms of enzyme efficiency and yield were at 37°C in pH 7.0 buffer.

In 2004, a patent was published to protect the enzymatic procedure of gaxylose synthesis (EP 1 408 118 B1). The authors reported a method that showed 50% yield and 71% regiospecificity. After several steps of purification, the final overall yield diminished to 22-25%.

Frame of the thesis

In the framework of the collaboration between the company INTERQUIM, S.A, and the Laboratory of Biochemistry at the IQS School of Engineering, the search of novel biocatalysts for the production of gaxilose was envisaged. The project is also embedded in the Industrial Doctorates Programme of the Generalitat de Catalunya, a co-funding scheme to pursue intersectoral PhD studies.

Interquim S.A. is integrated in the Ferrer HealthTech Division and manufactures active ingredients for the pharmaceutical industry. The company's core business is based on the development of highly competitive processes for the manufacture of high added-value active pharmaceutical ingredients that meet the strictest quality specifications, under European environmental and work safety regulations. The products and processes have been audited by international authorities, including the US FDA, EMA, and are sold in Japan, EU and US, indicating the excellence of the implemented standards. INTERQUIM is proprietary of the LacTest® and owns a patent for the enzymatic production of gaxilose. The company aims at improving gaxilose's current industrial production.

The Laboratory of Biochemistry at IQS is a research group with long expertise in the enzymology, protein engineering, and application of carbohydrate active enzymes, particularly glycosidases, glycosyltransferases, and carbohydrate esterases. The group developed the glycosynthase technology and currently is involved in the engineering of glycosidases into transglycosidases.

In this context, the main objective of the thesis is the contribution to the improvement of the efficiency of industrial gaxilose production focusing on the search and the engineering of β-galactosidases to obtain novel efficient biocatalysts.

RESULTS

Thesis organization

The results obtained are presented in three chapters.

CHAPTER 1 focuses on the study of the β-galactosidase from *E. coli* for gaxilose production (protein expression and protein engineering). The first chapter also includes the development of an analytical method by HPLC-MS for enzyme activity characterization (hydrolase and transglycosidase activities) and the description of analyzed activities in the whole work.

CHAPTER 2 focuses on the search of a new enzyme able to increase gaxilose yield. The search relies on the combined used of bibliographic analysis and bioinfomatic tools.

CHAPTER 3's objective is the study of the new enzyme. This chapter also includes the protein's 3D structure determination (crystallographic structure and model construction) and the set up of a HTS method for the selection of mutants obtained through protein engineering.

Some of the results obtained have been protected in a patent entittled *Enzymatic process for the preparation of 4*-O-*β*-*D*-*galactopyranosyl*-*D*-*xylose* (application number EP17382668.6) that was filled on October 2017. The patent is not included in the thesis due to confidential issues.

All the protocols used and the methods developed in this work are included in a final experimental section.

CHAPTER 1. β-galactosidase from *E. coli*

CHAPTER 1. β-GALACTOSIDASE FROM *E. COLI*

OPENING

Gaxilose is the active pharmaceutical ingredient (API) of LACTEST®. The last step in gaxilose synthesis is an enzymatic reaction catalysed by a β-galactosidase. β-galactosidases from diverse origins were studied as biocatalysts for that reaction, including *Aspergillus oryzae, Saccharomyces fragilis* or that of lamb small intestine amongst others. The selected enzyme was *E. coli* β-galactosidase [182], as explained in *i. INTRODUCTION*.

The *E. coli* β-galactosidase used in gaxilose industrial production was a commercially available enzyme, provided by Sigma (www.sigmaaldrich.com). The cost of the enzyme represents a high percentage of the total cost of goods for gaxilose production. That is the case for a large number of enzyme-catalysed APIs. One of the objectives of the present work was to increase the yield of gaxilose production and thus reduce production costs associated to the biocatalyst. Moreover, it was decided that the enzyme would be produced by Interquim, instead of purchasing it from a commercial source. The enzyme was cloned, expressed, purified and assayed at IQS premises at a mg scale. Then, the same enzyme was used at Interquim premises, where it was assayed at 10 g scale and the purification and isolation of gaxilose proved to comply with the regulatory standards. Once the enzyme proved to be an active biocatalyst, its industrial production was outsourced to a biotech company to produce enough amount of enzyme for industrial reaction at 10 kg scale. In a joint venture with that biotech company, the stability and reuse of the enzyme was investigated, using immobilization and different storage conditions to maintain and store the biocatalyst (liquid solution, lyophilized powder).

Enzymatic synthesis had been previously analysed and optimum conditions for enzyme activity had been determined [184], such as pH, temperature and use of co-solvents. Nevertheless, in this project several changes in the reaction conditions to optimize transglycosylation reaction were introduced: donor:acceptor ratio, substrates concentration or enzyme load. Those studies were performed at medium scale (10 g *o*NPG) and allowed the improvement of the reaction conditions to finally scale it up at industrial level (12 kg *o*NPG).

The β-galactosidase from *E. coli* is one of the most studied enzymes due to its history in the discovery of the *Lac* operon regulation. In addition, it has been extensively used as molecular biology tool. The mechanism, structure and enzymatic details of the enzyme have been described in previous studies (see *i. INTRODUCTION, i.4.2.3.1*). All that information was used to modify the enzyme by protein engineering in order to boost its

tranglycosylation and reduce its hydrolytic activity. The enzyme modification was focused on enzyme remodelling to increase gaxilose synthesis. In collaboration with Dr Alberto Marina and his group from Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas (IBV-CSIC), mutants were designed in order to have an impact in the movement of a loop (G794-P803),which is involved in substrate active binding mode.

This chapter also includes a description of all reactions carried out in this project in order to easy understand analysed activities. It is described how the double activity of enzymes (T and H) was compared.

In order to analyse enzymatic reactions, an analytical method was set up. Absorbance measurements methods allow detecting and quantifing the activity of enzymes but do not distinguish transglycosydation from hydrolysis. Direct quantification of synthesised disaccharides was needed to characterize the enzymes. A High Performance Cromatography Mass Spectrometry (HPLC-MS) method was developed to analyse reaction products at the Laboratory of Biochemistry, IQS School of Engineering.

1.1 Expression and purification of ß-galactosidase from E. coli

E. coli β-galactosidase from Sigma® is a commercially available enzyme able to catalyse gaxilose production. The enzyme showed high variability between batches and was too expensive, accounting for 50% of the total cost of goods of the gaxilose synthesis reaction. For both reasons, it was decided that the enzyme would be produced by Interquim.

Firstly, it was produced at laboratory scale (IQS School of Engineering facilities) and then it was outsourced to a biotech company (Biopolis, S.L.) to be produced in the large batches needed for industrial production (see *1.5.2.1*).

E. coli β-galactosidase gene was designed and subcloned in an expression vector (*pRSF-1b*) (Box 1. 1).

Box 1. 1 *pRSF plasmid*

The *pRSF-1b* plasmid includes coexpression capabilities as well as the ability to express fusion proteins with N-terminal His·Tag coding sequence that results in native protein after purification and cleavage (Figure 1. 1). The plasmid carries kanamycin resistance. It contains a strong T7*lac* promoter, an amino-terminal His·Tag coding sequence and a multiple cloning site (MCS) region. It also contains a cleavage site for the enterokinase (EK) protease and allows all amino-terminal vector-encoded sequences to be removed by EK digestion.

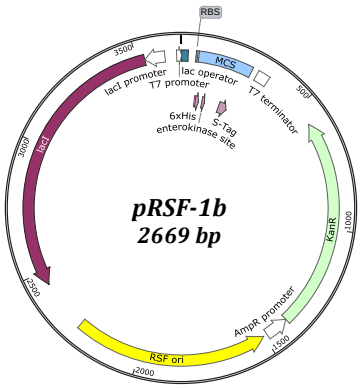


Figure 1. 1. *pRSF-1b vector map*

Protein expression system was carried out in *Escherichia coli* using the T7 system, which is included in the pRSF vector (Box 1. 2). Protein expression was performed in *E. coli* BL21 DE3 Star strain. Cells were grown in liquid LB in the presence of kanamycin (100 µg/mL) at 37ºC for 6-8 h and protein expression was induced adding IPTG (final concentration of 1 mM) at 37ºC for 16-18 h. This methodology allowed to obtain enough protein for laboratory scale experiments. However, protein expression was further studied in order to optimize the scale up during the industrial production trials.

Box 1. 2 *Expression system*

This expression vector with the gene cloned downstream of the T7 promoter was introduced into a T7 expression host (DE3 strains), which carries a chromosomal copy of the phage T7 RNA polymerase gene controlled by a *lac* promoter. When an inducer, e.g. IPTG, is added to the growth medium, T7 RNA Polymerase is expressed and thus the gene of interest is transcribed.

Protein expression was evaluated taking into account two main parameters: induction temperature and OD_{600} at the time of induction.

Induction temperature: after culture grew that 37ºC (OD_{600} 2-3), cultures were inducted (1 mM IPTG) at two different temperatures: 24ºC and 37ºC. Different samples were withdrawn at different times and after 16-18 h of induction, cultures were lysed and supernatants were analysed by SDS-PAGE. This study showed that samples induced at 37ºC had higher protein expression (*B*-Figure 1. 2) that expression at 24ºC, (*A*-Figure 1. 2) at all times evaluated. Therefore, 37ºC was selected as the optimal temperature for *E. coli* β-galactosidase expression.

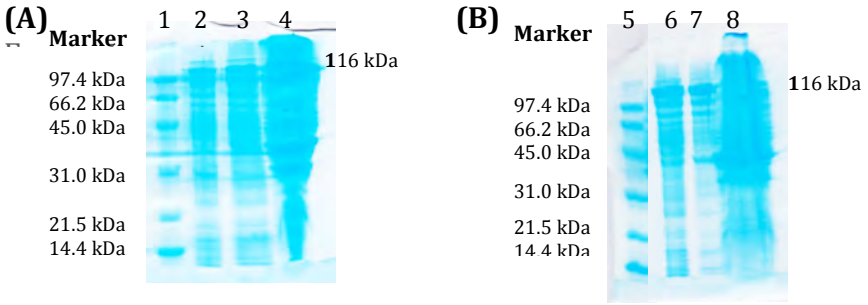


Figure 1. 2. *Analysis by SDS-PAGE (12%) of culture samples induced at different temperatures (analyses are supernatants of previously lysed samples). Protein expressed is 116kDa. (A) at 24ºC and (B) at 37ºC. Lanes: 1 and 5, marker proteins (molecular masses are indicated); 2 and 6, 2h induction samples; 3 and 7, 5h induction samples; 4 and 8, 22h induction samples. Protein expressed is 116 kDa*

Protein expression was also affected by culture's **optical density** at induction time (analysing the growth by measuring optical density at 600 nm, OD_{600}). It was detected that induction at early growth phase (OD_{600} 0.4 - 0.7) caused lower growth and less amount of expressed protein (Figure 1. 3). Induction at later growth phase allowed higher amount of cell growth and larger protein expression. Hence, protein expression was induced at the end of exponential phase (high OD_{600} of 2-3) at 37ºC.

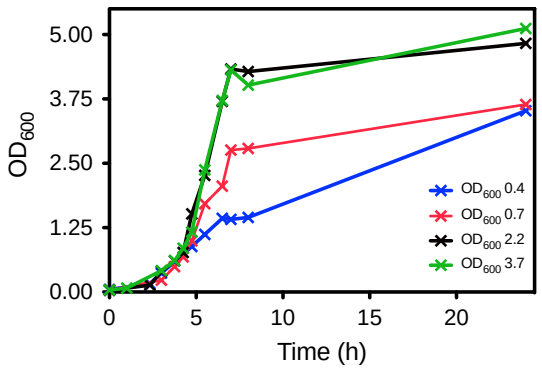


Figure 1. 3. *Growth curves of cultures (BL21 (DE3)) induced (by IPTG 1 mM) at different growth stage for the expression of the β-galactosidase from* E. coli *(subcloned on pRSF-1b). Colours: Blue - OD_{600} 0.4, red - OD_{600} 0.7, black - OD_{600} 2.2 and green - OD_{600} 3.7*

1.2 Analytic method development

Product quantification was required to characterize and describe gaxilose production by different proteins and mutants along the project. The quantification method should be quick, robust and reproducible.

Well-known and basic methods to characterize GHs are based on absorbance properties due to the hydrolytic activity of the enzymes on chromophoric substrates. On the other hand, transglycosidase activity is usually evaluated by HPLC or gas chromatography.

High performance liquid chromatography (**HPLC**) was selected as the technique to separate, identify and quantify different components in (reaction) mixtures. The Laboratory of Biochemistry at IQS School of Engineering has a collection of HPLC instruments with different detectors. Among them, two were selected to set up the analytic method:

- HPLC 1100 with UV absorbance detector and refractive index detector (RID).
- HPLC 1260 with a single quadrupole mass spectrometer (MS) detector.

Both methods are described below.

1.2.1 Absorbance and RID detector (HPLC 1100)

First, absorbance and RID detectors connected in serie were used to set up an analytical method (HPLC 1100 equipment). Absorbance detector would allow the dection and quantification of the *o*NPG substrate and therelease of *o*NP, meanwhile RID detector would determine mono-, and oligosaccharides.

The chosen stationary phase was an HILIC (*Hydrophylic Interaction Cromatography*) column with amide groups: *XBridge Amide* from Waters® (3.5 µm, 4.6 x 150 mm), (Box 1. 3). These columns are exceptional tools for the analysis of carbohydrates. The amide groups in the stationary phase compose the column and interact with carbohydrates' hydroxyls. The main drawback of this column is that it would not allow the analysis of molecules with apolar groups such as *o*NP due to the lack of interactions with the amide groups of the stationary phase.

Box 1. 3 *HILIC definition*

HILIC is a variant of normal phase liquid chromatography that mixes ion chromatography and reversed phase liquid chromatography. HILIC uses hydrophilic stationary phases with reversed-phase eluents. It provides an alternative approach to effectively separate small polar compounds on polar stationary phases.
HILIC phases can be grouped into five categories of bonded phases: unbounded silica silanol, amino or anionic, amide cationic or zwitterionic.

Different parameters were modified to optimize the method e.g. column temperature (40 - 80ºC), flow rate (0.2 - 0.6 mL/min), injection volume (2.5 - 5 µL) and mobile phase (diverse percentages of acetonitrile and water). Evaluation of the effect of these parameters was based in diverse features as resolution, retention time and efficiency (Box 1. 4).

Box 1. 4 *Chromatography parameters definitions*

Resolution, R_S, is one of the most important features to optimize. Resolution is the necessary distance to be able to distinguish between two peaks. It is calculated using the separation of two peaks in terms of their average peak width at the base ($t_{R2}>t_{R1}$). A resolution value of 1.5 or higher will ensure that the sample components are well separated, Eq. 1. 1.

$$R_s = \frac{(t_{R2} - t_{R1})}{(width_{base1} + width_{base2})/2}$$ **Eq. 1. 1.** *Resolution*

Retention time or retention factor (k) is a mean of measuring the retention of an analyte in the chromatographic column. Higher value indicates that the sample is highly retained and has spent a significant amount of time interacting with the stationary phase.

Efficiency of a chromatographic peak is a measure of the dispersion of the analyte band as it travels through the HPLC system and column; due to dispersion effects the peaks take on their "Gaussian" shape. There are many factors that contribute to peak broadening. The main one is the column itself: packing, particle size and column dimensions. Other factors have a huge influence too: injection volume, dead volume or flow rate.

Mobile phase	ACN:H_2O (0.2% TEA); 80:20
Colum temperature	70ºC
Flow rate	0.4 mL/min
Volume of injection	5 µL
Detector	RID

Table 1. 1. *Parameters of analytical method with HPLC 100 and column XBrige Amide (Waters, 3.5 µm, 4.6 x 150 mm)*

The developed methodology (Table 1. 1) let us to achieve a good peak resolution but with a long elution time (more than 20 min.) (Figure 1. 4). Moreover, broad width of peaks (3 min) hindered integration at low product concentration, feature that diminishes sensitivity. Furthermore, this method was not able to resolve different disaccharide regioisomers.

This methodology can only be used to detect transglycosidase activity at industrial conditions but not to characterize the enzyme kinetically (kinetic parameters).

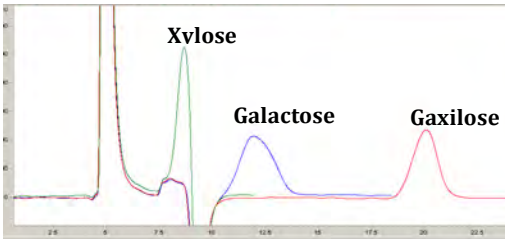


Figure 1. 4. *Chromatograms superposition of gaxilose (0.06 M), galactose (0.11M) and xylose (0.13 M) standards with RID detector, HPLC 100 (CONDITIONS: ACN:H_2O, 80:20, 0.2%TEA; 70ºC, 0.4 mL/min, volume 5 µL, XBridge Amide 3.5 µm, 4.6x150mm)*

1.2.2 Mass spectroscopy (HPLC-MS)

Mass spectroscopy offered higher sensitivity, specificity and speed of analysis. However, it is a more complex methodology and further parameters must be taken into account.

Ion mass detection

Mass spectrometry is an analytical technique that ionizes chemical species, it separates them according to their specific mass-to-charge ratio (m/z), and then records the relative abundance of each ion type (Box 1. 5).

Box 1. 5 *Ionize chemical species*

The first step in the mass spectrometric analysis is the production of gas phase ions of the compound by electrospray ionization (ESI). These ions are then separated according to their mass-to-charge ratio (m/z), by accelerating them and subjecting them to an electric or magnetic field. The instrument used was a single quadrupole analyzer (Box 1. 6). Finally, the ions are detected by a mechanism capable of detecting charge particles.

Box 1. 6 *Quadropole analyzer*

The quadrupole consists of four parallel metal rods; each opposing rod pair is connected together electrically, and a radio frequency (RF) voltage with a direct current (DC) offset voltage is applied between one pair of the rods and the others. Ions travel down the quadrupole between the rods. Only ions of a certain mass-to-charge ratio will reach the detector. This permits selection of an ion with a particular m/z or allows to scan for a range of a ms-values by continuous varying the applied voltage.

Mass spectroscopy detector was adjusted to detect diverse compounds. First, selection of the analyte ions produced in the ionization step was considered (detector in SCAN mode (Box 1. 7)). Most common analite ions detected in this work were $[M+Na^+]$ (*red*-Figure 1. 5) and $[M+NH_4^+]$ (*blue*-Figure 1. 5) (Box 1. 7. Different intensities were detected if samples were in water (*A*-Figure 1. 5) or in reaction buffer (*B*-Figure 1. 5*)*); this behaviour is explained by the presence of sodium ions in the buffer. On account of this, standards were also prepared in reaction buffer. In addition, NH_4^+ was added in the mobile phase to increase sensibility. For the analysis of gaxilose and galactose, the intensity of both adducts were summed up to build the chromatogram profile (*green*-Figure 1. 5*).*

Final detection parameters were selected to increase sensibility (Table 1. 3).

Box 1. 7 *MS detector*

MS detector can work in **scan mode** (it is used for identification of chemical components, qualitative analysis, or quantitative analysis and determination of parameters for SIM analysis) or **SIM mode** (*Single Ion Monitoring)* is used as quantitative scan whereby the molecular ion of the analyte is monitored, SIM offers more sensitivity and better selectivity); both modes were used according to the aim of the analysis.

Ion adducts are ions formed by interactions of two species, usually an ion and a molecule (for example, a Na^+ adduct of a molecule (M) is represented as $[M+Na]^+$.

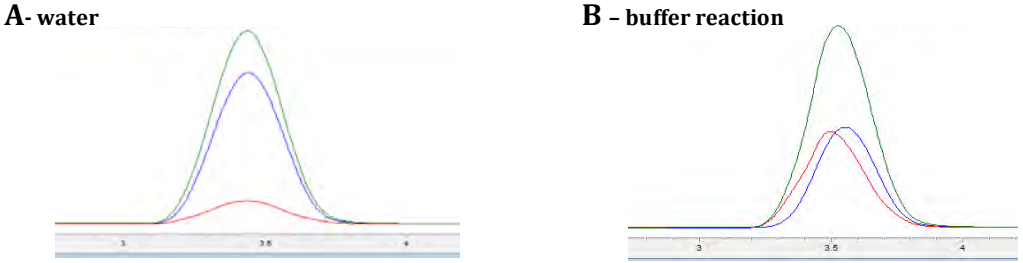


Figure 1. 5. *Chromatogram profile of gaxilose adducts by HPLC-MS. (Red-sodium adduct; blue-amonium adduct; green: sum of ammonium and sodium adducts)* A: *Gaxilose(2 mM) in water.* B: *Gaxilose (2 mM) in buffer reaction (1 mM $MgCl_2$, 100 mM phosphate, 50 mM β–mercaptoethanol, pH 7.0)*

Separation

The column selected was an HILIC (Hydrophilic Interaction Liquid Chromatography) column (Box 1. 3): *Xbrige Amide* (Waters; 2.5 µm, 2.1x100 mm), as used in the previous HPLC-UV/RID method. However, this column had smaller porous size and smaller diameter that allowed better efficiency; therefore, elution times could be reduced and resolution could be improved.

Different parameters (temperature, flow rate, mobile phase) were set in order to obtain good resolution (R_S), acceptable retention factor or retention time (k), suitable efficiency and peak symmetry (A_s) (Box 1. 4). Column temperature was increased to avoid separation of α/β anomers and it was set to 60ºC.

HILIC requires an organic mobile phase to isolate its analytes: acetonitrile and water were selected for the analysis (methanol would have been another option). Isocratic mode for analytes elution did not allow reaching good symmetry, efficiency nor acceptable resolution, even when different ratios were tested (from 95:5 to 70:30) (Figure 1. 6).

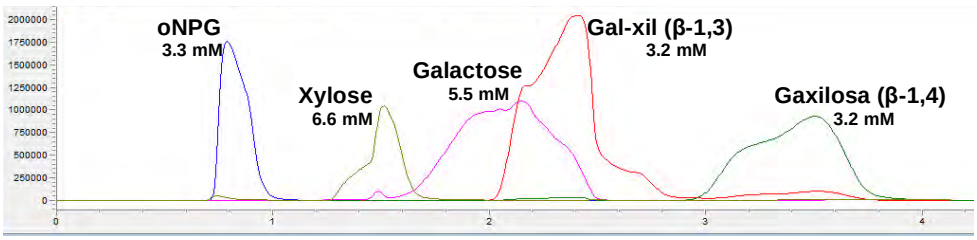


Figure 1. 6. *Superposition of chromatograms obtained with isocratic mode in HPLC-MS method Mass detector, on SIM mode, detected analite ions $[M+NH^+_4]$ and $[M+Na^+]$ for all compounds. Molecular masses (m/z) were:* oNPG *(319 and 324), xylose (168 and 173), galactose (198 and 203), gaxilose (330 and 335). Samples diluted in buffer reaction (1 mM $MgCl_2$, 100 mM phosphate, 50 mM β–mercaptoethanol, pH 7.0). (ACN:H_2O, 80:20, 60ºC,0.4 mL/min, volume 5 µL, XBridge Amide 2.5 µm, 2.1x100mm).*

Therefore, gradient elution was used to improve resolution, efficiency and symmetry (Box 1. 8).

Box 1. 8 *Gradient elution mode*

In this mode of separation, the mobile phase composition is altered during the analysis. The initial composition was chosen so that the solvent polarity is appropriate to retain and resolve early eluting analytes. The elution gradient is then increased in a predetermined way to elute compounds with optimum resolution. At the beginning of the analysis, the analyte will be wholly into the stationary phase at the head of the column. As the mobile phase changes (in the presented case, to higher aqueous content), the analyte will begin to cross into the mobile phase and move along the column.

Different gradient ratios, gradient steepness and gradient time were evaluated. Finally, the gradient used was from 95:5 to 80:20 (Table 1. 2). Flow rate was also evaluated and it was fix to 0.6 mL/min (Table 1. 3).

Time	Proportion		Description
	ACN	H_2O	
0:00	95	5	Initial conditions
0:10	80	20	Gradient (3 min)
5:10	80	20	All products have been eluted
5:11	95	5	Change to initial conditions
10:00	95	5	Recuperation of initial conditions

Table 1. 2. *Gradient program used in analytical method for HPLC-MS*

Temperature	60ºC
Flow rate	0.6 mL/min
Pressure	180 – 220 bar
Mobile phase	Gradient of ACN and water with amnium acetate

Table 1. 3. *Final conditionts of analytical method by HPLC-MS*

Xylose, galactose and three disaccharide regioisomers were the analytes to detect. Superposition of their chromatograms allow observing that analytes did not co-elute (Figure 1. 7). Avoiding peaks overlapping is mainly important for the disaccharide regiosomers which have the same molecular mass and the analyser cannot differentiate them. Even though galactose and the β-1,3 regiosomer presented partial overlap, it had not an effect on detection because the two molecules have different molecular mass. Peak symmetry was highly improved with respect to the isocratic method, as well as efficiency.

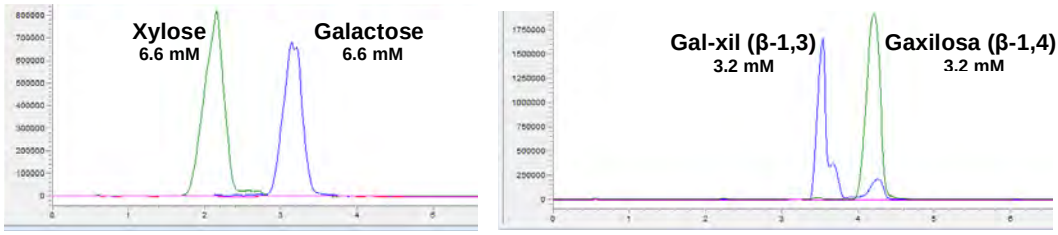


Figure 1. 7. *Superposed chromatograms of the different products (galactose, xylose, and β-D-galactopyranosyl-D-xyloses) detect by HPLC-MS , SIM mode, with gradient method (ACN:H_2O, 95:5→ 80:20, 60ºC,0.4 mL/min, volume injected 5 µL, XBridge Amide 2.5 µm, 2.1x100mm)*

Analysis by HPLC-MS would be set up for galactose, and β-D-galactopyranosyl-D-xyloses (ammonium and sodium adducts of each compound). Galactose would allow determining hydrolysis and disaccharides would allow quantifing transglycosylation.

Validation

The objective of validating an analytical procedure is to demonstrate suitability for its intended purpose. The most widely applied typical validation characteristics for various types of test are accuracy, precision, specificity, detection limit (LoD), quantification limit (LoQ), linearity, concentration range and robustness.

The validation method was applied to galactose, 4-*O*-β-D-galactopyranosyl-D-xylose (gaxilose), 3-*O*-β-D-galactopyranosyl-D-xylose and 2-*O*-β-D-galactopyranosyl-D-xylose. The analytical method had a linear range from 0.005 to 1 mM (sample concentration), when samples were diluted 50-fold (Figure 1. 8).

Standards of the three regioisomer were supplied by the Interquim S.A. team. It was checked that β-1,2, β-1,3 and β-1,4 regiosisomers had the same response factor.

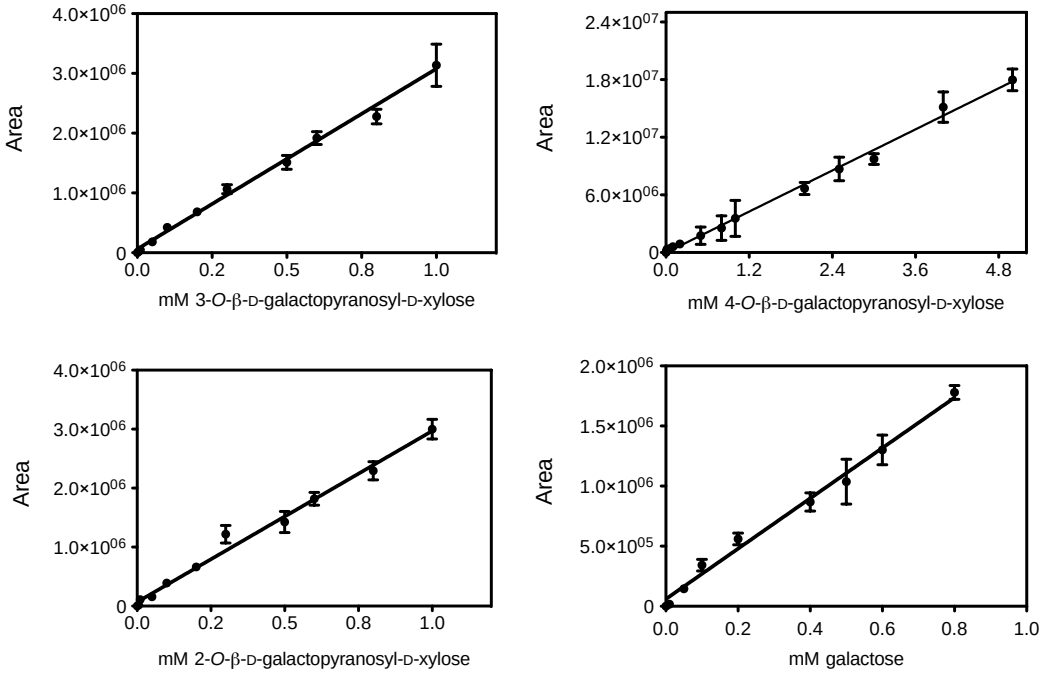


Figure 1. 8. *Standard curves of 4-O-β-D-galactopyranosyl-D-xylose, 3-O-β-D-galactopyranosyl-D-xylose, 2-O-β-D-galactopyranosyl-D-xylose and galactose*

Injection reproducibility had a standard deviation of 3-15% for galactose, β-1,2, β-1,3 and β-1,4 regiosomers. Limit of detection and quantification, LoD and LoQ (Box 1. 9), were also determined for the three compounds (Eq. 2.2 and Eq. 1.3).

$$LoD = 3 \cdot {SD_{blank}}/{slope}$$

Eq. 1. 2 *Limit of detection (SD, single deviation. Slope, area/mM)*

$$LoQ = 10 \cdot {SD_{blank}}/{slope}$$

Eq. 1. 3 *Limit of quantification (SD, single deviation. Slope, area/mM)*

Box 1. 9 *LOD and LOQ*

The limit of detection (LOD) is the lowest quantity of a substance that can be distinguished from the noise in the absence of that substance and limit of quantification (LOQ) is the lowest amount of analyte that can be quantified. LOQ and LOD were estimated based on signal to noise ratio (Eq. 1.2 and Eq.1.3).

LOQ was 3.6 µm for galactose and 1.6-1.9 µm for the disaccharides (Table 1. 4). These values were acceptable for the ranges used during enzyme characterization (The less concentrated reaction (analytic conditions) was at 5 mM, where limit of gaxilose quantification would represent 0.0038% of reaction conversion.)

	Galactose	Gaxilose	β-1,3	β-1,2
LoD	1.08 µM	0.57 µM	0.05 µM	0.48 µM
LoQ	3.6 µM	1.9 µM	1.7 µM	1.6 µM

Table 1. 4. *Limit of detection and quantification of galactose and the three disaccahriades determined by method described in Table 1. 2 and Table 1. 4*

Different reactions were performed during the project at different substrate concentrations: different dilutions were used in each reaction to adjust the linear range to each reaction requirement.

1.3 Enzyme characterization

This part has the main goal of describe reactions used for the enzyme characterization during the entire work. As previously described, GHs present two activities and therefore both activities should be taken into account to describe and compare enzymes. Two reactions will be monitored for each enzyme studied in this work: glycosidase and transglycosylase reaction.

1.3.1 Glycosidase reaction

It refers to the hydrolase activity of the enzyme with *o*NPG susbtrate (Figure 1. 9). Activity for this process is referred as *glycosidase activity* or *glycosidase specific activity* and it will be measured by the rate of aglycon release (*o*NP) by the spectrophotometric method (absorbance at 405 nm).

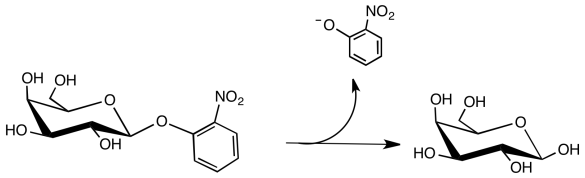


Figure 1. 9. *Activity on oNPG substrate. Detection by* o*NP release (Abs 410 nm)*

Glycosidase reaction would allow defining activity in *Units, U*. One unit hydrolyses 1.0 µmole of *o*-nitrophenyl β-D-galactoside to *o*-nitrophenol and D-galactose per minute at pH 7.0 at 37 °C.

Enzyme units (U) are determined with overall activity on *o*NPG and they are used to stablish the amount of enzyme needed in the reaction at industrial scale. The amount of enzyme *per* mg of protein would not assure activity needed; it would modify time reaction from batch to batch or even it would modify gaxilose reached.

1.3.2 Transglycosylase reaction

The transglycosylase reaction uses *o*NPG as donor and xylose as acceptor (Figure 1. 10). For this reaction, different competing processes take place and therefore expression of activities and specific activities need to be clarified

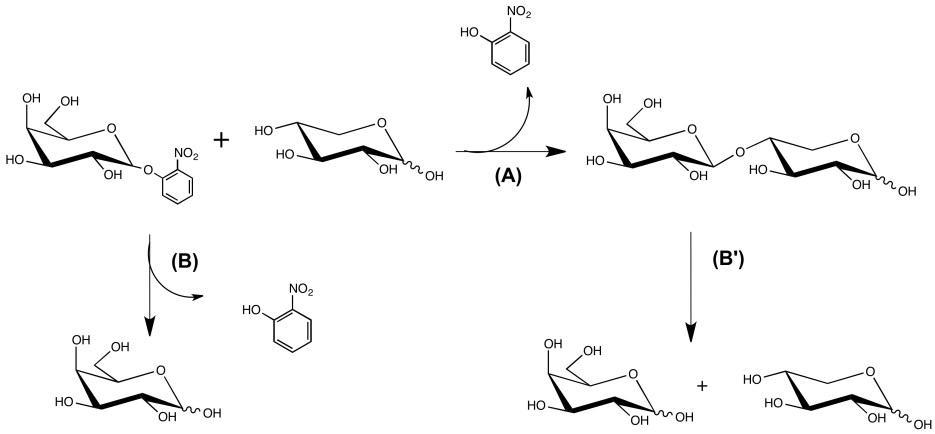


Figure 1. 10. *β-Galactosidase reaction of transglycosidation (A), primary hydrolsysis (B) and secondary hydrolysis (B') using* o*NPG and xylose as a substrates*

- *Transglycosidase activity* or *transglycosidase specific activity* refers to the rate of gaxilose product formation, or other regioisomer, monitored by HPLC-MS (*A*-Figure 1. 10).
- *Hydrolase activity* or *hydrolase specific activity* refers to the galactose produced during the *transglycosylation reaction*, corresponding to both hydrolysis of the *o*NPG substrate (primary hydrolysis, *B*-Figure 1. 10) and hydrolysis of the gaxilose transglycosylation product (secondary hydrolysis, *B'*-Figure 1. 10). Galactose formation is monitored by HPLC-MS. To distinguish between

hydrolase activities in glycosidase or transglycosylase reaction, we will use *glycosidase activity* for the reaction with only *o*NPG substrate (the native hydrolase activity of the enzyme), and *hydrolase activity* for the primary and secondary hydrolysis in the transglycosylation reactions.

Transglycosylase reactions were performed at different substrate concentration during the project to evaluate the enzymes under different conditions.

- *LABORATORY CONDITIONS* (50 mM *o*NPG, 250 mM xylose)
- *ANALYTICAL CONDITIONS* (5 mM *o*NPG, 25 mM xylose)
- *INDUSTRIAL CONDITIONS*
    - 0.1 g *o*NPG scale (0.1 M *o*NPG, 0.4 M xylose).
    - 10 g or 50 g *o*NPG scale at Interquim facilities
    - 12 kg *o*NPG scale at Interquim facilities (1000 L reactor)

1.3.3 Transglycosylation to hydrolysis ratio (T/H ratio)

The term *transglycosylation to hydrolysis ratio*, T/H ratio, will then refer to the ratio between *transglycosylation activity* or *transglycosylation rate* and *hydrolase activity or rate* (meaning primary and secondary hydrolysis in the transgycosylation reaction, not to confuse with *glycosidase activity*). It can be determined as ratio of specific activities, ratio of initial rates or as the representation of transglycosylase product (gaxilose or other regioisomer) *versus* hydrolysis product (galactose).

1.4 Structural analysis

E. coli β-galactosidase has been extensively studied, mainly for its importance in molecular biology (e.g. the *lac* operon model). Its catalytic mechanism is fully understood and the amino acids involved in catalysis are well described (Box 1. 10).

The protein 3D structure has been solved under different conditionsfor the wild type enzyme and different mutants, with or without substrates or with diverse inhibitors mimicking transition states (e.g. galactonolactone, PDB nº 3VDB). PDB database (Protein Data Bank, www.rcsb.org [185,186]) includes 59 structures of *E. coli* β-galactosidase (as of October, 2017).

Box 1. 10 *Mechanism details of the β-galactosidase from* E. coli

Substrate binding occurs in two steps, as described in *i. INTRODUCTION* (see *i.1.4.2.3.1.1*). Lactose, the natural substrate, binds in what is defined as *shallow* mode, stacking to Trp999 with the galactoside oxygen more or less centred over the indole. This binding mode is non-productive, and in order to proceed with the reaction, the substrate has to move 3 Å and rotate 90º. This position is described as *deep* binding mode. In this geometry, the galactose moiety of the lactose interacts with Trp568 and is correctly positioned between the two catalytic residues [168]. Conformational changes occur between the two conformational binding modes: 60º rotation of Phe601 and a movement of 11 Å of a mobile loop (Gly794-Pro803) that closes over the active site.

E. coli β-galactosidase is a bifunctional enzyme catalysing the hydrolysis of lactose to galactose and glucose as well as the intramolecular isomerization of lactose to allolactose. Even at lower concentration of lactose, this enzyme reacts with lactose to produce 50% allolactose (transglycosylation). The mobil loop helsp maintain glucose inside the catalytic site and achieving high transglycosidation yields. This mobile loop has paramount effect of the T/H ratio.

1.4.1 Closed loop

Several residues are involved in loop movement and stabilisation of the closed and open conformations, as described in *i. INTRODUCTION* (see *i.4.2.3.1.1*): Gly794 [174,187], Ser796 [175], Arg599 [176,177], Met542, Phe601 [178] and Glu808 [175]. The movement of this loop is favored by Gly794, located at the beginning of the loop, and it is described as a hinge for the loop. It has been proved that mutations of this residue to alanine or less flexible residues favour the closed conformation [174].

Dr Alberto Marina's group from IBV-CSIC proposed the construction and crystallyzation of a mutant with the mobile loop mainly closed. G794A mutant was constructed with two main objectives. A) Structural purposes: closed loop would help maintaining gaxilose in the catalytic site and it could be trapped during crystallization assays. Therefore, it would describe enzyme-gaxilose interactions (useful to design mutants to increase transglycosidase activity). B) Transglycosidase analysis: how would transglycosylation and hydrolysis reactions, and consequently its T/H ratio, be affected by the closed loop that could hinder substrates' entrance.

Juers and co-workers had studied G794A mutant previously [174]. Their kinetic analysis showed that mutant G794A reduces glycosidase activity (*o*NPG) and transglycosidase activity, using lactose as substrate, with respect to the wild type enzyme. Its crystal structure demonstrated that the 794-803 loop was closed. Glucose was not entrapped in any crystal structure; however, kinetic evaluation proved that G794A mutant binds glucose better than the wild type enzyme. This relatively good binding of glucose to the free enzyme could indicate that the closed loop creates a preformed site for glucose binding and that the acceptor binding site is more complementary to the structure of glucose (the loop must cause the residues that bind glucose are slightly different positioned so that binding is better). However, Ala to

Gly794 substitution should draw glucose out of a reactive position and thus decreases transglycosidase activity.

Transglycosylase reaction for gaxilose production did not use glucose, but xylose. Xylose, compared to glucose, does not have the 6-hydroxymethyl group: this change could allow binding on different way in preformed subsite +1. Will this *different* binding be in a productive conformation? Could this binding imply improved transglycosidase activity?

Glycosidase, hydrolase and transglycosidase activities

Firstly, glycosidase specific activity on *o*NPG of G794A was evaluated and compared to the wild type protein (Table 1. 5). Mutant G794A showed a 15-fold decrease of activity compared to the wild type enzyme, as described previously [174]. It could be caused by the blockage of the catalytic centre due to the closed loop that impeded *o*NPG entrance. Simultaneously, it could also impede release of products formed (galactose and *o*NP); which would delay continuous activity of the enzyme.

	Glycosidase activity (V_o/[E], s^{-1})
Wild type	2,902.9 ± 132.5
G794A	190.1 ± 9.8

Table 1. 5. *Glycosidase specifc activity of the wild type and the G794A mutant of the* E. coli *β-galactosidase CONDITIONS:* o*NPG (13 mM), buffer 1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol, pH 7.0, 37ºC*

Transgalactosidase and hydrolytic specific activities of wild type and G794A were determined. ***Transglycosylation specific activity*** was nearly 75-fold higher for the wild type than for the G794A mutant. Meanwhile ***hydrolysis specific activity*** was only 5.5-fold higher for the wild type (Table 1. 6). Differences between two proteins were also reflected, in the T/H ratio: it was 13-fold higher for the wild type enzyme.

	Specific activity (min^{-1})		**T/H**
	Transglycosylation (V_o gaxilose/[E])	**Hydrolase** (V_o gaxilose/[E])	
Wild type	2.25 ± 0.12	177.90 ± 10.67	0.012
G794A	0.030 ± 0.001	32.20 ± 4.17	0.0009
Wild type / G794A	74.9	5.5	

Table 1. 6. *Specific activity of transglycosylation and hydrolysis of the wild type enzyme and G794A mutant of the* E. coli *β-galactosidase CONDITIONS:* o*NPG (5 mM), xylose (25 mM), 0.01-0.1 U of enzyme buffer 1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol, pH 7.0 at 37ºC*

For the industrial production of gaxilose, the amount of enzyme used (the units of enzyme/ weight of substrate) is determined by the measurement of glycosidase activity on *o*NPG (one unit hydrolyses 1.0 μmole of *o*-nitrophenyl β-D-galactoside to *o*-nitrophenol and D-galactose per minute at pH 7.0 at 37 °C).

	Specific activity ($mM \cdot min^{-1} \cdot U^{-1}$)		T/H
	Transglycosidase (V_o gaxilose/U of enzyme)	**Hydrolase** V_o gaxilose/U of enzyme)	
Wild type	$(1.58 \pm 0.13) \cdot 10^{-3}$	0.13 ± 0,01	0.012
G794A	$(1.98 \pm 0.13) \cdot 10^{-3}$	2.14 ± 0.28	0.0009
G794A / Wild type	1.2	16.5	

Table 1. 7. *Specific activity of transglycosidase and hydrolase activities of the wild type enzyme and G794A mutant of the* E. coli *β-galactosidase taking into account Units of enzyme employed. CONDITIONS: oNPG (5 mM), xylose (25 mM), 0.01-0.1 U of enzyme, buffer 1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol, pH 7.0 at 37ºC*

Initial rates could also be compared using the same number of glycosidase units (Table 1. 7 and Figure 1. 11). Transglycosidase activity per unit of glycosidase activity showed slightly higher activity for the G794A enzyme than for the wild type enzyme, 1.2 fold higher (*right*-Figure 1. 11). On the other hand, hydrolase activity per unit of glycosidase activity (*left*-,-Figure 1. 11) exhibited that G794A had more than 10-fold higher activity than the wild type enzyme. Therefore, the T/H ratio was favourable for the wild type enzyme (d*own*-Figure 1. 11).

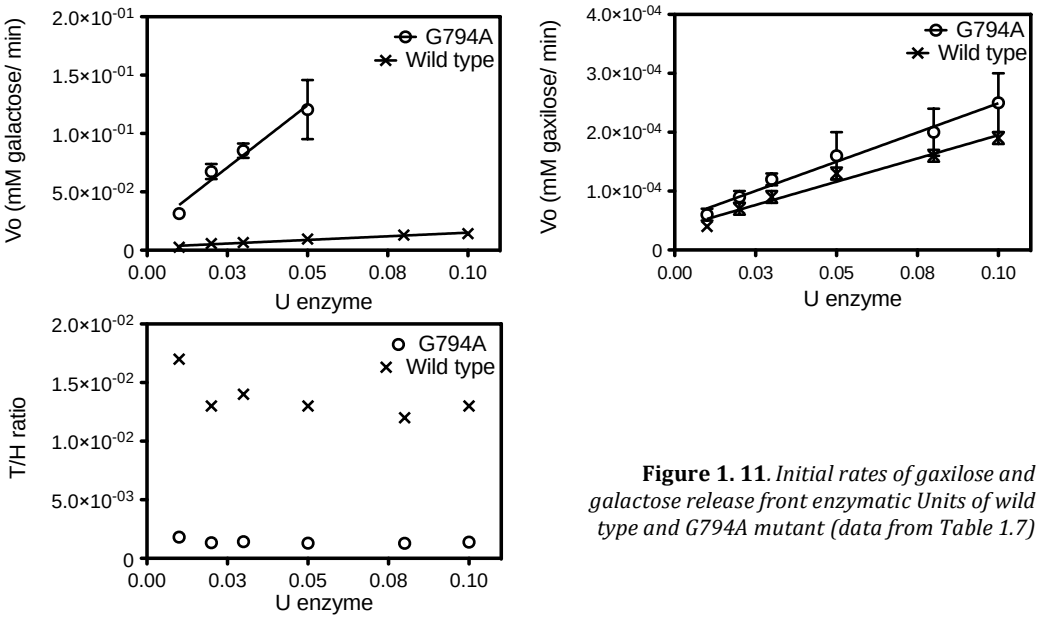


Figure 1. 11. *Initial rates of gaxilose and galactose release front enzymatic Units of wild type and G794A mutant (data from Table 1.7)*

After the evaluation of initial rates under analytical conditions, **industrial conditions** (0.1 M *o*NPG, 0.4 M xylose and 2.5 enzyme units (U)) were assayed. The amount of gaxilose with time showed that G794A synthesised 3 to 4-fold less gaxilose than the wild type enzyme (*left*- Figure 1. 12*).* Transgalactosidase/hydrolysis ratio, calculated as the quotient of gaxilose concentration *versus* galactose concentration, assessed that G794A mutant has a low T/H ratio (*right*- Figure 1. 12) as it was shown in the specific activity analysis.

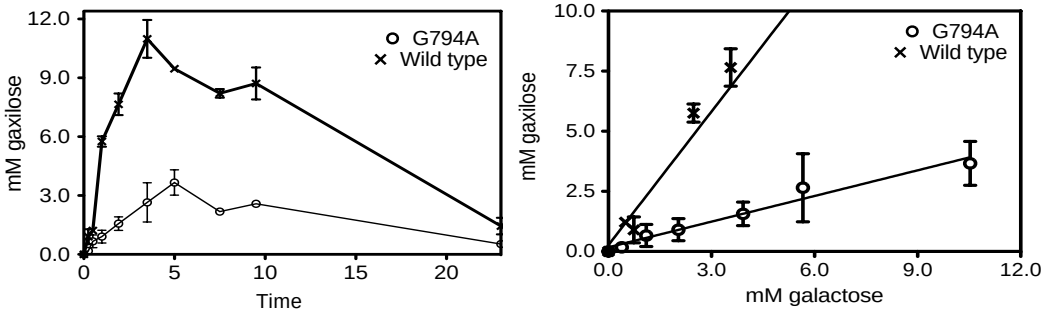


Figure 1. 12. *Industrial transglycosylation reaction catalyased by* wild type *enzyme and G794A mutant of* E. coli *β-galactosidase.* Left: *gaxilose amount on time.* Righ*t: amount of gaxilose front amount of galatcose by* wild type *enzyme and G794A mutant. CONDITIONS:* o*NPG 0.1 M, xylose 0,4 M, 1 U of enzyme, buffer 1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol, pH 7.0 at 37ºC*

G794A has the mobile loop mainly closed. It was hypothesised that preformed subsite +1 could change xylose binding and it could improve transglycosidase activity. Initial rates and long reaction times showed that G794A did not improve gaxilose production mainly due to its low T/H ratio and notable activity reduction. Decrease of activity could be caused by the lack of xylose binding or by a xylose binding in a non-productive mode as hypothesized with glucose.

Michaelis-Menten

Michaelis-Menten analysis was performed in order to understand G794A behaviour (Table 1. 8). The study was carried out calculating transglycosidase initial rate (*left*- Figure 1. 13) and hydrolase initial rate (*right*- Figure 1. 13*).* Determination of kinetic parameters for gaxilose synthesis was only feasible the for wild type enzyme; where it had a value of k_{cat} of 1,044 ± 56 min^{-1} and a K_M of 0.39 ± 0.01 mM.

	Transglycosylation		**Hydrolysis**	
	k_{cat} (min^{-1})	**K_M (mM)**	**k_{cat} (min^{-1})**	**K_M (mM)**
Wild type	1,044 ± 56	0.39 ± 0.01	84,481 ± 6,896	1.06 ± 0.28
G794A	-	-	20,491 ± 1599	1.17 ± 0.28

Table 1. 8. *Kinetic parameters k_{cat} and K_M for transglycosidase and hydrolase activity of enzyme wild type of* E. coli *and G794A mutant. CONDITIONS: 0.05-10 mM* o*NPG, 25 mM xylose, 0.01-0.03 U of enzyme, 37ºC, pH 7.0, buffer:100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$*

On the other hand, evaluation of the hydrolase activity was possible for both enzymes and differences in catalytic efficiency were found. k_{cat} values were 84,481 ± 6,896 min^{-1} for wild type enzyme and 4-fold less for G794A mutant (20,491 ± 1599 min^{-1}). K_M values were the same for both enzymes (1.06 ± 0.28 mM for wild type and 1.17 ± 0.28 mM for G794A). This reflected that loop movement did not affect *o*NPG binding, as had been described previously [174]. Therefore, slower rate of G794A could be affected by the difficulty on product release (galactose and *o*NP) and not for *o*NPG entrance.

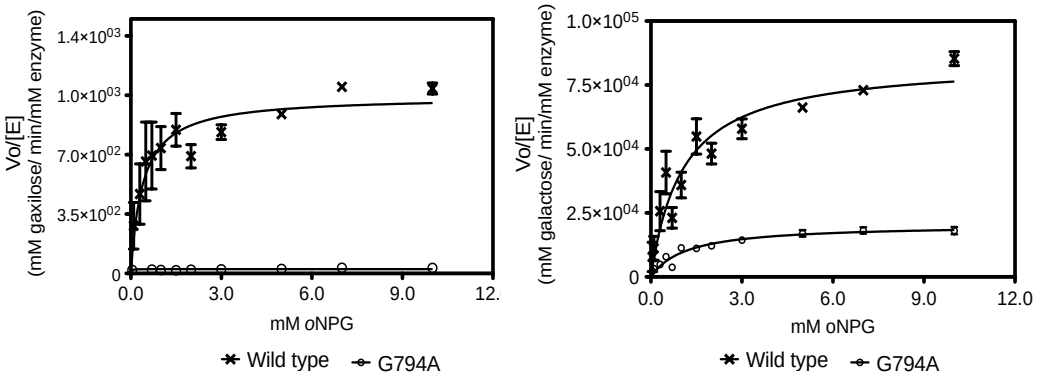


Figure 1. 13. *Michaelis-Menten profile of the G794 mutant and the wild type enzyme of the β-galactosidase from* E. coli. Left) transglycosidase activity. Right) hydrolase activity. *CONDITIONS: 0.05-10 mM* o*NPG, 25 mM xylose, 0.01-0.03 U of enzyme, 37ºC, pH 7.0, buffer:100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$

G794A had slower transglycosidase specific activity (min^{-1}) than wild type and it had lower T/H ratio: both features discarded this mutant as a good candidate to improve gaxilose production.

It was concluded that the closed loop did not improve transglycosylation, nor did it increase the T/H ratio. Moreover, G794A also had 5 to 7-fold lower hydrolase and glycosidase specific activity than the wild type. It was assumed that *o*NPG substrate could enter to catalytic subsite, but xylose could not or it could be binding in a non-reactive mode.

1.5 Industrial application

Modifying the reaction conditions could optimize the process of industrial production of gaxilose. Improvement of product yields and decrease of costs were goals of the project that were developed at Interquim premises.

Enzymatic activity had been evaluated previously in the project analysing different process parameters (temperature, pH and use of co-solvents). However, some features could be still modified, i.e. substrate concentration, donor:acceptor ratio and enzyme amount. Different reaction conditions were analysed at 10 g (*o*NPG) scale and then they were evaluated at 12 kg (*o*NPG) scale.

1.5.1 Scale up: 10 g

A ten-gram (*o*NPG) reaction scale was used to evaluate different reaction parameters: donor:acceptor ratio, substrate concentration, enzyme units added and reaction time. The actual substrates (xylose and *o*NPG), temperature (37ºC) and reaction time (~24h) were invariable.

The first approach to set up reaction conditions was performed with commercial β-galactosidase from *E. coli* (Sigma) (*exp. 1-4* Table 1. 9). Different substrate ratios,

Per motius de confidencialitat les pagines 87-89 han estat eliminades d'aquest manuscrit.

CLOSING REMARKS

This chapter includes the evaluation of the *E. coli* β-galactosidase at different levels.

- Analytical method for transglycosidase reaction was set up to evaluate galactopyranosyl-D-xylose regioisomers production and galactose release. The method allowed quantifying hydrolase and transglycosidase activity by HPLC-MS. LOD and LOQ allowed evaluate reactions at used concentrations.
- The best found conditions for *E. coli* β-galactosidase protein expression were set: 1 mM IPTG at OD_{600} of 2-3 for 18h at 37°C.
- The *E. coli* β-galactosidase mutant G794A showed 15-fold less activity on *o*NPG than wild type. It had nearly 80-fold less transglycosidase specific activity than wild type and 5.5-fold less hydrolase specific activity. T/H ratio evaluation showed that G794A has a 13-fold lower ratio compared to wild type.
- G794A mutant was also evaluated at industrial level and it showed 3 to 4-fold lower synthesis of gaxilose.
- Michaelis-Menten transglycosylation kinetic parameters of G794A could not be determined due to its negligible activity at low substrate concentration. Michaelis-Menten parameters of hydrolysis activity showed 4-fold lower activity compared to the wild type protein. Both enzymes showed the same K_M. This could be indicating that the loop in the closed conformation had no effect on *o*NPG binding.
- It is tempting to speculate that the almost closed loop of G794A mutant could obstruct xylose entrance in the catalytic site or rather bind xylose in an un-reactive mode and thus impedes transglycosylation.
- G794A was discarded for gaxilose production due to its low transglycosidase activity and low gaxilose yield reached.
- [redacted]

CHAPTER 2. Selection of candidates

CHAPTER 2. SELECTION OF CANDIDATES

OPENING

Gaxilose, the active pharmaceutical ingredient (API) in Lactest®, a non-invasive method for hypolactasia diagnosis, is produced enzymatically by β-galactosidase from *E. coli* (*i. INTRODUCTION*). In this thesis, the industrial production of gaxilose has been optimized achieving a yield of 22% of pure product after several purification steps (*CHAPTER 1: 1.5.2.2. Industrial production)*. Reaction conditions were modified to increase production and to reduce costs (*CHAPTER 1,* see *1.5.2*). The catalyst, β-galactosidase enzyme, would be the next reaction factor to be optimized. In this research project, a **new β-galactosidase** was searched to increase gaxilose production yield.

The goal in this chapter is to find a new enzyme with improved characteristics for transglycosylation compared to the current enzyme (*E. coli* β-galactosidase): higher transglycosidase activity and regioselectivity and lower hydrolase activity. **Search** for candidates was chosen as the first step to select a better biocatalyst. Expression, characterization and evaluation of candidates were necessary for **final selection**.

2.1 Search criteria

The search for new β-galactosidases able to catalyze gaxilose formation started with a selection of enzymes in the CAZY database-(Carbohydrate-Active enZYmes) [80,188]. This database describes a classification of enzymes involved in the breakdown, biosynthesis, modification and metabolism of carbohydrates. It includes: glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA), and carbohydrate-binding (CBM) families. CAZy is a dynamic database that is constantly updated: novel enzymes are included whenever a new carbohydrate acting enzyme is described or annotated. CAZy was used as source to find potential β-galactosidase candidates.

β-Galactosidases are GH enzymes (E.C. 3.2.1.23). CAZy classifies GHs in 149 families (Box 2. 1). β-Galactosidase enzymes can be mainly found in four families: GH1, GH2, GH35 and GH42 †. All of them are grouped in clan GH-A and share an 8 TIM-barrel module. In general, GH1 and GH2 galactosidases are found in mesophilic organisms and they have lactose as natural substrate, while enzymes belonging to families GH35 and GH42 are usually found in thermophiles and preferentially act on other galacto-derivates [166]. Within families GH1, GH2, GH35 and GH42, 29,670 proteins are included (as of August 2017). In order to reduce the number of candidates, only enzymes described and characterised (with an assigned E.C. number) were chosen; those were 619 in total. To further filter the candidates, only those defined as β-galactosidases were selected for further analysis (137 in total, Table 2. 1).

Box 2. 1 *CAZy therminology*

Families are enzymes grouped based on amino acid sequence similarities. Furthermore, there is a direct realtionship between sequences and folding similarities

Clans are groups of families sharing a fold and catalytic machinery.

	GH1	GH2	GH35	GH42	Total
All	15,727	9,905	1,865	2,173	29,670*
Characterised	336	147	72	64	619*
β-galactosidase	7	63 (86*)	39 (65*)	28 (64*)	137 (221*)
Transglyco. activity described	5	19	9	12	45

Table 2. 1 *Summary of evaluated galactosidases enzymes found in families GH1, GH2, GH35 and GH42 on June 2014 (*Data as of August 2017)*

† Families GH59 and GH147 also include one β-galactosidase (one in each family), however these β-galactosidases had not been expressed neither characterised, they have been only identified [394].

2.2 Bibliographic search

2.2.1 Transglycosidase activity reported

A total of 137 enzymes characterized as β-galactosidases were found in CAZy. A more detailed evaluation was performed with the aim of reducing the number of candidates and selecting a manageable amount of enzymes to express, characterise and evaluate. A bibliographic study was developed to search for enzymes with described transglycosydase activity in the literature.

Among 137 β-galactosidases, 45 enzymes were described to have transglycosylation capacity. Their main characteristics are summarized (Table 2. 2) and a more detailed description is available in the *Appendix* (*A. 1*). Most of the enzymes were described both by hydrolase and transglycosydase activities. *o*NPG and lactose were found to be the most commonly employed substrates to characterize those enzymes. This was very useful for our candidate selection because of the coincidence of *o*NPG as substrate and because of the structural similarity between lactose and gaxilose.

Organism	GH	Hydrolysis Product(s)	Transglycosylation Product(s) (substrate used)	Ref.
Pyrococcus furiosus	GH1	Lactose (K_M 186 mM; k_{cat} 3,800 s^{-1}) *o*NPGal (K_M 5.3 mM; k_{cat} 10,000 s^{-1})	40% tri- and tetrasaccharide: β-1,3> β-1,6 (70%, w/v, lactose)	153,155, 189–194
Pyrococcus woesei		*o*NPGal (K_M 1.2 mM)	Galβ(1,4)GlcNAc-linker-tBoc (also β-1,6) (lactose as donor, GlcNac-link-tBoc as acceptor)	195–199
Sulfolobus acidocaldarius		*o*NPFuc (K_M 0.12 mM; k_{cat} 26.4 s^{-1}) *p*NPFuc (K_M 0.14 mM; k_{cat} 18.6 s^{-1}) *o*NPGlu (K_M 0.14 mM; k_{cat} 16.7 s^{-1}) *o*NPGal (K_M 3.19 mM; k_{cat} 31.5 s^{-1})	25-30 g/L tri- and terasaccharides (pNPGal as a donor and Gal as a acceptor or lactose)	200
Sulfolobus solfataricus		*o*NPGal (K_M 1.1 mM; k_{cat} 1,300 s^{-1}) Lactose (K_M 196 mM; k_{cat} 1,500 s^{-1})	52% GOS: tri- and tetrasaccharides (600 g/L lactose) (major product: β -D-Galp-(1–6)- β -D-Galp-(1–4)- β -D-Glcp) (58% F359Q; 62% F441Y)	153,154, 201–205
Sporobolomyces singularis (Bullera singularos)		*o*NPGal (K_M 5.40 mM; V_{max} 3.07 µmol min^{-1}·mg^{-1}) *p*NPGlc (K_M 1.96 mM; V_{max} 2.30 µmol $min^{-1}mg^{-1}$)	90 g/L GalOS (180 g/L) 55% oligosaccharides. Tetra saccharides or higher GlcOS (cellobiose) > di-, tri- and tetra- GalOS (300 g/L lactose)	206 207–210
Arthrobacter psychrolactophilus strain F2	GH2	Lactose (K_M 42.1 mM; k_{cat} 3.02 s^{-1}) *o*NPGal (K_M 2.7 mM; k_{cat} 12.07 s^{-1}) pNPGal	Trisaccharides	211,212
Arthrobacter sp. C2-2		Lactose (K_M 53.1 mM; k_{cat} 1,106 s^{-1})	34 mM trisaccharides 6 mM tetrasaccharides (680 mM lactose)	213

Arthrobacter sp. 32cB	Lactose (K_M 1.52 mM; k_{cat} 30.55 s^{-1}) *o*NPGal (K_M 16.56 mM; k_{cat} 31.84 s^{-1})	Trisaccharides, tetrasaccharides, low amount pentasaccharides (292-582 mM lactose) Aklyl glucosides (2-propyl, 1-butyl, 1-hexyl, cyclohexyl) Heterooligosaccharides (fructose, xylose, arabinose)	214
Bacillus circulans	Lactose (V_{max} 158.5 U·min^{-1}·mg^{-1}) *o*NPGal (K_M 45.9 mM; V_{max} 277.7 U·mg^{-1}	64% yield (β-1,4) 65% (β-1,3 and β-1,4) changing product synthesise	215–219 220
Bacillus megaterium	Lactose (K_M 12.6 mM; k_{cat} 54.4 mM/min) *o*NPGal (K_M 9.5 mM; V_{max} 16.6 mM/min)	Heterooligosaccharides: fructose, glucose, sorbitose, sucrose, cellobiose, sorbitol, mannitol, alkyl alcohol among others (with *o*NPG as donor)	221
Bifidobacterium bifidum (DSM 20215)	*o*NPGal	GOS	222,223
Bifidobacterium bifidum (NCIMB 41171)	*o*NPGal	59% di- and trisaccharides (43% w/w lactose)	109,224–227
Enterobacter agglomerans	Lactose (K_M 114 mM; k_{cat} 43 s^{-1}) *o*NPGal (K_M 0.06 mM; k_{cat} 180 s^{-1})	38% GalOS (13%, w/v, lactose) (*o*NPGal as donor and difernts acceptors inlcudincs hexose, pentose, alchohol, etc..)	228
[redacted]	*o*NPGal (K_M 0.01 mM; V_{max} 2 mM/min) Lactose (K_M 0.30 mM; V_{max} 0,21 mM/min)	49% GOS (275 g/L lactose) Heterooligosaccharides: galactose,, glucose, sucrose, trehalose, sorbitol, salicin(with *o*NPG as donor)	229–231
Klebisella pneumoniae	*o*NPGal (K_M 0.72 mM; k_{cat} 121 s^{-1})	43.2 g/L 1-lactulose (400 g/L lactose and 200 g/L fructose)	232–234
Klyveromyces lactis		Lactulose 14.4 g/L; allo-lactulose 47.5 g/L (200 g/L lactose and 20 g/L fructose)	235–238
Lactobacillus acidophilus R22	*o*NPGal (K_M 0.73 mM; V_{max} 28.8 µM/min/mg prot) Lactose (K_M 4.04 mM; V_{max} 361 µM/min/mg prot)	39% GOS :16% disaccharides, 19% trisaccharides, 4% tetrasaccharides	239
Lactobacillus delbrueckii sbsps. bulgaricus DSM-20081	Lactose (K_M 19.2 mM; k_{cat} 234 s^{-1}) *o*NPGal (K_M 0.92 mM; k_{cat} 603 s^{-1})	GOS (102 g/L): disaccharides and trisaccharides (β-1,6> β-1,3)	240
Lactobacillus plantarum WCFS1	Lactose (K_M 29 mM; k_{cat} 98 s^{-1}) *o*NPGal (K_M 0.9 mM; k_{cat} 92 s^{-1})	41% GOS (w/w); β-1,6> β-1,3 (600 mM lactose)	241
Lactobacillus sakei	Lactose (K_M 20 mM; k_{cat} 43 s^{-1}) *o*NPGal (K_M 0.6 mM; k_{cat} 180 s^{-1})	41% GalOS (lactose 215 g/L): disaccharides and trisaccharides (β-1,6> β-1,3)	242
Pseudoalteromonas sp 22b	Lactose (k_{cat}/K_M 47.5 mM^{-1}·s^{-1}; k_{cat} 157s^{-1}) *o*NPGal (K_M 0.06 mM; k_{cat} 180 s^{-1})	Alkyl galactosides: 32 g/L 2-propyl galactopyranoside, 30 g/L 1-butyl galactopiyranoside (lactose and alcohols), 30% water	243–245
Saccharopolyspora rectivirgula	Lactose (K_M 0.75 mM; k_{cat} 63.1 s^{-1}) *o*NPGal (K_M 0.03 mM; k_{cat} 47.6 s^{-1}) *p*NPGal (K_M 0.04 mM; k_{cat} 55.8 s^{-1})	41% (w/w) (1.75 M lactose)	246–250

Streptococcus thermophilus			Lactose (K_M 6.90 mM) *o*NPGal (K_M 0.98 mM)	40% oligosaccharides: di- and trisaccharides (initial lactose 70% (w/w))	251–253
Thermotoga maritima			Lactose (K_M 1.6 mM; V_{max} 63.3 µmol·min^{-1}·mg^{-1}) *o*NPGal (K_M 0.33 mM; V_{max} 79.6 µmol·min^{-1}·mg^{-1})	97 g/L GOS: tri- and tetrasaccharides , β-1,3 (500 g/L lactose)	156,157, 254,255
			*o*NPGal (K_M 0.33 mM; V_{max} 79.6 µmol min^{-1}mg^{-1}) Lactose (K_M 11.6 mM; V_{max} 27.8 µmol min^{-1}mg^{-1})	GOS: Two di- and two tri-saccharides (30%, w/v, lactose)	254,256, 257
Aspergillus niger	GH35			Mainly disaccharides and low amount of trisaccharides (lactose)	258
Aspergillus candidus (GCMCC3.2919)				32% GOS (400g/L lactose)	259
Aspergillus oryzae				27% GOS: trisaccharides (50% lactose) Galactosyl-polyols (*o*NPG as donor)	260–263
Bacillus circulans			*p*NPGal (V_{max} 48.7 U·mg^{-1}) Lactose (V_{max} 61.0 U·mg^{-1}) *o*NPGal (V_{max} 12.7 U·mg^{-1})	Trisaccharides, β-1,3 (lactose)	264
Carnobacterium maltaromaticum (Carnobacterium piscicola)			*o*NPGal (K_M 1.7 mM V_{max} 450 µmol·mim^{-1}·mg^{-1})	Hypothesis	265,266
**Hypocrea jecorina (Trichoderma reese)*			*o*NPGal (K_M 0.36 mM; k_{cat} 17.31 s^{-1}) Galactobiose (K_M 9.06 mM; k_{cat} 5.68 s^{-1})	Disaccharides	267
Paenibacillus thiaminolyticus			*p*NPFuc (K_M 1.18 mM; V 1.7 µmol min^{-1} mg^{-1}) *p*NPGal (K_M 250 mM; V 270 µmol min^{-1} mg^{-1}) *p*NPGc (K_M 77 mM; V 0.1 µmol min^{-1} mg^{-1})	*p*NPFuc as donor with different acceptors: 69% (*p*NPMan; β-1,3), 42% (*p*NPGal; β-1,6), 94% (methanol), 72% (ethanol)	268
Penicillium expansum				44 % GOS (100 g/L lactose)	269
Penicillum sp.				70% Gal-β-1,6-Gal (AcGal as substrate)	270
Alicylcobacillus acidocladarius				wt: *p*NPGal (K_M 4 mM; k_{cat} 2,400 s^{-1}) Lactose (K_M 60.8 mM; k_{cat} 5,212 s^{-1}) E157G: *p*NPGal (K_M 0.1mM; k_{cat} 0.30 s^{-1}) Lactose (K_M 0.35 mM; k_{cat} 16.8 s^{-1})	271,272
Arthrobacter sp.				Lactulose (40% lactose, 20% fructose , w/v)	273
Bacillus circulans	GH42		*p*NPGal (V_{max} 169.8 U·mg^{-1}) Lactose (V_{max} 44.3 U·mg^{-1}) *o*NPGal (V_{max} 47.2 U·mg^{-1})	Tri- and tetrasaccharides (lactose)	264,274, 275
			*p*NPGal (V_{max} 46.4 U·mg^{-1}) Lactose (V_{max} 45.3 U·mg^{-1}) *o*NPGal (V_{max} 11.8 U·mg^{-1})	Tri- and tetrasaccharides (lactose)	
Bifid adolescentis.	*ATCC 15703/ DSM 20083/ NCTC 11814 /E194a*		*p*NPGal (K_M 2.2 mM; V_{max} 5 U mL^{-1}) β-D-Galp(1,4)- β-D-Lactose (K_M 2.2 mM; V_{max} 93 U·mL^{-1}) β-D-Galp(1,4)- β-DGal (K_M 3.7 mM; V_{max} 95.3 U·mL^{-1})	From di- up to octamers	276,277

	ANB-7	*o*NPGal (V_{max} 23.81 U·mL^{-1}) Lactose (V_{max} 133.01U·mL^{-1})	43% Gos (30%, wt/wt, lactose)	278
Bifidobacterium bifidum NCIMB 41171			20% (w/w) GOS: 25% di-, 35% tri. 25% tetra- and 15% pentasaccharides (50%, wt/wt, lactose)	224,226, 227,279, 280
Bifidobacterium longum subsp. infantis			48% oligosaccharides whole cell Differents GOS with TLC	222,278, 281,282
Bifido. longum sbsp. longum	*BRC 15708*		33% GOS (40% lactose)	283
	JCM 7052	*p*NPGal (K_M 0.42 mM; V_{max} 5.21 µmol·min^{-1}·mg prot) *o*NPGal (K_M 0.78 mM; V_{max} 2.16 µmol·min^{-1}·mg prot)	(*p*NPGal as donor; Gal or Mel as acceptor)	284
Caldicellulosiruptor saccharolyticus (stran ATCC 43494/ DSM 8903/ Tp8T 6331)		*p*NPGal (K_M 0.11 mM; k_{cat} 32 s^{-1}) *o*NPGal (K_M 1.21 mM; k_{cat} 149 s^{-1}) Galactobiose (K_M 30.0 mM; k_{cat} 42 s^{-1})	0.171 g/L lactulose (1 mM *p*NPGal, 50 mM Gal) 408 g/L lactulose , yield 74% (700 g/L lactose)	285,286 287
Geobacillus kaustophilus (Geobacillus starothermophilus)		Lactose (K_M 114 mM; V_{max} 0.05 U/mg) Lactulose (K_M 160 mM; V_{max} 0.03 U/mg)	82 mM (23% wt/wt) Gal-β-1,3-Lac; 30% GOS (500 mM lactose)	288,289
Klebisella pneumoniae			23.8 g/L of 1-lactulose (400 g/L lactose and 200 g/L fructose)	232–234
Thermus neapilitana			17%-39% GOS (cellobiose and lactose as donors. Different acceptors)	

Table 2. 2. *Summary of β-galactosidase found in CAZY database with reported transglycosidase activity. Name of the microorganism origin, GH family from CAZy, hydrolase and transglycosidase detailes reported and references*

2.2.2 Further selection: phylogenetic trees

Phylogenetic trees analysis (Box 2. 2) showed to be a good criterium to further select among the pre-selected forty-five enzymes. This tool would allow discarding enzymes with high percentage of identity and select those with better characteristics reported.

Box 2. 2 *Phylogenetic trees*

A phylogenetic tree is a branching diagram or "tree" that depicts the lines of evolutionary descents of different species, organisms, or genes from a common ancestor. These trees are useful for organizing knowledge of biological diversity to structure classifications.

Five enzymes were found in **family GH1** labelled as β-galactosidases and with reported transgalactosidase activity (Figure 2. 1). Glucosidases were more common across this family, and only five galactosidases were described as able to synthesise any kind of galacto-derivate. *Sulfolobus solfataricus* enzyme was chosen as a final candidate because it presented higher yields and, in addition, the enzyme preferred β-1,4 linkages.

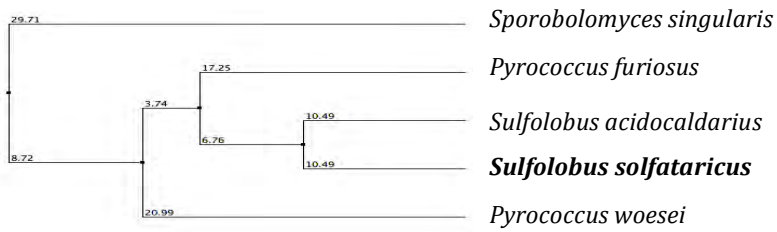


Figure 2. 1. *Phylogenetic tree of GH1 β-galactosidase candidates constructed with catalytic domains of the proteins. Alignment was made by T-coffee algorithm and tree was build according % of identity using Jalview software.*

Family GH2 had nineteen enzymes with described transglycosidase activity. Its phylogenetic tree was built only with the catalytic domain of the enzymes (being the 3rd domain in all proteins (Box 2. 3) (Figure 2. 2). Enzymes were grouped by genus (e.g. [redacted] *Arthrobacter* or *Lactobacillus).* In addition, four subgroups could be built according to their percentage of identity. In order to reduce the final number of candidates, it was decided that only one enzyme per group would be selected. Those enzymes with desirable characteristics (higher yields, β-1,4 linkage) within a group were selected as new candidates.

Box 2. 3 *Domain architecture of family GH2*

There are diverse domain architectures in GH2 based on the presence and distribution of protein domains. All of them share a central TIM barrel (catalytic domain) with two β-sandwich domains at the N-terminal end, but differ in the nature of addinional non-catalytoc modules at the C-terminal [166].

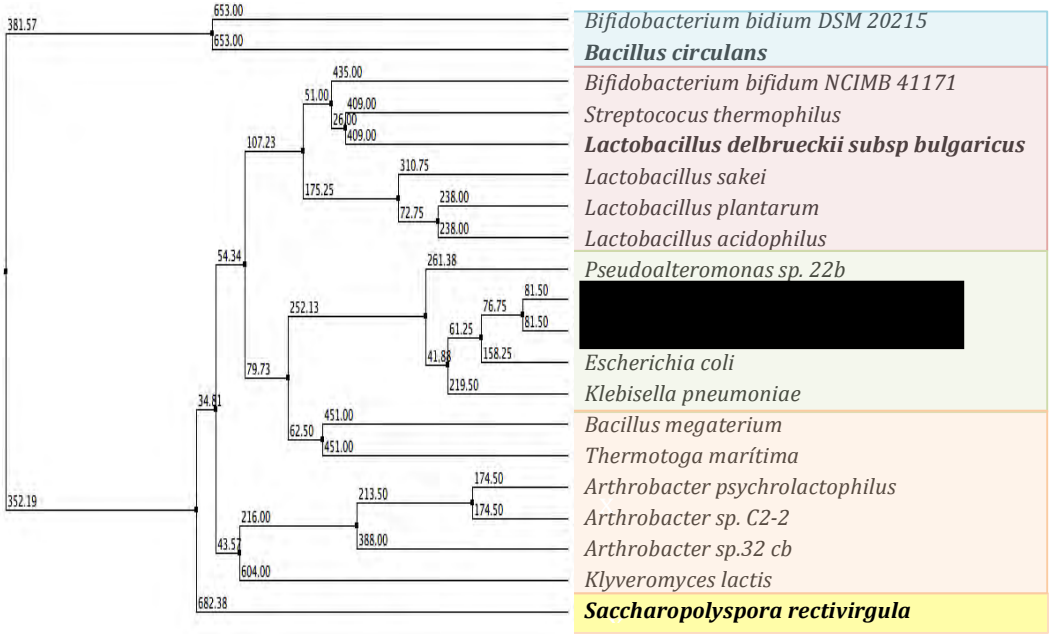


Figure 2. 2. *Phylogenetic tree of GH2 β-galactosidase candidates constructed with catalytic domains of the proteins. Alignment was made by T-coffee algorithm and tree was build according Blosum2 matrix using Jalview software*

It is interesting to highlight that one enzyme was not grouped in any branch, i.e. *Saccharopolyspora rectivirgula*. Even though this enzyme presents 25% of sequence identity with *E. coli* β-galactosidase, it has some characteristics that make it quite different from enzymes from family GH2 (for further explanation see section *2.2.3*). No candidate was selected from the *Arthrobacter* group because of low yields of transglycosylation and scant information of the linkages' nature. To sum up, β-galactosidases from *Bacillus circulans*, [redacted] *Lactobacillus delbrueckii* and *Saccharipolyspora rectivirgula* were the selected enzymes from family GH2 because they were described to have the more interesting characteristics (high transglycosidase activity and/or β-1,4 preference)

Galactosidases from **family GH35** are well known to be active on galacto-derivates but not on lactose. Half of the selected enzymes from this family belong to the fungi kingdom. They are used in whole cell processes, which would not be the first option for the gaxilose production, although they would be taken into account if yields were very high. *Paenibacillus thiaminolyticus* was selected as the best option given its high yields and different linkages. The other two enzymes of this family were discarded because *Bacillus circulans* galactosidase was described to form β-1,3 linkages and *Carnobacterium maltaromaticum* had little information of the product formed (Figure 2. 3).

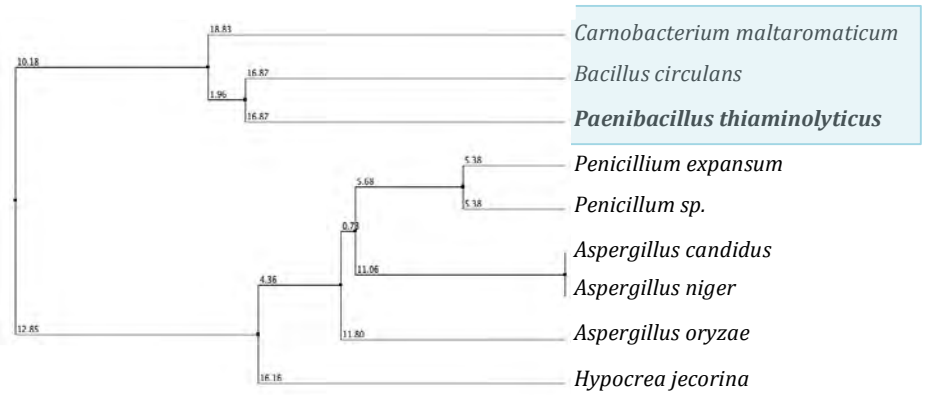


Figure 2. 3. *Phylogenetic tree of GH35 β-galactosidase candidates. Alignment was made by T-Coffee algorithm and tree was build according % of identity using Jalview software*

Twelve enzymes were found as β-galactosidases from **family GH42** with described transglycosidase activity. The phylogenetic tree organized those enzymes in two subgroups: one formed, mainly, by *Bacillus* genus and another by *Bifidobacterium* genus (Figure 2. 4).

Among the enzymes of the *Bacillus* subgroup (blue, Figure 2. 4), *Geobacillus kaustophilus* was selected. In the *Bifidobacterium* subgroup, the enzyme from *Bifidobacterium longum subsp. infantis* was chosen. Both of them presented better yields or preferable reaction conditions. Additionally, one enzyme that was in an isolated branch of the phylogenetic tree, *Thermotoga neapolitana*, was selected as candidate. The idea behind its selection was to include an enzyme with low sequence similarity to test if it could provide better characteristics.

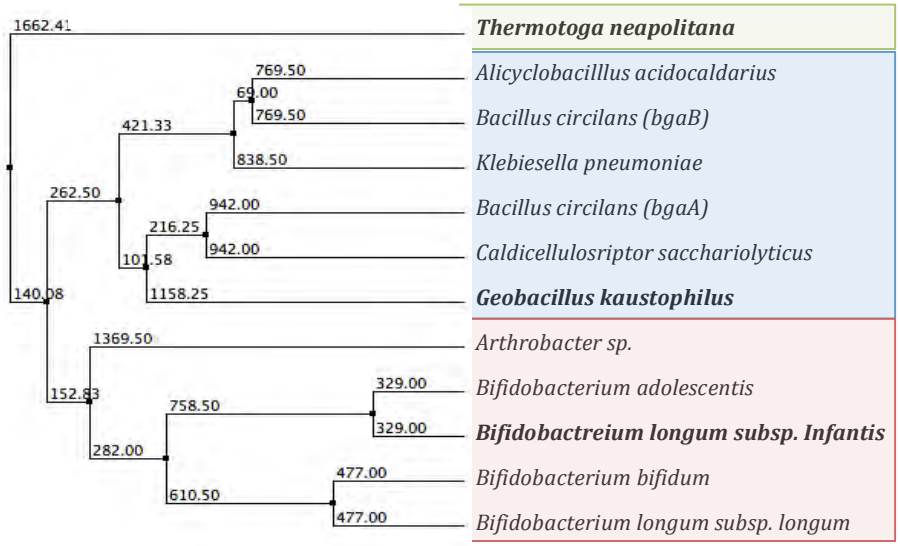


Figure 2. 4. *Phylogenetic tree of β-galactosidase candidates from family GH42. Alignment was made by T-coffee algorithm and tree was build according Blosum62 matrix using Jalview software*

Clear differences exist among GH families that contain galactosidases: one of the most distinctive one is the length of their sequences. GH1 family members are approximately 500 amino acids long, GH2 between 700-1,000 aa, GH35 above 400 aa and GH42 around 700 aa. Different lengths result in different domain organization and diversification of their structural organization. There are few available structures of GH families members on the Protein Data Bank (PDB, www.rcsb.org [185,186]). Only twenty-two β-galactosidases from different species have been solved (six from GH2, eight from GH35 and eight from GH42). Due to the low number of structures, this relevant information could not be used as criteria to select candidates.

2.2.3 Final selection: nine candidates

From the above considerations, nine enzymes were finally selected for experimental evaluation (Table 2. 3): one from family GH1, four from GH2, one from GH35 and three from GH42. (Gene and amino acid sequences are in *Appendi A.2*).

Sulfolobus solfataricus	GH1
Bacillus circulans [redacted] ***Lactobacillus delbrueckii*** ***Saccharopolyspora rectivirgula***	GH2
Paenibacillus thiaminolyticus	GH35
Bifidobacterium longum subsp. infantis ***Thermotoga neapolitana*** ***Geobacillus kastophilus***	GH42
E. coli	GH2

Table 2. 3. *Origin of the nine β-galactosidases selected as final candidates*

SULFOLOBUS SOLFATARICUS β-GALACTOSIDASE (GH1)

S. solfataricus thermostable glycosyl hydrolase from family GH1 is formed by two domains and 489 amino acids. Each monomer of 56 kDa is assembled in a tetramer (Figure 2. 5*)* (Uniprot: G8GCT7; PDB: 1uwi).

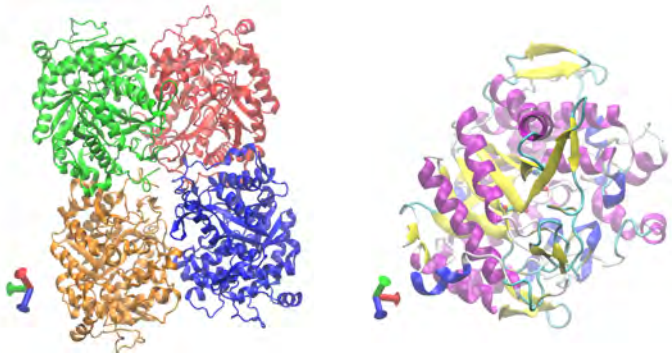


Figure 2. 5. *Homotetramer (left) and monomer (right) of β-galactosidase from* Sulfolobus solfataricus *(1uwi from PDB). Viewed by VMD: NewCartoon representation*

The enzyme has been produced in different expression systems (*Saccharomyces cerevisae* [202], *E. coli* [153,154,203,290], *Pichia pastoris* [204], *Lactococcus lactis* [205]) and it is highly active with different substrates (*p*NPGal, *p*NPFuc, *p*NPGlu) [201].

S. solfataricus β-galactosidase was able to synthesise 50 g/l lactulose (using 40% (w/v) lactose as a donor and 20% (w/v) fructose as a acceptor) [287]. It has also been used for GOS production due its ability to resist high temperatures (needed for lactose solubility and increase reaction velocity [154]).

Other studies had obtained up to 53% of GOS with wild type enzyme [154] and 61% with F441Y mutant [290,204] using lactose as a substrate. It has been analysed that the enzyme shows a marked preference to generate β-1,3 and β-1,6 linkages [153]. Even though these are not the desired linkages, it was hypothesized that a β-1,4 glycosidic

linkage might be formed with a five-carbon pentose acceptor (xylose) as opposed to β-1,6 obtained with hexoses.

LACTOBACILLUS DELBRUECKII SUBSP. BULGARICUS β-GALACTOSIDASE (GH2)

The β-galactosidase from *L. delbrueckii* (Uniprot: P0C1Y0) belongs to family GH2 and it is encoded by the *lacZ* gene. This enzyme is composed by five domains (1001 aa) and its quaternary structure is a homodimer. This is a differential feature compared to most of galactosidases of GH2 found in *Lactobacilli,* which are heterodimeric proteins encoded by two overlapping genes, *lacL* and *lacM*.

L. delbrueckii β-galactosidase was overexpressed in *L. plantarum* [240,292,293]. The transgalactosidase activity was studied and galacto-oligosaccharides reached a maximum of 102 g/L (50%) when 90% of initial lactose had reacted (consumed) (600 mM initial lactose). The main products obtained were di- and trisaccharides with a β-1,6 linkage and a minimum amount of β-1,3.

The structures of the synthesised oligosaccharides are similar to those obtained by other galactosidases from the *Lactobacillus* genus: β-1,6 linkage is the most common bond. Although this is not the desired bond, it has been described that enzymes that prefer β-1,6 linkage are able to synthesise β-1,4 when the 6-hydroxymethyl group is

absent (e.g. *E. coli* can generate β-1,6 bonds with hexose acceptors; however, β-1,4 is the most abundant linkage with xylose as acceptor, with 80% regioselectivity).

SACCHAROPOLYSPORA RECTIVIRGULA β-GALACTOSIDASE (GH2)

This GH2 family β-galactosidase belongs to a thermophilic actinomycete isolated from hay (Uniprot: Q9ZBD1). The enzyme is unique among family GH2 in different ways. It is a monomeric enzyme of 145 kDa and their 1247 amino acids are organized in 6 domains.

It is a multimetal enzyme with eight specific binding sites for divalent metal ions. These ions play important roles in maintaining the native structure essential for activity and thermostability [249,250]. It has an unusual structural feature in this family: it contains a stretch of 192 amino acid residues with no similarity to known proteins. These residues resemble an insertion into its active-site domain (detected with homology study with the *E. coli* enzyme). Deletion of this sequence lets the enzyme catalytically inactive and induces protein dimerization that retains multimetal binding characteristics, pointing at this sequence as an important element for maintaining structure and activity [247,248].

This galactosidase was able to synthesise oligosaccharides up to four monomers ($(Gal)_n$-Glc (n=1,2 3 or 4)) using lactose as substrate (1,75 M) with a maximum yield of 41% (w/w) [246].

BACILLUS CIRCULANS β-GALACTOSIDASE (GH2)

B. circulans has β-galactosidases beloging to three families: GH2 (E5RWQ2, *bgaD* [215–220]), GH35 (O31341, *bgaC* [264]) and GH42 (Q45092, *bgaA* and Q45093, *bgaB* [274,275,294,295]). Those enzymes present different lengths, sequences, physical properties and enzymatic characteristics. The enzyme selected belongs to family GH2 and it is found in Biolacta N3 preparation [217].

It has a multiple domain architecture. The amino terminus contains a signal peptide followed by three domains that are common in GH2, the third one the catalytic domain; then, a bacterial IG-like domain of group 1 and four Ig-like domains of group 4. Finally, a F5/8 type domain (Discoidin domain) located at the carboxyl terminus (Figure 2. 7)

Figure 2. 7. *Domain organization of* B. circulans.*β-galactosidase from GH2*

The DS domain is present in various eukaryotic and prokaryotic proteins. Various functions of this domain have been determined. It has been reported that the presence of the DS domain have been reported, for example, the enhancement of the hydrolytic activity of various glycosidases due to high interaction between the substrate and enzyme molecules [216,296].

Discoidin (DS) domain has been deeply studied in *B. circulans* enzyme, as well as Ig-like domains from group 4. It was determined that the DS domain played an important role in repressing the production of GOS. In particular, the production yield of tri- and tetrasaccharides was increased by deletion of the DS domain (from 7% to 30%). It was suggested that one possible reason for GOS yield increase was a more open fragile structure of the deletion mutant. Moreover, some point mutations inside DS domain were analysed and GOSs yields were also increased, probably as consequence of conformational changes [216] (other examples in bibliography were found where C-terminal deletion increasing transglycosylation activity (*Bifidobacterium bifidum* DSM20215 [297]). In this work, the enzyme was designed without the peptide signal and without the DS domain.

Several studies have analysed residues in the catalytic centre [217], thermostability [218] and modification of product specificity using enzyme engineering [220]; therefore a lot of information to improve activity is available. Another reason to select this enzyme was that it catalysed, mainly, the formation of β-1,4 linkages.

PAENIBACILLIS THIAMINOLYTICUS β-GALACTOSIDASE (GH35)

This monomeric enzyme belongs to family GH35 and it is formed by 583 amino acids (66 kDa). It possesses not only the β-galactosidase activity but also β-fucosidase activity; it is active with different substrates such as *p*NP-β-Fuc, *p*NP-β-Gal and *p*NP-β-Glc. Transfucosylase activity (65ºC, pH 5.5) was detected using β-fucosylated compounds (*p*NP-β-glycopyranosides and alcohols as acceptors). Yields were up to 69% (*p*NP-β-Man as acceptor) or 94% (methanol as acceptor) and β-1,3 and β-1,6 were the main linkages detected [268].

BIFIDOBACTERIUM LONGUM SUBSP. INFANTIS β-GALACTOSIDASE (GH42)

Bifidobacterium genus possesses high amount of galactosidases, mainly from families GH2 and GH42. Many studies have found and characterised galactosidases from this bacterial genus: *B. bifidium DSM 20215* [223], *B. bifidum NCIMB4N71* with β-gal in GH2 and GH42 [224,226,227], *B. infantis HL96* also with galactosidases in GH2 and GH42

[298–300]. Other galactosidases analysed from the same selected strain, *B. longum subsp. infantis DSM20088*, with β-gals from families GH2 and GH42 (*BIF4*-Q9F4C8 which share up to 64% of identity with selected enzyme) and additionally transgalactosidase activity was detected. Even using the same strain whole cell fermentation was used for GOS production [278].

This bacterium has three galactosidases in GH42 (Bga42A, Bga42B and Bga42C). While Bga42A hydrolysed β-1,3 bonds, Bga42B preferred β-1,4 and Bga42C acted on β-1,4 with lower efficiency [281,282]. Enzyme Bga42B was selected due to its preference for β-1,4 linkages. Different percentage of transglycosylation was achieved by GH42 galactosidases of this strain (48% oligosaccharides (w/w) [278] or detected TLC analysis).

THERMOTOGA NEAPOLITANA β-GALACTOSIDASE (GH42)

The enzyme selected belongs to family GH42 and is composed by 648 amino acids organized in three domains. The enzyme is able to produce 20-30% of GOS using lactose (500 g/L) or cellobiose as donors and different acceptors such as galactose, fructose or sucrose.

Enzymes belonging to thermophilic microorganisms can be heterologously expressed in a mesophilic organism and purified from the rest of cellular enzyme by heat inactivation and denaturation of the host's proteins. This inexpensive, simple and quick purification procedure diminishes the price of industrial enzymes, which is a major contributor to the cost of goods of biocatalytic processes.

GEOBACILLUS KAUSTOPHILUS β-GALACTOSIDASE (GH42)

This enzyme is a galactosidase encoded by the *bgaB* gene that belongs to family GH42 (Uniprot: P19668). It is a thermostable enzyme formed by 672 aa (78 kDa) organized in three domains, the first being the catalytic domain, and forming a homotrimer.

This enzyme is able to produce oligosaccharides, mainly the trisaccharide 3'-galactosyl-lactose, from lactose. Wild type enzyme was capable of producing up to 2% (weight/weight) of that trisaccaharide. New mutants with increased yield, 21% (R109V) or 23% (R109W), have been described [288,289].

ESCHERICHIA COLI β-GALACTOSIDASE (GH2)

β-galactosidase from *E.coli* has been extensively described previously in *CHAPTER 1*. This is the current enzyme that is being used for industrial gaxilose production.

Therefore, this enzyme is included in all enzyme's evaluation as a standard: it would be considered the threshold for the new enzyme selection.

2.3 Evaluation

All the selected enzymes have been described to synthesise some kind of oligo-, polysaccharide or galacto-derivatives. Capacity to synthesise gaxilose needs to be evaluated in the transglycosylation reaction with specific substrates (xylose and *o*NPG).

2.3.1 Effect of pH and temperature

The hydrolase activity of the selected enzymes has been reported under different reaction conditions. Whereas some have been evaluated at different pH values, others have only been reported at a single pH value (Table 2. 4).

Single pH value			**Several pH values**		
	pH	**T**		**pH**	**T**
S. solfataricus (GH1)	6.0	85ºC	***B. circulans***	6.0-7.0	40ºC
P. thiaminolyticus (GH35)	5.5	65ºC	[redacted]	6.5-10.2	35ºC
B. longum (GH42)	6.0	40ºC	***L. delbrueckii***	6.0-9.0	50ºC
G. kaustophilus (GH42)	6.5	37ºC	***S. rectivirgula***	6.5-7.2	high
T. neapolitana (GH42)			***E. coli***	7.0	37ºC

Table 2. 4. *Origin of candidate enzymes studied at a single pH value and enzymes studies at several pH values. It also includes pH and temperature used to evaluate the enzymes*

In order to compare the enzymes and check their behaviour under the same conditions (used for the reference the *E. coli* β-galactosidase), their pH profiles for the glycosidase acivity (*o*NPG) were studied (**¡Error! No se encuentra el origen de la referencia.**).

Only the enzyme from *T. neapolitana* had its maximum activity at pH 7.0. The other enzymes (*B. longum*, *G. kaustophilus*, *P. thiaminolyticus* and *S. solfataricus*) shifted its optimum to 6.3 – 6.6, closer to the optimum activity of *E. coli* enzyme. All enzymes maintained more than 75% of their activity at pH 7.0. For this reason all enzymes were evaluated at pH 7.0 in future reactions.

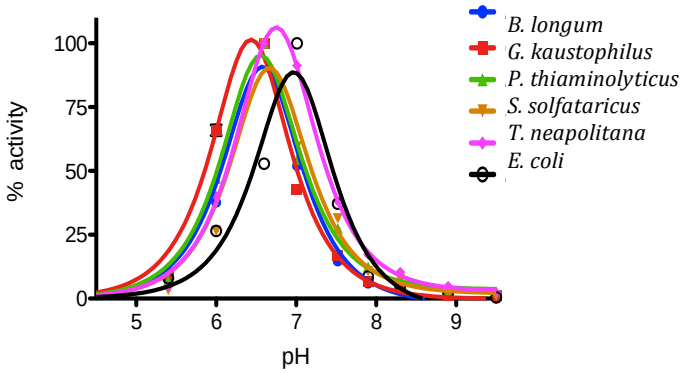


Figure 2. 8. *pH profile of glycosidase activity of some galactosidases candidates. CONDITIONS: 13 mM* oNPG, *0.01-0.03 U of enzyme, 37ºC, pH 7.0, buffer:100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$ *(release of* oNP*)*

Temperature could not be modified largely due to spontaneous *o*NPG hydrolysis. Therefore, evaluation of optimum temperature was not performed.

2.3.2 Glycosidase activity

Glycosidase activity was evaluated under standard conditions (13 mM *o*NPG, pH 7, 37ºC), as defined in *CHAPTER 1*. All enzymes presented activity with *o*NPG, as expected, with different glycosidase specific activities (Table 2. 5)

The enzyme with the highest specific activity using *o*NPG as substrate was galactosidase from *E. coli* followed by *L. delbrueckii*, [redacted] and *B. circulans*: all of them belong to family GH2, described as the most active group on simple saccharides like lactose.

On the other side, enzymes from *G. kaustophilus*, *T. neapolitana* and *S. rectivitugula* presented two orders of magnitud lower activity. *P. thiaminolyticus* and *B. longum* from families GH35 and GH42 show intermediate activity.

		Glycosidase specific activity	
		(s^{-1}) (umol oNP·s^{-1}·umol protein)	U/mg (umol oNP·min^{-1}· mg $prot^{-1}$)
S. solfataricus	GH1	10.90 ± 0.52	11.5 ± 0.5
B. circulans	GH2	268.82 ± 9.56	141.0 ± 5.1
[redacted]	GH2	305.50 ± 29.96	156.0 ± 14.9
L. delbrueckii	GH2	989.32 ± 81.83	513.1 ± 42.4
S. rectivirgula	GH2	20.83 ± 2.44	8.8 ± 1.0
P. thiaminolyticus	GH35	57.70 ± 2.70	52.4 ± 2.1
B. longum subsp. infantis	GH42	67.50 ± 7.83	50.3 ± 5.9
T neapolitana	GH42	8.48 ± 1.14	6.8 ± 0.9
G. kaustophilus	GH42	7.66 ± 1.21	5.9 ± 0.9
E. coli	GH2	1,455.52 ± 103.60	753.7 ± 52.8

Table 2. 5. *Glycosidase specific activity (s^{-1}) and U/mg of candidates (hydrolysis of* oNPG *13 mM, 0.01-0.1 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$*).). Determined following absorbance of* oNP *release. U, Units are defined as µmol substrate catalysed per minute*

2.3.3 Transglycosylation reaction

The selected enzymes had reported transglycosidase activity in different conditions (T, pH, substrate concentration, amount of enzyme). However, enzyme evaluation should be performed in standardized conditions to compare all the enzymes under the same reaction conditions and with the same product quantification method (HPLC-MS).

Transglycosylation need to be studied and optimized under the industrial conditions reported for the *E. coli* β-galactosidase. However, under these conditions proper enzymatic characterisation (i.e. determination of catalytic constants) is difficult due to insolubility of substrate *o*NPG. (In industrial settings, it is frequent to add insoluble substrates at once for ease of operation. The substrate dissolves as the reaction proceeds.) In order to calculate catalytic constants, other conditions were set up, referred as *Laboratory conditions*.

Regarding the amount of enzyme used in the transglycosylation reactions, the same number of *glycosidase activity units* were added to each reaction in order to compare the results of the different candidates (Units determined by glycosidase activity quantifying the amount of oNP generated in the reaction (*1.3.1*).

2.3.3.1 First selection: Laboratory conditions

All nine selected enzymes were analysed for their transglycosidase activity in laboratory conditions, with *o*NPG (50 mM) as a donor substrate, xylose (250 mM) as acceptor substrate, at pH 7.0 and 37ºC. Substrates and products concentrations were analysed by HPLC-MS as well as by TLC. TLC analysis allowed to identify the presence or absence of *o*NPG (limiting reagent) during the reactions and, therefore, evaluate substrate conversion. Different amounts of *o*NPG could be detected at 20 hours depending on the amount of enzyme used and its activity (Figure 2. 9).

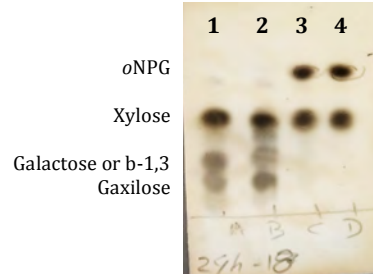


Figure 2. 9. *TLC analysis of transglycosylase reaction, in laboratory conditions, at 20 h of reactions with different amount of enzyme* [redacted] *1.:15 U; 2: 5 U; 3: 0.5 U; 4: 0,015 U. Mobile phase: 7.5 ipa, 2.5 water, 0.5* [redacted]*onium. Reveletor* H_2SO_4*:MeOH:*H_2O

All enzymes presented transglycosidase activity with *o*NPG and xylose as substrates. However, as expected, with different initial rates (Figure 2.10). Some enzymes

hydrolysed and transglycosylased at the same level and others with different order of magnitude (e.g. A- *B. longum*, B- *G. kaustophilus*, E- *T. neapolitana* (Figure 2.10)*:* those enzymes were analysed with two ordinate axes).

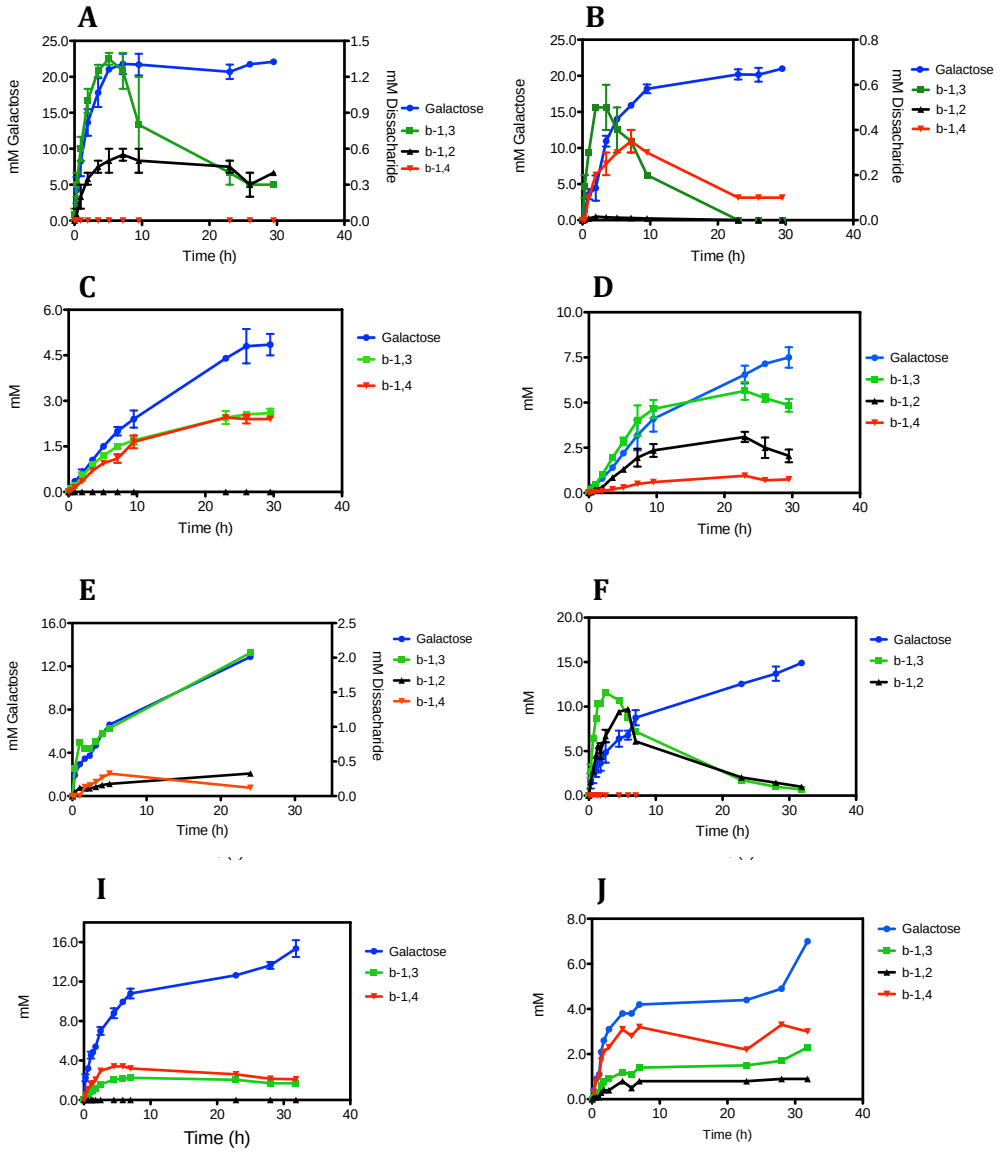


Figure 2.10. *Product concentration (mM)* versus *time of the nine candidate enzymes in transglycosylation reactions under* Laboratory conditions*. A)* B[redacted]. kaustophilus; *C)* P. thiamiloyticus; *D)* S. solfataricus*; E)* T. neapolitana*; F)* B. circulans; [redacted]*H)* L. delbrueckii; *I)* S. rectivirgula*; J)* E. coli*. Blue-galactose, green- β-1,3 disaccharide, red β-1,4 and black β-1,2. CONDITIONS: 50 mM* oNPG*, 250 mM xylose, 0.3-5 U of enzyme, 1.25 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$*). Products quantified by HPLC-MS*

Different features were used to evaluate these reactions: regioselectivity, gaxilose formation (yield and T/H ratio) and initial rates.

2.3.3.1.1 Regioselectivity of transglycosylation

Candidates synthesise various regiosomers, although only gaxilose (4-*O*-β-D-galactopyranosyl-D-xylose) is the desired product. In order to evaluate the enzymes' transglycosidase capacity, three regiosomers were analysed in every reaction (i.e. 2-*O*-β-D-galactopyranosyl-D-xylose, 3-*O*-β-D-galactopyranosyl-D-xylose and 4-*O*-β-D-galactopyranosyl-D-xylose, abbreviated as β-1,2, β-1,3 and β-1,4, respectively). Those regioisomers are the ones synthesised by the *E. coli* β-galactosidase, the standard biocatalyst. The candidate with higher transglycosydase activity were *B. circulans*, *S. solfataricus* and [redacted] (Figure 2. 12). *W*hile the first two synthesised β-1,3 and β-1,2, [redacted] assembled β-1,4 linkage (maximum yield at Table 2.6). T/H ratio also showed that these three enzymes had higher transglycosylation capacity than the *E. coli* enzyme (*right*-Figure 2. 12*). B. circulans* galactosidase exhibited a 8-fold higher T/H ratio than the *E. coli* enzyme, while [redacted] and *S. solfataricus* presented a 1.6-fold higher ratio.

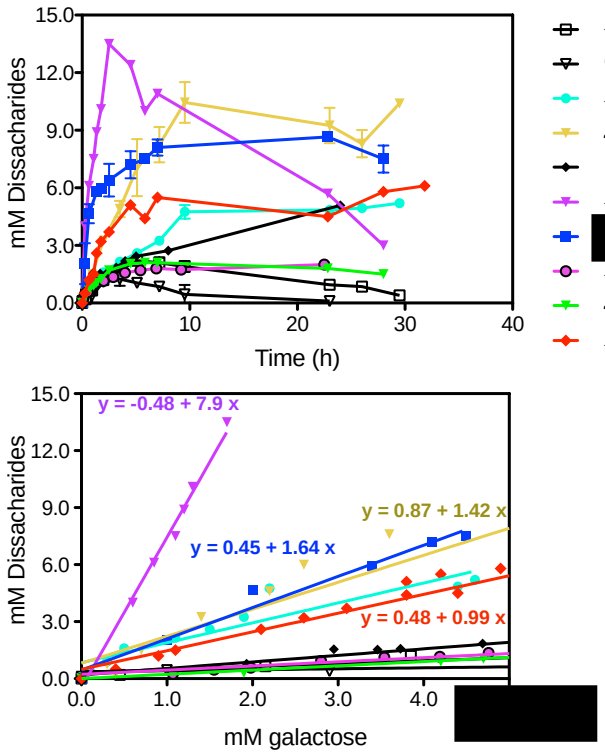


Figure 2. 11. *Transglycosidase capacity of β-galactosidase candidates.* Left-*Disaccharides (β-1,2+ β-1,3+ β-1,4) produced by different candidates on time.* Right- *Disaccharides* vs *galactose concentration: T/H ratio evaluation. . CONDITIONS: 50 mM* oNPG, *250 mM xylose, 0.3-5 U* of enzyme, *37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

	Maximum disaccaride
B. circulans	13.5 mM
[redacted]	8.6 mM
S. solfataricus	10.4 mM

Table 2. 6. *Maximum disaccharide achieved by β-galactosidase candidates under laboratory conditions (data from Figure 2.11)*

During the enzymatic reactions other products were simultaneously synthesised and could be detected, abeit not quantified, using HPLC-MS in Scan-mode (see *CHAPTER*

1). It was determined that all enzymes synthesised some amount of Gal-Gal-*o*NP and Gal-Gal disaccharide. Even some of them synthesise longer oligosaccharides (*Appendix, A.3).*

2.3.3.1.2 Gaxilose synthesis

Some enzymes did not synthesise gaxilose, or minimal amounts, such as β-galactosidases from *B. longum* (A), *S. solfataricus* (D) or *B. circulans* (F) (Figure 2.10). *B. longum* had been described to have β-1,4 preference, but using xylose as acceptor, the β-1,3 was the preferred isomer; same behaviour occurred with *B. circulans* enzyme and *S. solfataricus* (D): both enzymes achieved higher yields for β-1,3.

Other enzymes such as galactosiases from [redacted] *L. delbrueckii* (H) or *S. rectivirgula* (I) synthesised gaxilose as the main disaccharide product (Figure 2. 11).

It should be pointed out that some enzymes produced more hydrolysis product, galactose, than any kind of disaccharide such as *B. longum* (A), *G. kaustophilus* (B), *T. neapolitana* (E), *L. delbreuckii* (H) or *S. rectivirgula* (H). This indicates that those candidates are predominantly hydrolytic enzymes (their natural activity) and that they would not present high T/H ratio.

Reaction catalysed by the *T. neapolitana* enzyme (E, Figure 2.10) did not reach yet the maximum yield of disaccharide products formation in the monitored time scale; probably because the enzyme is from a thermophile and is more active at higher temperatures. For this reason the reaction was set up at higher enzyme concentration, but higher product concentration was not detected (not shown). Reaction was not performed at higher temperatures, although enzymatic activity would probably increase, because *o*NPG would hydrolyse spontaneously and the product stability may be at stake.

From all the analysed candidates, one enzyme stood out: β-galactosidase from [redacted] synthesised three times more gaxilose than *E. coli* enzyme (blue in *left*- Figure 2. 12). Other two enzymes presented the same synthesis yield as *E. coli* enzyme: *L. delbrueckii* and *S. rectivirgula*. The other six enzymes did not produce significant amounts of gaxilose.

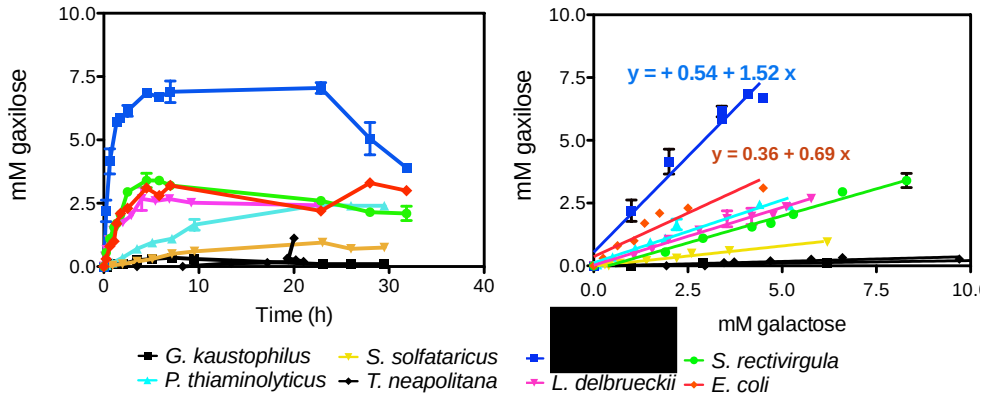


Figure 2. 12. *Gaxilose synthesis capacity of β-galactosidase candidates.* Left-*amount of gaxilose* vs. *time of all the enzymes.* Right-*amount of gaxilose* vs. *amount of galactose (ratio T/H). Under laboratory conditions CONDITIONS: 50 mM* oNPG, *250 mM xylose, 0.3-5 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

Plotting gaxilose concentration *versus* galactose concentration (transglycosylation product *vs.* hydrolysis product, respectively) transglycosylation to hydrolysis (T/H) ratio could be calculated (*right*-Figure 2. 12): higher slope means higher T/H ratio and consequently more gaxilose formation. [redacted] enzyme showed the highest T/H ratio followed by *E. coli* galactosidase. This is a clear advantage of the [redacted] enzyme.

It is interesting to highlight that, for the enzyme from *P. thiaminolyticus* which presented a high T/H ratio (clear blue, *right*-Figure 2. 12), gaxilose production was low. This is due to the fact that this enzyme hydrolysed the substrate with a very low rate (very low amount of galactose release). On the other hand, it synthesised different kinds and yields of other oligosaccharides (not only β-1,4 disaccharide): in addition to the three regioisomers for galactopyranosyl-D-xyloses, other compounds were detected by HPLC-MS scan or TLC (13- Figure 2. 13) such as GalGal-*o*NP or GalGal. Therefore, this enzyme was not a good candidate despite its remarkable T/H ratio.

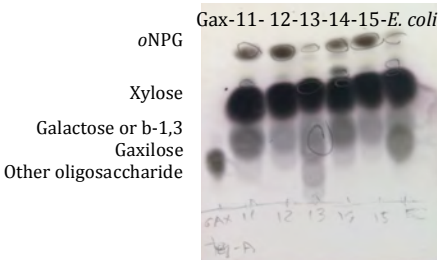


Figure 2. 13. *TLC analysis of transglycosylation reaction of different candidates: 11*-B. longum *;12*-G. kaustophilus *;13*-P. thaminolyticus *;14*-S. solfataricus *;15*-T. neapolitana *and* E. coli *at laboratory conditions. CONDITIONS: 50 mM* oNPG, *250 mM xylose, 1.25 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Mobile phase: 7.5 isopropanol, 2.5 water, 0.5 amonium. Reveletor H_2SO_4:MeOH:H_2O*

2.3.3.1.3 Transglycosidase and hydrolase initial rates

Transglycosidase and hydrolase initial rates were calculated (Table 2. 7). β-galactosidases from *B. longum* (GH42) and *T. neapolitana* (GH42) exhibited the highest hydrolase initial rate, while *P. thiaminolyticus* (GH35) and *S. solfataricus* (GH1) the lowest.

	V_o (mM/min) ($V_{o,i}$ /V_t %)				MAX. mM
	Galactose	**β-1,3**	**β-1,2**	**β-1,4**	**β-1,4**
B. longum	$(6.68 \pm 0.01)\cdot10^{-2}$ (90%)	$(4.99 \pm 0.29)\cdot10^{-3}$ (7%)	$(2.08 \pm 0.06)\cdot10^{-3}$ (2%)		
G. kaustophilus	$(3.99 \pm 0.22)\cdot10^{-2}$ (92%)	$(2.94 \pm 0.01)\cdot10^{-3}$ (3%)	$(2.73 \pm 0.02)\cdot10^{-3}$ (3%)	$(1.43 \pm 0.02)\cdot10^{-3}$ (7%)	0.4
P. thiaminolyticus	$(4.90 \pm 0.04)\cdot10^{-3}$ (39%)	$(4.94 \pm 0.33)\cdot10^{-3}$ (39%)		$(2.76 \pm 0.17)\cdot10^{-3}$ (22%)	2.7
S. solfataricus	$(6.85 \pm 0.22)\cdot10^{-3}$ (30%)	$1.04 \pm 0.05)\cdot10^{-2}$ (45%)	$(4.83 \pm 0.27)\cdot10^{-3}$ (22%)	$(8.70 \pm 0.99)\cdot10^{-4}$ (4%)	1.0
T. neapolitana	$(1.58 \pm 0.90)\cdot10^{-1}$ (85%)	$(7.38 \pm 0.09)\cdot10^{-3}$ (4%)	$(1.80 \pm 0.01)\cdot10^{-2}$ (10%)	$(2.90 \pm 1,25)\cdot10^{-3}$ (2%)	1.1
B. circulans	$(3.24 \pm 0.77)\cdot10^{-2}$ (19%)	$(9.23 \pm 0.25)\cdot10^{-2}$ (54%)	$(4.64 \pm 1.00)\cdot10^{-2}$ (27%)		
[redacted]	$(2.59 \pm 0.13)\cdot10^{-2}$ (33%)	$(6.44 \pm 0.99)\cdot10^{-3}$ (8%)	$(5.04 \pm 0.75)\cdot10^{-3}$ (6%)	$(4.16 \pm 0.30)\cdot10^{-2}$ (53%)	7.2
L. delbrueckii	$(3.48 \pm 0.55)\cdot10^{2}$ (56%)	$1.06 \pm 0.11)\cdot10^{-2}$ (17%)		$(1.72 \pm 0.29)\cdot10^{-3}$ (28%)	2.7
S. rectivirgula	$(4.51 \pm 0.31)\cdot10^{-2}$ (58%)	$(1.01 \pm 0.01)\cdot10^{-2}$ (13%)		$(2.21 \pm 0.01)\cdot10^{-2}$ (29%)	3.3
E. coli	$(1.44 \pm 0.22)\cdot10^{-2}$ (30%)	$(6.24 \pm 0.56)\cdot10^{-3}$ (13%)	$(3.31 \pm 0.38)\cdot10^{-3}$ (7%)	$(2.40 \pm 0.16)\cdot10^{-2}$ (50%)	3.3

Table 2. 7. *Initial rates of all products analysed (V_o, mM/min) (galactose and the three regiosomers), below the percentatge of the initial rates determined by ($\frac{V_{o\,(i)}}{Vo_{gal}+Vo_{\beta1,4}+Vo_{\beta1,3}+Vo_{\beta1,2}}$)·100.) and the third column is the maximum gaxilose reached in transglycosylase reaction by β-galactosidases candidates under laboratory coditions. CONDITIONS: 50 mM oNPG, 250 mM xylose, 0.3-5 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

If transglycosidase initial rates were high, gaxilose yield at low substrate conversion would be higher too; e.g. enzymes from [redacted] *L. delbrueckii, S. rectivirgula* and *E. coli*.

Additionally, because of product hydrolysis (secondary hydrolysis) at long reaction times, enzymes with similar transglycosidase initial rates (red-*left*- Figure 2. 14) such as *E. coli*, [redacted] or *S. rectivirgula*, produced different final yields of gaxilose formation. The extent of secondary hydrolase activity depends on the affinity of the enzyme for the transglycosylation product. Therefore, initial rates could not be used as the sole indicator for candidate selection. Moreover, other factors e.g. product

inhibition or other product(s) formation (gaxilose can act as donor and/or acceptor) also could affect gaxilose yield at long time reactions with no effect in initial rates.

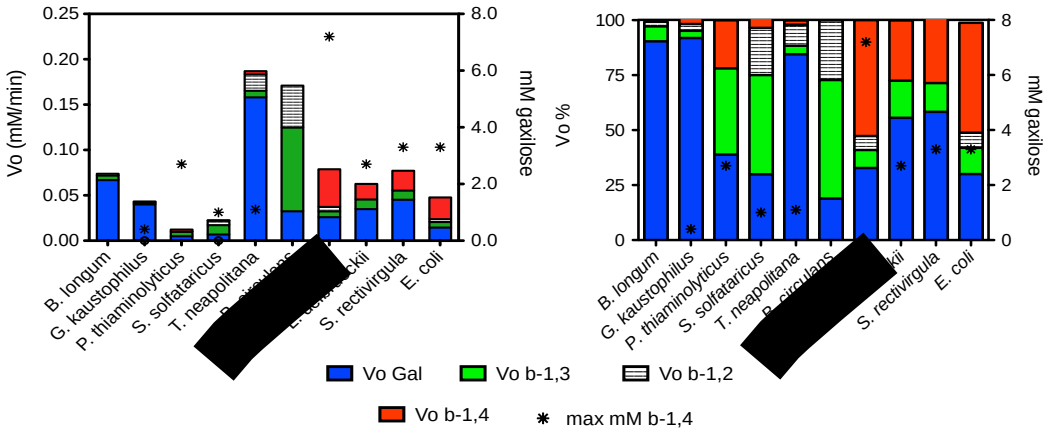


Figure 2. 14. *Comparative analysis of transglycosylation of the nine candiates by representanting initial rates (Blue-galactose, green- β-1,3 disaccharide, red β-1,4 and black β-1,2) and maxim gaxilose reached.* Left-*initial rates of galactose and the three regioisomers synthesised (left-Y) and gaxilose yields (right-Y)) obtained by each enzyme.* Right-*Percentatge of initial rates of(galactose and the three regioisomers synthesised left-Y) and gaxilose yield (eight Y)*

Taking into account the percentages of initial rates (*right*- Figure 2. 14 and Table 2. 7), *E. coli* β-galactosidase presented higher transglycosidase intial rate and [redacted] had higher hydrolase initial rate, but it was the second enzyme that achieved higher yield of gaxilose. This indicates that it could have lower secondary hydrolysis and/or no product inhibition.

Laboratory conditions used in the above experiments were chosen for kinetic characterization under initial "homogeneous conditions". However, during the course of the reaction some turbidity appeared. Reactions were monitored until turbidity appeared and reactions did not reach high reaction conversion (40-50%). In order to re-evaluate the outstanding enzymes at higher reaction conversions, new conditions were set up: *analytical conditions* (*2.3.3.2*). These conditions would have lower substrate concentration and they would allow to describe the system in homogeneous conditions along the whole reaction.

2.3.3.2 Further screening: Analytical conditions

Under these new evaluation conditions (5 mM *o*NPG-25 mM xylose), only three candidates were selected for further analysis. The [redacted] enzyme was selected because it was able to synthesise gaxilose in higher yield and, besides, presented the best T/H ratio. ***L. delbrueckii*** and ***S. rectivirgula*** enzymes reached similar yields than the galactosidase from *E. coli* despite their lower T/H ratios. Other candidates were

discarded because low amount of gaxilose was produced due to low transglycosylation activity or because of the preferential synthesis of others regioisomers (e.g. *B. circulans, S. solfataricus* or *P. thiaminolyticus.*

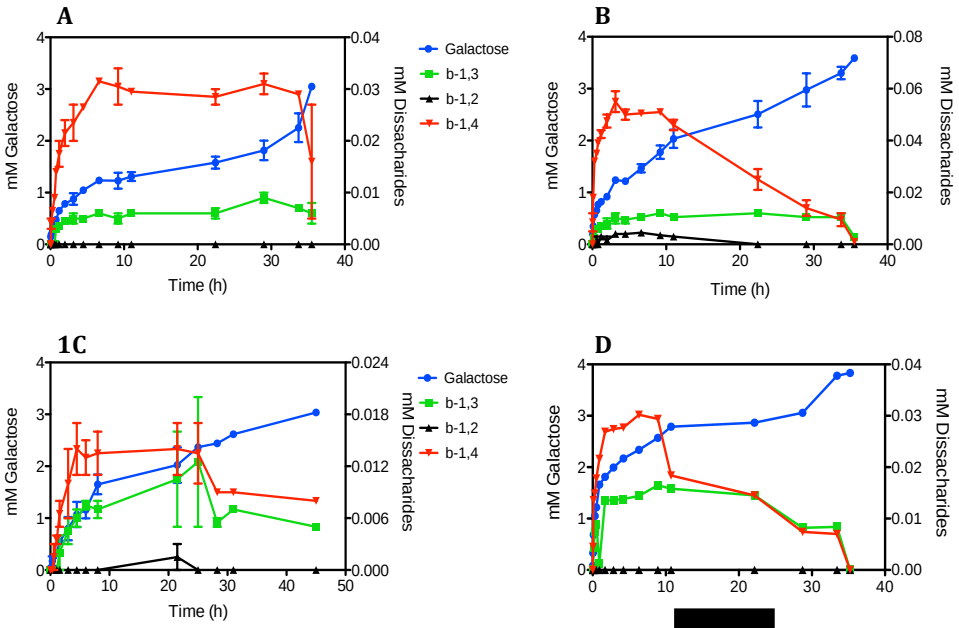


Figure 2. 15. *Time course reaction in* analytical conditions o*f: A:* E. coli *B:* [redacted] *C:* S. rectivirgula, *D:* L. delbrueckii *(this enzyme has no replicates in this reaction) Blue-galactose, green- β-1,3 disaccharide, red β-1,4 and black β-1,2. CONDITIONS: 5 mM oNPG, 25 mM xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

Time course of the reactions was evaluated for the *four* enzymes (three candidates and control enzyme from *E. coli*) during more than 24h (Figure 2. 15). All enzymes showed a maximum in gaxilose production and from this time point enzymes hydrolysed the product at different rates. In addition, once enzymes began to hydrolyse gaxilose, galactose amount increased: indicating that the decrease of gaxilose is due to hydrolysis and not because of the synthesis of other oligosaccharides (gaxilose could act as a donor and/or acceptor for further tranglycosylation reactions).

The candidate enzymes produced gaxilose with different initial rates and they achieved different yields (Table 2. 8). Lower substrate concentration (5 mM donor, 25 mM acceptor) decreased transglycosidase activity and it is reflected in the percentage of initial rates and gaxilose yield achieved. [redacted]nzyme showed the highest T/H ratio: high % of transglycosidase initial rate and low % of hydrolysis rate.

	V_o (mM/min) (V_i/V_{total}·100)			MAX. mM β-1,4
	Galactose	β-1,3	β-1,4	
E. coli	$(9.11 \pm 0.16)\cdot10^{-3}$ (97%)	$(4.88 \pm 0.31)\cdot10^{-5}$ (<1%)	$(2.83 \pm 0.16)\cdot10^{-4}$ (3%)	0.03
[redacted]	$(2.28 \pm 0.86)\cdot10^{-3}$ (92%)	$(1.76 \pm 0.07)\cdot10^{-5}$ (<1%)	$(1.81 \pm 0.03)\cdot10^{-3}$ (7%)	0.06
L. delbrueckii	$(5.66 \pm 0.36)\cdot10^{-2}$ (98%)	$(2.29 \pm 0.32)\cdot10^{-4}$ (<1%)	$(1.19 \pm 0.91)\cdot10^{-3}$ (2%)	0.03
S. rectivirgula	$(4.48 \pm 0.57)\cdot10^{-3}$ (98%)	$(3.29 \pm 0.33)\cdot10^{-5}$ (<1%)	$(7.38 \pm 1.98)\cdot10^{-5}$ (2%)	0.02

Table 2. 8. *Hydrolysis and transglycosydase initial rates of the four final candidates (β-galactosidases from* E. coli, [redacted] L. delbrueckii *and* S. rectivirgula *under analytic conditions). CONDITIONS: 5 mM* oNPG, *25 mM xy[redacted]8 U of enzyme , 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

Additionally, *S. rectivirgula* and *L. delbrueckii* enzymes presented lower regioselectivity (50-70%) (Table 2. 9); meanwhile, *E. coli* and [redacted] galactosidases maintain β-1,4 disaccharide at 80% respect all galactopyranosyl-D-xyloses synthesised, (Figure 2. 16). None of the enzymes produce significant amount of β-1,2 regioisomer.

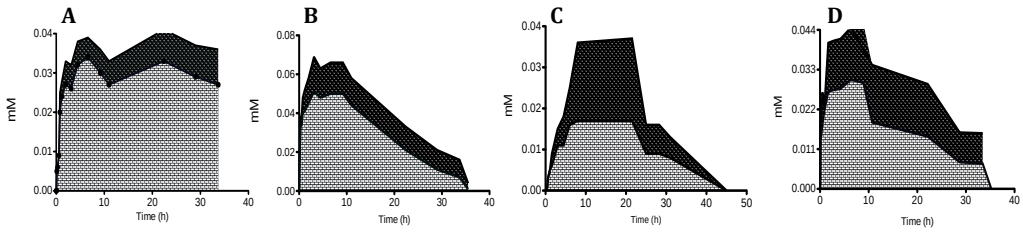


Figure 2. 16. *Production on time of disaccharides (dark grey) and β-1,4 regiosomer (light grey) of the final candidates. A:* E.coli*; B.* [redacted]*:* S. rectivirgula*; D:* L. delbrueckii. *CONDITIONS: 5 mM* oNPG, *25 mM xylose, 0.01-0.08 U of enzyme, 37º[redacted]mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

	Regioselectivity β-1,4
E. coli	80 – 87%
[redacted]	73 – 89%
L. delbrueckii	50 – 75%
S. rectivirgula	45 - 73%

Table 2. 9. *Regioselectivity for β-1,4 regioisomer by the final four β-galactosidase candidates under analytic conditions (data from Figure 2.6)*

Under these conditions, the [redacted]enzyme showed, again, a higher gaxilose yield (Table 2. 8): two-fold higher than *E. coli* and *L. delbrueckii*, (*left*-Figure 2. 17). On the other hand, *S. rectivirgula* galactosidase exhibited nearly 2.5-fold lower gaxilose yield than *E. coli*; while in *laboratory conditions* reached nearly the same yield. In addition, the same decrease was also observed in T/H ratio, (*right*-Figure 2. 17). One possible

explanation is that *S. rectivirgula*'s drop of transglycosylic activity is due to less thermodynamically favoured conditions.

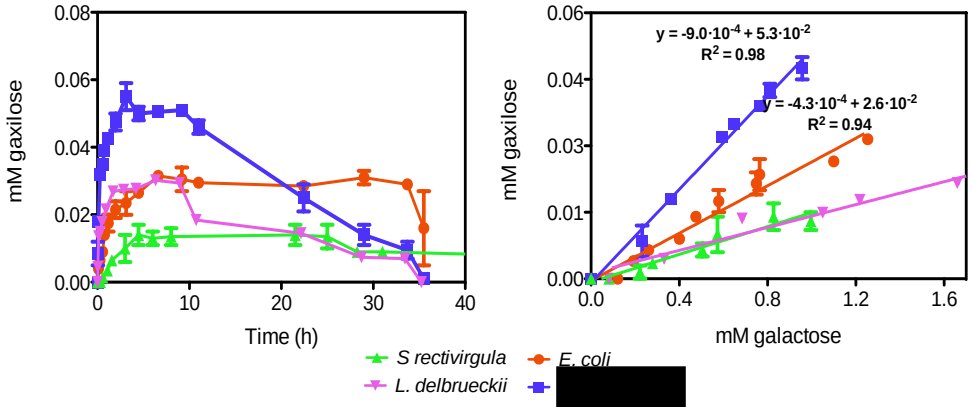


Figure 2. 17. *Analysis of gaxilose production by the four final candidates under analytical conditions.* Left- *amount of gaxilose produced on time* Right- *T/H ratio of candidates. CONDITIONS: 5 mM oNPG, 25 mM xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

2.3.3.3 Industrial conditions

Even though the enzymes were assayed in two different conditions, it was also needed to study them under industrial conditions, i.e. higher substrate concentration (0.1 M *o*NPG and 0.4 M xylose). Industrial condition reactions were not homogeneous. (samples of the heterogeneous reactions were diluted and totally dissolved before their analysis).

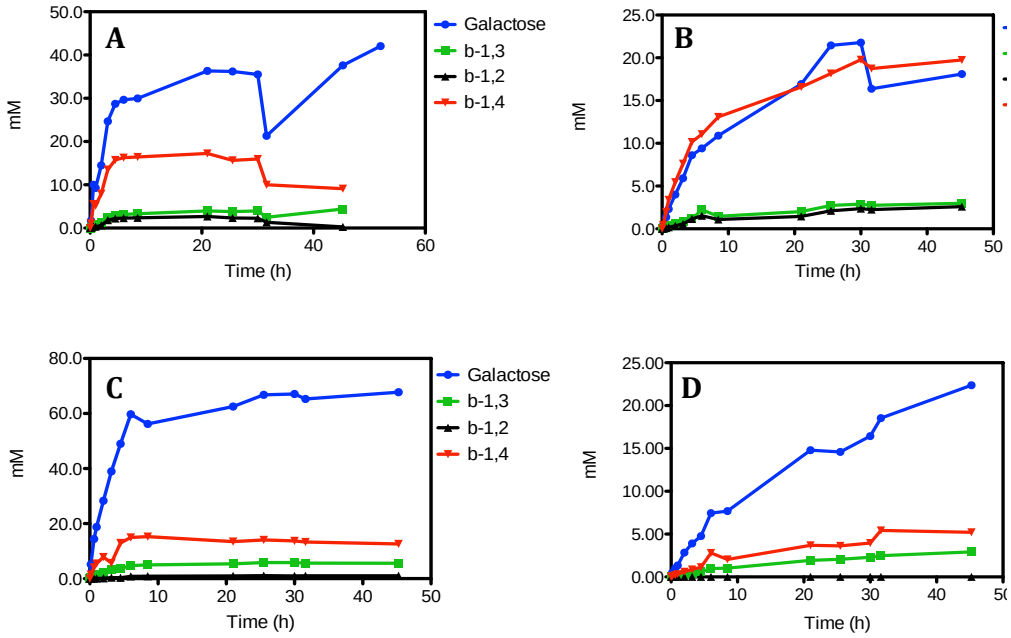


Figure 2. 18. *Products on time of the transglycosidic reaction under industrial conditions by the four final candidates A-* E.coli *;B-* [redacted] *; C-* L. delbrueckii*; D-* S. rectivirgula Colours: *Blue-galactose, green- β-1,3 disaccharide, red β-1,4* [redacted] *-1,2. CONDITIONS:* 0.1 *M* oNPG*, 0.4 M xylose, 1.25-5 U of enzyme 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

Time course of the reactions was also evaluated for more than 40 h. Different behaviours were observed: while the galactosidase from [redacted] did not hydrolyse gaxilose (it seemed it had arrived in a *plateau* zone), the other three enzymes, after achieving a maximum, began to hydrolyse gaxilose (Figure 2. 18). However, the decrease in gaxilose concentration was not as sharp as in *laboratory* or *analytical conditions*. This behaviour is maybe due to product inhibition (higher concentrations) and/or enzyme inactivation (longer reactions times).

L. delbrueckii (C) and *S. rectivirgula* (D) enzymes achieved nearly the same gaxilose (red) as galactose (blue), showing its higher hydrolytic activity.

It can be observed (*left*- Figure 2. 19) that all the enzymes rendered similar yields of gaxilose (Table 2. 10). Nevertheless, the [redacted] enzyme did not begin to hydrolyse gaxiose; probably as a result of its lower secondary hydrolysis and/or because of enzyme inactivation. Enzymes did not hydrolyse gaxilose in such a high rate as they did in *laboratory* or *analytical* conditions. Different hypothesises could explain this behaviour: longer reaction times could be inactivating the enzymes or higher substrate concentration could be inhibiting the enzymes (e.g. galactose release).

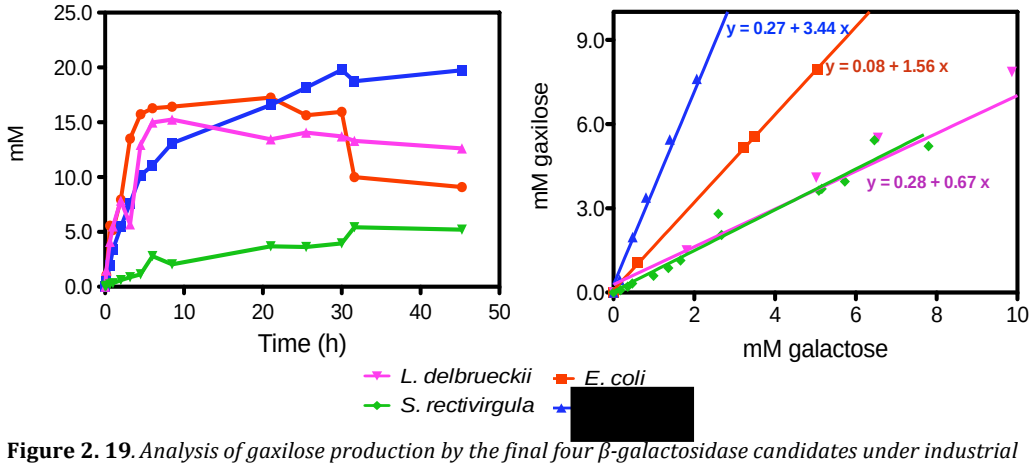


Figure 2. 19. *Analysis of gaxilose production by the final four β-galactosidase candidates under industrial conditions.* Left-*gaxilose produced on time.* Right- *gaxilose* vs. *galactose concentration (T/H ratio).* *CONDITIONS:* 0.1 M oNPG, 0.4 M xylose, 1,25-5 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). *Products quantified by HPLC-MS*

	E. coli	[redacted]	*L. delbrueckii*	*S. rectivirgula*
MAX. β-1,4	17.2 mM	19.3 mM	15.2 mM	5.4 mM
(% regioselct.)	(72%)	(80%)	(70%)	(65%)

Table 2. 10. *Maximum gaxilose achieved by β-galactosidase candidates and % of regioselectivity under industrial conditions (linked to Figure 2. 19)*

The analyses of the enzymes' T/H ratio (*right*-Figure 2. 19) showed that the highest T/H ratio was that of the [redacted] β-galactosidase: nearly two-fold higher than that of

the *E. coli* enzyme. The enzymes from *L. delbrueckii* and *S. retcivirgula* presented similar ratios.

The [redacted]β-galactosidase presented the highest T/H ratio and did not seem to hydrolyse gaxilose under those conditions. To understand this behaviour, an additional experiment was performed. A set of reactions was carried out with increasing amounts of enzyme (Figure 2. 20). If the enzyme dose was doubled (5 U), the initial rate was higher and the achieved yield increased accordingly. However, when higher enzyme concentration was used, initial rates were higher but gaxilose hydrolysis began earlier (in 10h reaction for 10 U used and in 2.5h for 20 U and 30 U). Higher enzyme concentration seemed to increase reaction yield, but it needs shorter times to avoid product hydrolysis. In conclusion, the amount (or concentration) of enzyme must be evaluated under industrial conditions (higher volume and additional pH control) to determine the optimal number of enzymatic units.

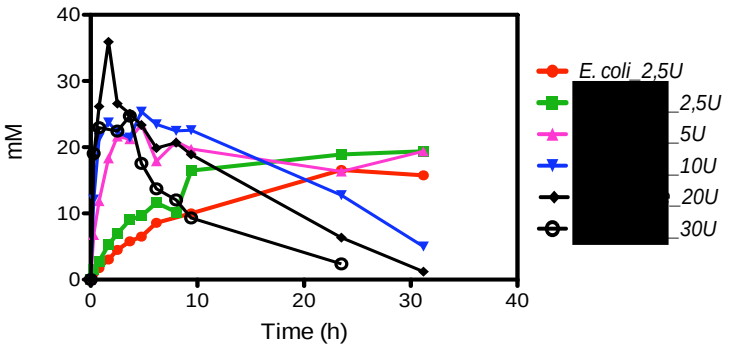


Figure 2. 20. *Gaxilose concentration on time under industrial conditions and different amounts of the* [redacted] *enzyme. CONDITIONS: 0.1 M oNPG, 0.4 M xylose, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Product quantified by HPLC-MS*

2.4 Best candidate

After analysing the candidates under three different conditions (*laboratory, analytical* and *industrial*), β-galactosidase from [redacted] showed the best characteristics. The enzyme achieved the highest yields under the three conditions, and moreover showed the highest T/H ratios.

Laboratory conditions allowed selecting three of the nine candidates: those enzymes presented the same or higher yields than the β-galactosidase from *E. coli*. Furthermore, the enzyme from [redacted] reached three-fold higher yield than the *E. coli* enzyme.

Analytical conditions were used to evaluate three of the nine enzymes in order to monitor the reactions along time up to high degree of conversion in homogeneous conditions. These conditions exhibited the same behaviour as in *laboratory* conditions:

[redacted] synthesised larger amount than *E. coli. L. delbrueckii* and *S. rectivirgula* produced, approximately, the same gaxilose amount than *E. coli*.

Finally, *industrial conditions* confirmed that the enzyme from [redacted] synthesised the highest amount of gaxilose and furthermore the extent of gaxilose hydrolysis was lower than that of the other candidates.

In addition, T/H ratio was higher for [redacted] β-gal under the three analysed conditions (Table 2. 11). Industrial conditions showed the highest T/H ratio, followed by *laboratory* conditions. This behaviour is a result of most favourable conditions (higher substrate concentration) for transglycosylation (thermodynamic control).

	T/H ratio		
	Lab conditions	**Analytical conditions**	**Industrial conditions**
[redacted]	1.52	$5.3 \cdot 10^{-2}$	3.44
E. coli	0.69	$2.6 \cdot 10^{-2}$	1.57

Table 2. 11. *Comparation of the T/H ratio at the different reaction conditions of enzymes from* E.coli *and* [redacted]

It must be take into account that the [redacted] enzyme also showed higher regioslectivity under all conditions analysed (Table 2. 12).

	Regioselectivity		
	Lab conditions	**Analytical conditions**	**Industrial conditions**
[redacted]	79-81%	73-89%	70%
E. coli	60-70%	75-87%	80-90%

Table 2. 12. *Comparation of regioselectivity under different reaction conditions of* E. coli *and* [redacted] *enzymes*

All these reasons made [redacted] β-galactosidase the new selected biocatalyst for gaxilose production.

CLOSING REMARKS

- This chapter used CAZy database to search for enzymes classified and characterised as β-galactosidases.
- A total of 123 enzymes were selected and among them 45 β-galactosidases with reported transglycosylation activity were detected.
- Bibliographic search and phylogenetic analyses allowed to reduce the number of enzymes. Finally, nine enzymes were selected to be expressed, purified and characterised.
- Glycosidase, hydrolase and transglycosydase activities of the nine candidates were quantified under different reaction conditions. Firstly, *laboratory conditions* discarded six enzymes and one enzyme, the β-galactosidase from [redacted] stood out. Under *analytical* and *industrial conditions* [redacted] β-galactosidase confirmed its ability to synthesise gaxilose in higher yields and better T/H ratios than the currently used *E. coli* β-galactosidase. Therefore, β-galactosidase from [redacted] has been the selected biocatalyst for gaxilose synthesis.

CHAPTER 3. New enzyme

CHAPTER 3. NEW ENZYME

OPENING

The β-galactosidase from [redacted] was selected among nine candidate enzymes to be used in the industrial production of gaxilose. This new enzyme presented better characteristics under laboratory conditions (greater yields, higher T/H ratio) compared to the *E. coli* enzyme.

Next step was **characterizing** β-galactosidase from [redacted] in order to describe it and understand its behaviour. Furthermore, the production of the enzyme has to be scaled up to test the enzyme under *semi*-industrial conditions, to finally reach industrial production level (1000 L bioreactor).

The hydrolytic activity (primary and secondary) of [redacted] β-galactosidase impeded to achieve higher yields of gaxilose production. In this scenario, **protein engineering** appeared as a solution to modify, and improve, enzyme activity. It offers several tools to change enzyme characteristics and finally optimize gaxilose production.

Protein engineering is based on the creation of altered forms of a known enzyme that exhibit one or more of the following properties: i) an increased catalytic function, ii) an altered substrate specificity or regioselectivity (the engineered protein is capable of catalysing the conversion of substrates other than the wildt type enzyme), iii) an increased in stability in the reaction environment (thermostability, organic solvent resistance), iv) novel properties.

Protein engineering can be carried out by two main strategies: rational design or directed evolution. In *rational design*, specific changes in amino acid sequence are proposed based on a detailed knowledge of protein structure, function and mechanism, and then introduced using site directed mutagenesis [301]. Therefore a deep understanding of the mechanism is needed. On the other hand, *directed evolution* does not require information about enzyme structure or mechanism. This technique employs random processes to create libraries of mutagenized genes. The main limitation of this technique is the prerequisite for a sensitive and efficient method for screening a large number of potential mutants [302,303]. Combining both strategies, a semi rational design offers a suitable solution. This approach utilizes information of protein sequence, structure and/or function to preselect promising target sites and limited amino acid diversity. Focusing on specific amino acid position translates into dramatically reduced library sizes [304–306].

Directed evolution strategies are subject to genetic diversity and high throughput screening or selection (HTSOS) methods. Genetic diversity can be generated by advanced molecular biology methods including random mutagenesis, gene recombination or semi-rational mutagenesis (focused mutagenesis) [302,307–309]: the mutant libraries quality plays a key role in finding improved mutants.

The **High-throughput screening (HTS)** method is one of the crucial steps in any directed evolution experiment. Screening refers to evaluation of every protein for the desired property, while selection eliminates automatically non-functional variants. Most high-throughput screening processes used in GHs are based on spectroscopically monitoring the progress of a chemical reaction. Several fluorogenic or chromogenic substrates can be used to detect and monitor cleavage reactions, but they are not associated with bond formation (transglycosylation). Moreover, another difficulty for HTS is the need of using substrate and reactions as similar as possible to the actual ones in the real biocatalytic reaction.

3.1 Characterization

CHAPTER 2 showed that β-galactosidase from [redacted] catalysed gaxilose synthesis in higher yield than the *E. coli* β-galactosidase and that the other eight candidates (Table 3. 1). [redacted] enzyme was able to improve *E. coli* enzyme features in all analysed conditions.

	[redacted] characteristics (*E. coli* characteristics)		
	MAX. β-1,4	**% regioselec.**	**T/H ratio**
Lab conditions	7 - 8 mM (3-4 mM)	79 - 86% (60-70%)	1.54 (0.6-0.7)
Analytical conditions	0.06 mM (0.03 mM)	73 - 89% (75-87%)	0.05 - 0.06 (0.02-0.03)
Industrial conditions	19 - 35 mM (17 mM)	80 - 90% (70-75%)	3 - 4 (1.9)

Table 3. 1. *Summary of maximum gaxilose, regioselectivity and T/H ratio achieved by* [redacted] *enzyme under different conditions and compared to the* E. coli *enzyme*

In this chapter, [redacted] β-galactosidase was further characterised in order to describe and understand its behaviour. Different properties were evaluated:

- temperature and pH for optimal transglycosidase activity (comparison between glycosidase and hydrolase activity)
- evaluation of primary and secondary hydrolysis
- transglycosylase reaction under industrial conditions (50 g *o*NPG and 12 kg *o*NPG)

3.1.1 Transglycosidase and hydrolase specific activities

Transglycosidase and hydrolase specific activity were determined under analytical conditions (5 mM *o*NPG, 25 mM xylose).

Transglycosidase specific activity was evaluated and compared to *E. coli* β-galactosidase (Figure 3. 1 and Table 3. 2). Transglycosidase specific activity was analysed by two procedures: using enzyme concentration (mM) or using enzymatic units (U) (corresponding to the hydrolase activity with oPNG substrate, glycosidase as described previously (see CHAPTER 1, *1.3.1*)).

Analysis of transglycosidase activity using enzyme concentration (mM) showed that [redacted] β-galactosidase was 12-fold slower than *E. coli* β-galactosidase (*right*- Figure 3. 1), a reduction also observed in its hydrolytic activity (see CHAPTER 2, *2.3.2* (Table 2. 5)).

Analysis using enzymatic units exhibited same transglycosidase activity for [redacted] and *E. coli* β-galactosidase (*left*-Figure 3.1 and Table 3. 2*).*

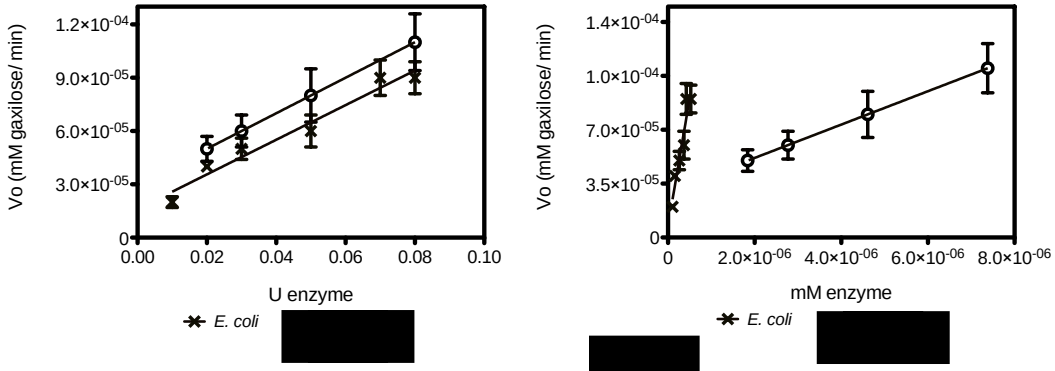


Figure 3. 1. *Transglycosidase specific activity of* E. coli *and* [redacted] *β-galactosidases under analytical conditions.* Left: *initial rate of gaxilose production at different Units of enzyme used*. Right: *transglycosylase initial rate at different enzyme concentration (mM). CONDITIONS:* 5 *mM* o*NPG, 25 mM xylose, 37°C, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

	Transglycosidase specific activity	
	V_o *vs* U enzyme (mM gax·min^{-1}·U^{-1})	**V_o *vs* mM enzyme** (min^{-1})
E. coli	(9.75 ± 0.97)·10^{-4}	165.80 ± 24.88
[redacted]	(1.00 ± 0.19)·10^{-3}	10.84 ± 0.01

Table 3. 2. *Transglycosidase specific activity of* [redacted] *and* E. coli *β-galactosidases under analytical conditions: determined by enzyme units added (left) and by mM of enzyme (right) CONDITIONS:* 5 *mM* o*NPG, 25 mM xylose, 0.01-0.08 U of enzyme, 37°C, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

Hydrolase activity is also present in transglycosylase reaction (as explained CHAPTER 1, *1.3.2*), (*B- and B'* Figure 3. 2). Galactose release is the result of *o*NPG and gaxilose hydrolysis. Initial rates calculation (less than 10% conversion) assumed that secondary hydrolysis was negligible.

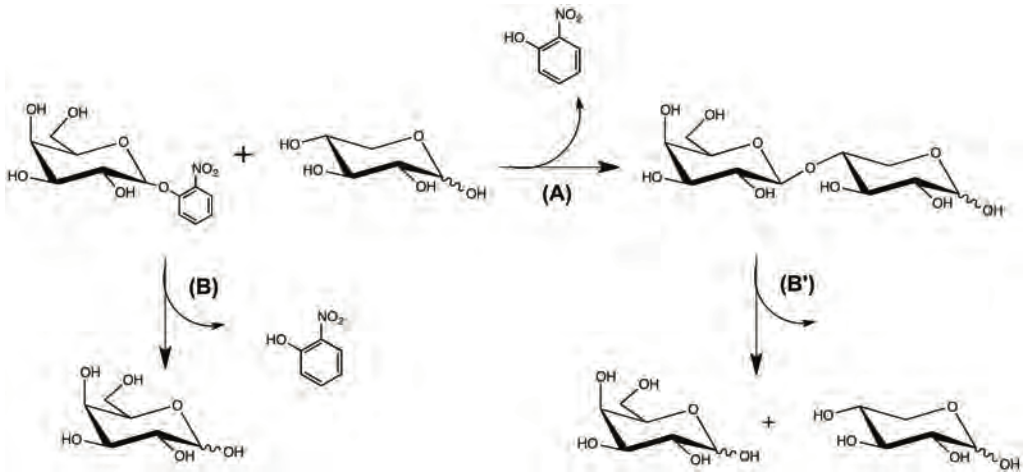


Figure 3. 2. *Paralel reactions using* o*NPG as acceptor and xylose as a donor. (A) Reaction of transglycosylation. (B) Primary hydrolysis of donor. (B') Secondary hydrolysis of product (gaxilose)*

Therefore, **hydrolase specific activity** could be also determined under transglycosylase reaction (Figure 3. 3 and Table 3. 3). Activity could be evaluated using enzyme concentration (mM) or using enzymatic units (U), as did in transglycosidase specific activity. It showed that *E. coli* β-galactosidase hydrolyses *o*NPG with higher initial rates tha[redacted]enzyme in both comparative analysis (1.7-folds and 27.1-folds, respectively).

T/H ratio exhibited higher values for [redacted], as shown in all conditions tested in *selection procedure* (CHAPTER 2) (*right*-Figure 3. 3). This feature is essential in order to obtain higher gaxilose production.

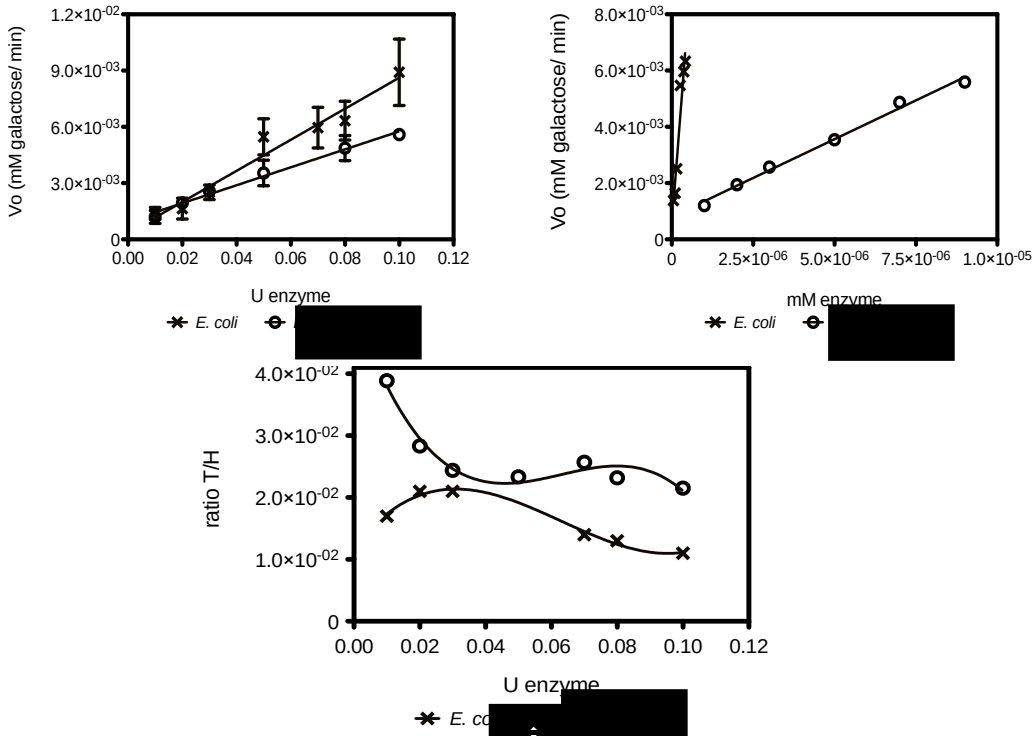


Figure 3. 3. *Hydrolase specific activity by* E. coli *and* [redacted] *enzymes.* Left: *initial rate of galactose release at different Units of enzyme.* Right: *initial rate of galactose release at different concentration (mM) of enzyme.* Down: *Ratio of initial rates of trangslycosylation and hydrolysis ($V_{o,transgl.}/V_{o.hydro}$) at different Units used. CONDITIONS:* 5 *mM* o*NPG, 25 mM xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

	Hydrolase specific activity	
	V_o *vs* U enzyme (mM gal·min^{-1}·U^{-1})	**V_o *vs* mM enzyme** (min^{-1})
E. coli	(8.23± 0.73)·10^{-2}	14,969 ± 1,922
[redacted]	(4.79± 0.24)·10^{-2}	551.5 ± 24.3

Table 3. 3. *Hydrolase specific activity o*[redacted]*and* E. coli *β-galactosidases of substrate (*o*NPG) and product (gaxilose) (data from Figure 3. 3)*

3.1.2 pH and temperature profile

The effects of pH and temperature on the transglycosidase activity of [redacted] β-galactosidase were evaluated.

Optimum pH for hydrolase activity was reported previously. It was described that the enzyme was highly active and stable in the pH range of 6.5-10.5 [291]. In this project pH effects were evaluated for the transglycosylase and glycosidase reactions**.**

- k_{cat}/K_M values were calculated for transglycosylation reactions. In those reactions two activities could be evaluated (Figure 3. 2): A) transglycosidase activity (*green*-Figure 3. 4) and B) hydrolase activity (*black*-Figure 3. 4).
- k_{cat}/K_M for glycosidase activity (*o*NPG) was also evaluated (*blue*-Figure 3. 4).

The pH profiles of the two activities follow a double ionization curve. Each activity shows a bell-shaped profile that does not overlap. Transglycosidase pH curve was quite displaced to acidic values and nucleophile pK_a value was different for both activities (Table 3. 4). Nucleophile pK_a value was 5.6 ± 0.1 for transglycosidase activity and 6.2 ± 0.1 for hydrolase. Hydrolase and transglycosidase activities share the same mechanism, but they did not have the same energy barriers.

	Transglycosylation reaction		**Glycosidase reaction (*o*NPG)**
	Gaxilose (transglyco.)	**Galactose (hydro)**	
pK_a (nucleophile residue)	5.6 ± 0.1	6.2 ± 0.1	6.1 ± 0.1
pK_a (A/B residue)	7.6 ± 0.2	7.2 ± 0,2	7.9 ± 0.3

Table 3. 4. *pK_a values of nucleophile and acid/base residues determined in transglycosylase and hydrolysis reactions of [redacted] β-galactosidase and by analysing its double activity. CONDITIONS: Tranglycosylase reaction:* 5 *mM o[redacted]M xylose, 0.01-0.8 U of enzyme 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS. Glycosidase reaction:* 13 *mM* o*NPG, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$).* o*NP released determined by absorbance (410 nm)*

Optimum pH for transglycosidase activity slightly shifted to acidic pH: the optimum was 6.5 instead of 7.0 for hydrolysis. However, both activities were maintained above 75% in the 6.0-7.5 pH range (Figure 3. 4). Industrial production could be set up at pH 6.0-6.5 to diminish hydrolytic activity (to 50-75%) and maintain transglycosidase activity up to 80%.

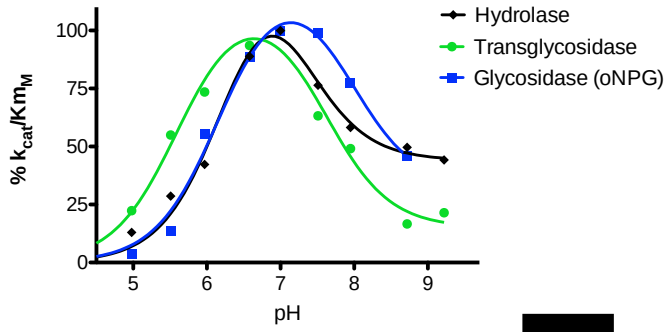


Figure 3. 4. *pH profile of glycosidase and transglycosylase reactions of* [redacted] *β-galactosidase: glycosidase, hydrolase and transgycosidase activities. CONDITIONS: Tranglycosylase reaction:* 5 *mM* oNPG, *25 mM xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$*). Products quantified by HPLC-MS. Glycosidase reaction:* 13 *mM* oNPG, *37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$*). oNP released determined by absorbance (410 nm)*

In addition, pH analysis was also tested at industrial conditions (0.1 M *o*NPG, 0.4 M xylose). In this case, 5.0-7.5 pH range was studied. Monitoring gaxilose concentration with time showed that pH 6.0 and pH 6.5 gave higher yields, similar to the results at pH 7.0 (*left*-Figure 3. 5). These results were in accordance with the pH profile determined previously. T/H ratio showed some but small dependence on pH (*right*-Figure 3. 5). The T/H ratio was higher at pH 6.0-7.0 than at pH 5.0-6.0.

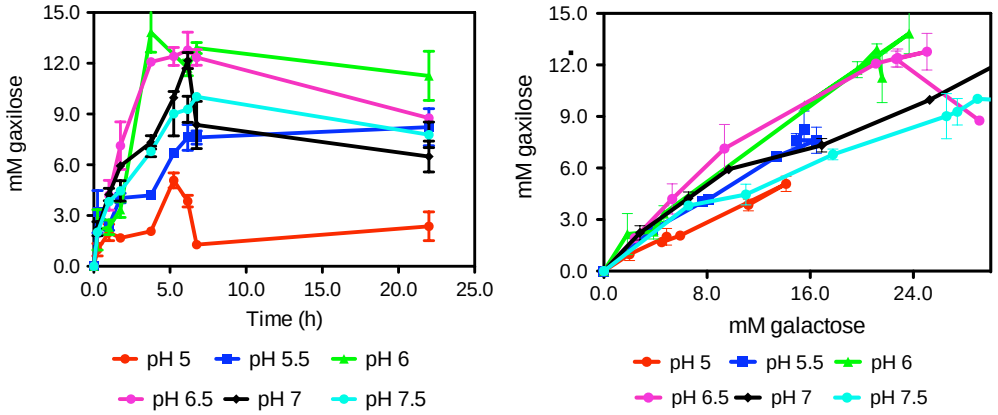


Figure 3. 5. *pH effect on transgalactosidase activity at industrial conditions.* Left: *mM gaxilose on time.* Right*: T/H ratio by representing amount of gaxilose* vs *amount of galactose. CONDITIONS: 0.1 M* oNPG, *0.4 M xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM* $MgCl_2$*). Products quantified by HPLC-MS*

Despite the spontaneous hydrolysis of *o*NPG and the thermal instability of gaxilose at high temperatures, a study of the effect of temperature on enzymatic activity was analysed, albeit at a relative narrow range. Transglycosylase reaction was evaluated in the 25-60ºC range; transglycosidase and hydrolase specific activities were determined (Figure 3. 6). Transglycosidase specific activity increased with temperature until 50-60ºC, where it remained constant. Hydrolase specific activity also increased with temperature (*cross*- Figure 3. 6). T/H ratio *versus* temperature is plotted in *right*-

(Figure 3. 6). It showed how the T/H ratio increases between 25ºC and 40ºC but then decreases. This feature will be taken into account for future reaction optimization at industrial level.

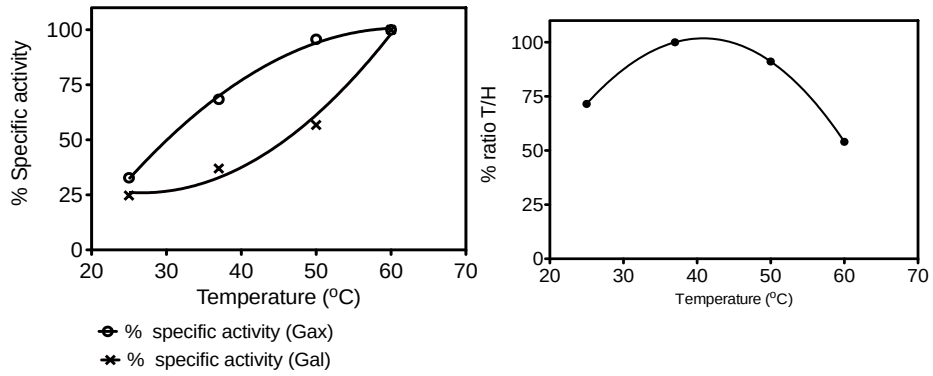


Figure 3. 6. *Temperature effect of initial rates of galactose and gaxilose production. CONDITIONS: 5 mM* o*NPG, 25mM M xylose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

3.1.3 Primary and secondary hydrolytic activity

Hydrolysis is always present in transglycosylase reactions (Figure 3. 2). β-Galactosidases can hydrolyse their substrate (primary hydrolysis, *o*NPG) and also hydrolyse the formed products (secondary hydrolysis, gaxilose), as explained in CHAPTER 2 (see *1.3.2*).

A hydrolytic evaluation (glycosidases activities) was carried out using *o*NPG and gaxilose as substrates. The main goal was to compare primary and secondary hydrolysis of β-galactosidases of *E. coli* and [redacted] on these substrates. Both enzymes were active on *o*NPG and gaxilose with different rates (Table 3. 5). Both enzymes hydrolyse *o*NPG faster than gaxilose. *E. coli* β-galactosidase hydrolyses *o*NPG 1.3-fold faster than gaxilose, in contrast to the [redacted]nzyme that does it 8.7-fold faster. This difference in substrate affinity could explain why the [redacted] enzyme presents higher productivity compared to the one from *E. coli*: whenever there may be *o*NPG in the transglycosylation reaction, gaxilose would remain *un*-hydrolyzed.

Although these hydrolysis reactions did not exactly mimic the actual reaction (not all substrate and products were present), the analysis allowed to conclude that *E. coli* β-galactosidase shows higher secondary hydrolysis than [redacted] β-galactosidase (6.7-fold).

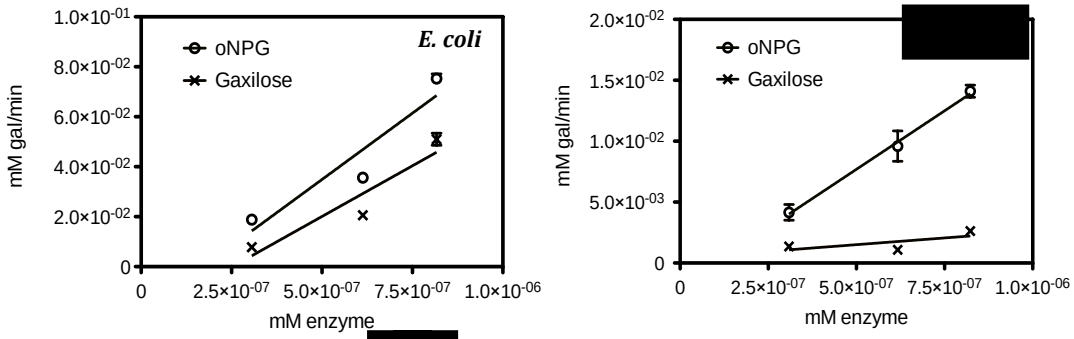


Figure 3. 7. *Hydrolytic activity of* [redacted] *and* E. coli *enzymes using* o*NPG or gaxilose as a substrates. CONDITIONS: 5 mM* o*NPG or gaxilose, 0.01-0.08 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). Products quantified by HPLC-MS*

	Activity (min^{-1})		*o*NPG/gaxilose
	***o*NPG**	**Gaxilose**	
E. coli	106,119 ± 38,223	81,107 ± 2,935	1.3
[redacted]	19,206 ± 1,157	2,207 ± 230	8.7
E. coli/[redacted]	5.5	36.7	

Table 3. 5 *Hydrolase activities of oNPG or gaxilose by β-galactosidases from* E. coli *and* [redacted]

E. coli β-galactosidase hydrolyses *o*NPG up to five-fold faster than the [redacted] enzyme. This activity could decrease *o*NPG susbtrate in *E. coli* catalysed reactions, and consequently decrease transglycosylation yields. A similar behaviour was observed for gaxilose hydrolysis but with larger differences: β-galactosidase from *E. coli* hydrolysed gaxilose up to 36-fold faster than [redacted] β-galactosidase. This characteristic, together with the difference in *o*NPG hydrolysis, explains the better yields in gaxilose production observed for the [redacted] β-galactosidase. The increase of gaxilose stability makes the process easy to control at industrial scale.

3.1.4 Semi-industrial approach

A test under industrial conditions was indispensable to fully evaluate [redacted] β-galactosidase and to be able to calculate the yield and quality of the purified isolated product. Gaxilose production scale up included an intermediate step at 10 g-scale. This evaluation was performed under controlled conditions of temperature and pH (addition of NaOH) during the complete reaction time (Figure 3. 8).

Figure 3. 8. *Reaction assembly at 10 g oNPG scale*

The whole process (upstream reaction and downstream purification) was performed in a similar manner to that performed with *E. coli* β-galactosidase (see *1.5*). While with β-galactosidase from *E. coli* the gaxilose yield was 21-22%, with [redacted] the yield was 35%. Moreover, the [redacted] enzyme had a higher regioslectivity as previously shown: under this conditions [redacted] presented a regioselectivity of 95%, while *E. coli* had it around 85%. Altogheter, these results indicate that [redacted] β-galactosidase is a better enzyme for the production of gaxilose than that of *E. coli*, and thus one of the goals of project had been completed.

Next steps for the scale-up of the process to an industrial scale requires the industrial production of the [redacted] β-galactosidase. This task has proven to be very challenging. At the point where the experiments of this thesis were interrupted, the production of the enzyme was achieved and the industrial synthesis of gaxilose is currently ongoing (Box 3. 1).

Box 3. 1 *Industrial production of gaxilose: enzyme amount*

[redacted]

3.2 Protein engineering

[redacted] β-galactosidase allowed to achieve higher amount of gaxilose than *E. coli* β-galactosidase. Nevertheless, there is still room for improvement of the enzymatic reaction, e.g. decreasing hydrolysis and increasing transglycosylation. In this scenario, protein engineering provides useful tools to modify and improve enzymes.

Protein engineering on [redacted] β-galactosidases was performed using three methodologies:

A) RATIONAL APPROACH. The [redacted] β-galactosidase structure was analysed searching for suitable positions to modify enzymatic activity. Point mutations were rationally designed and obtained by *site directed mutagenesis*. Their hydrolase and transglycosydase activities were evaluated.

B) SITE SATURATION MUTAGENESIS was carried out in previously studied positions. The main goal is searching for the optimum amino acid change. The libraries obtained by site saturation mutagenesis were analysed by a HTS assay (set up in *3.2.1*).

C) RANDOM MUTAGENESIS in the catalytic domain was carried out to find non-rationalizable unexpected mutations. The libraries obtained were analysed by a HTS assay (set up in *3.2.1*).

3.2.1 Screening method

In order to be able to analyse the mutant versions of the enzyme, an appropriate screening method needs to be implemented. Most of HTS methods for GHs are based on monitoring spectroscopically the progress of a chemical reaction. However, bond formation is difficult to detect in HTS mode. The screening method used in this project was based on a method developed by Prof. Charles Tellier group [310] (Unité de Fonctionnalité et Ingénierie des Protéines, Université Nantes, France).

3.2.1.1 Basis of the method

The selected method was developed in 2009 by Prof. Charles Tellier group [310]. The method is based on the detection of enzyme activation/deactivation by acceptor addition.

β-Galactosidase activity is measured using a substrate that, when hydrolysed, releases an aglycone that can be detected colorimetrically. Two substrates had been used for activity measurement: *o*NPG or *X*-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). *o*NPG hydrolysis releases *o*NP that will be detected by absorbance at 405 nm. *X*-gal hydrolysis will release 5-bromo-4-chloro-3-hydroxyindole, that spontaneously dimerizes and is oxidezed into 5-5'-dibromo-4,4'-dichloro-indigo, a blue product which is insoluble (Figure 3. 9).

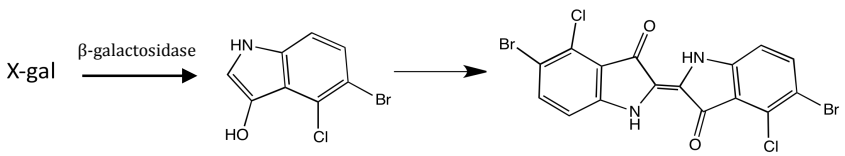


Figure 3. 9. *X-gal reaction*

Activity on X-gal and *o*NPG substrates measures glycosidase activity (*A*, *blue*- Figure 3.10). It is assumed that using low donor concentration, aglycone released is only linked with hydrolytic activity and not with self-condensation reaction. Acceptor (xylose) addition partially inhibits the hydrolytic activity (*red* Figure 3.10) and therefore a decrease of *o*NP, or indigo from X-gal, may be detected. Concurrently, the presence of xylose adds another reaction, transglycosylation (*B*, Figure 3.10) that also involves *o*NP or indigo release. If transglycosylation reaction is high enough, *o*NP or indigo release is higher than hydrolysis inhibition rate and therefore an increase of *o*NP or indigo is detected in the presence of xylose (*bright green*- Figure 3.10). On the other hand, if transglycosidase is moderate the increase of activity would remain covered up by the inhibition effect (*dark green*- Figure 3.10).

The main goal is to search for mutants that increase aglycone release in the presence of the acceptor. Some studies describe enzyme activation when acceptor is added and it is proved that this activation is related to higher transglycosidase activity [130,310]. Mutants showing an increase of *o*NP or indigo release upon xylose addition, may correspond to variants with higher **T/H ratio**.

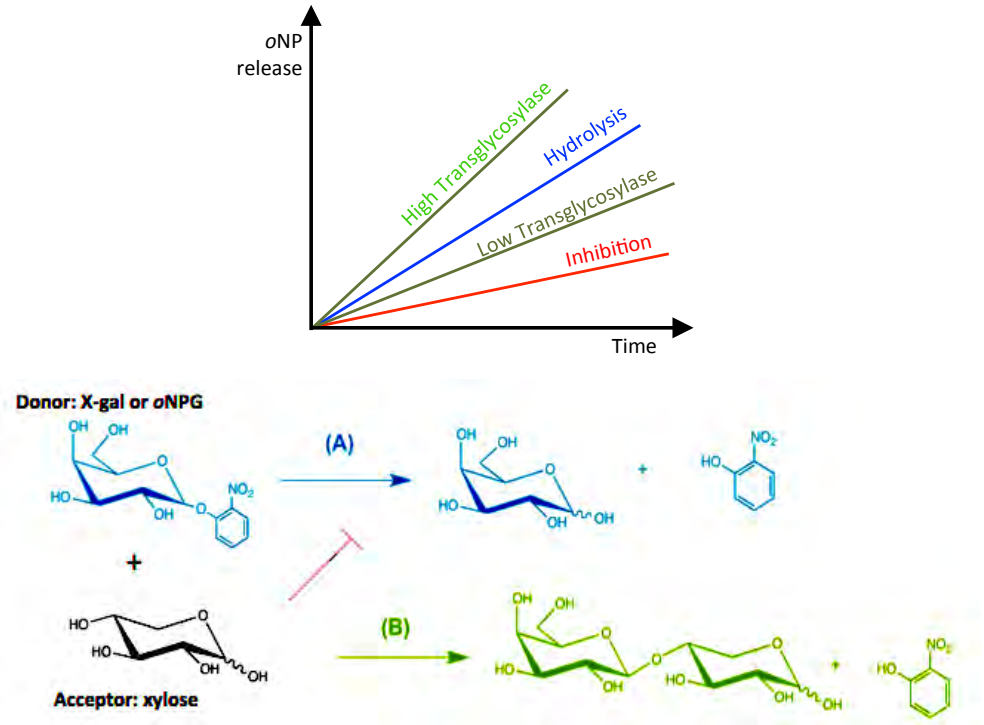


Figure 3. 10. *Xylose effect on β-galactosidase activity due to parallel reactions and the inhibition effect*

The screening method would be performed on whole cells and lysed cells. Assay in agar plate would be carried out without the lysis step, analysing colonies directly on agar plates using X-gal as substrate (*3.2.1.3*). Liquid screening assay would be done with a clarified cell extract after lysis step, using *o*NPG as substrate.

3.2.1.2 Strain and vector

The screening methods presented above will be performed on cells expressing the library of mutant proteins. Therefore, since the enzymes are not purified, endogenous *E. coli* galactosidase activity may interfere with that of the recombinant enzyme. In order to avoid false positives in the screening of mutants, an Δ*lacZ E. coli* strain was used. In addition to β-galactosidase deficiency, the *E. coli* strain should be able to express recombinant proteins. Among different strains, **TOP10** was selected. This strain has non-functional β-galactosidase (mutation *lac*ZΔM15), it is able to express recombinant proteins and besides it can be transformed with double or single-stranded DNA (Box 3. 2).

Box 3. 2 *TOP10 genotype and pBAD system*

F-mcrA Δ*(mrr-hsdRMS-mrBC)*Φ*80lacZ*Δ*M15* Δ *lacX74 recA1 araD139* Δ*(araleu)7697 galU galK rpsL* (StrR) endA1 *nupG*

This strain is efficient for transformation of unmethylated DNA from PCR amplifications (*hdsdR*). It is also useful for clean preparations of DNA due to the elimination of a non-specific digestion by Endonucelase I (*endA1*) and reduced occurrence of unwanted recombination in cloned DNA (*recA1).* These characteristics allowed used this strain to **transform complex libraries** from PCR constructions or ligated products, which used to be obtained in low amounts.
LacZ deletion (*lacZ*ΔM15) allows **screening for galactosidases** with no mask its activity.
pBAD vector was used to express protien. Recombinant protein expresion is possible using the *ara*BAD promoter (P_{BAD}) from *E. coli* and *araC* gene (encodes the regulatory protein for tight regulation of P_{BAD} promoter [411]).
Furthermore, TOP 10 strain is capable of transporting L-arabinose, but not metabolizing it: important to mantain **L-arabinose constant** inside the cell (*ara*BADC$^-$ and *ara*EFGH$^+$).
This strain is also *xyl-5*; that blocks catabolism of xylose.
The **pBAD plasmid**, induced by L-arabinose, expresses fusion proteins with a C-terminal polyhistidine (6xHis) tag for purification with a nickel-chelating resin. The plasmid carries ampicillin resistance as a marker for bacterial transformation.

***pBAD* expression vector** was used to subclone and express the [redacted] gene in TOP10 cells. *pBAD* vector allows the inducible expression of cloned proteins using L-arabinose as inducer.

Before setting up the screening conditions, it was checked that xylose and L-arabinose were able to enter the cell and were not metabolised. TLC analysis of a cell extract of a culture pre-incubated with xylose and L-arabinose revealed their presence within the cell (Figure 3.11).

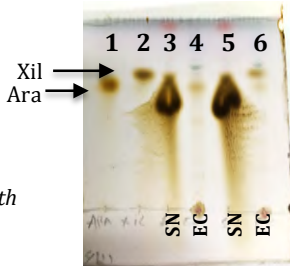
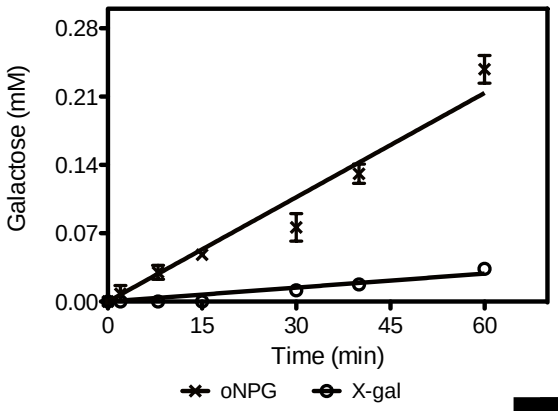


Figure 3. 11. *TLC analysis of cultures incubated with arabinose (0.2%) and xylose (20mM).Lane: 1 and 2- xylose and L-arabinose standards; 3 and 5 supernatant of cultures; 4-6 cell extract of cultures previously incubated with with arabinose and xylose. Mobile phase: 5 metanol, 5 ethanol, 3 water amonium. Reveletor H_2SO_4:MeOH:H_2O*

3.2.1.3 Set up conditions in agar plates

Several studies have applied similar screening methods to transglycosylating glycosidases in LB agar plates [129,130,310]. The agar plates assay used X-gal as a substrate. X-gal is not the substrate used for gaxilose production. Therefore, some differences between X-Gal and *o*NPG reactivity can occur. Glycosidase activity with X-Gal substrate was analysed to assess suitability of this substrate for the intended assay (Figure 3. 12).



V_H (*o*NPG)	$3.56 \cdot 10^{-3}$ mM/min
V_H (*X*-Gal)	$4.69 \cdot 10^{-4}$ mM/min
$\frac{V_H\ (oNPG)}{V_H\ (XGal)}$	7.6

Figure 3. 12 *Hydrolysis of* o*NPG and X-Gal on time by* [redacted] *β-galactosidase. Stock substrates were dissolved in DMSO due to the extreme low solubility of X-gal in water. oNPG was also dissolved in DMSO to perform both reactions under the same experimental conditions CONDITIONS: 13 mM substrate, 5 nM enzyme, 100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH7.0, 37ºC). Galactose quantified by HPLC-MS*

β-Galactosidase from [redacted] hydrolyses nearly eight-fold faster *o*NPG than *X*-Gal. Nevertheless, this difference has no effect for the experiment purpose. Even though the activity was lower, it may be high enough to detect activity and detect acceptor effect. Thus, transglycosylase reaction (adding xylose) should have been performed to compare X-gal and *o*NPG activity as donors.

Screening in agar plates was performed using two methodologies: **replica-plating assay** and **digital imaging assay**.

3.2.1.3.1 Replica-plating assay

Replica plating assay allowed the analysis of xylose effect in a parallel approach. The main purpose of this approach was to qualitatively compare intensities of blue (indigo blue, X-gal hydrolysis product). An LB-Ap-plate with colonies to be analysed (master plate) was imprinted onto a sterilised absorbent sheet of filter paper (Figure 3. 13). The imprint plate was transferred onto two plates with different compounds:

- One plate contained X-gal and L-arabinose for protein induction and substrate hydrolysis: glycosidase activity was evaluated.
- A second plate with X-gal, L-arabinose and xylose for protein induction and detection of transglycosidase activity.

Those mutants showing intense blue colour on the second plate (high transglycosidase activity), and with pale blue or white colonies (low glycosidase activity or inactivity) would be selected as potential mutants for further analysis.

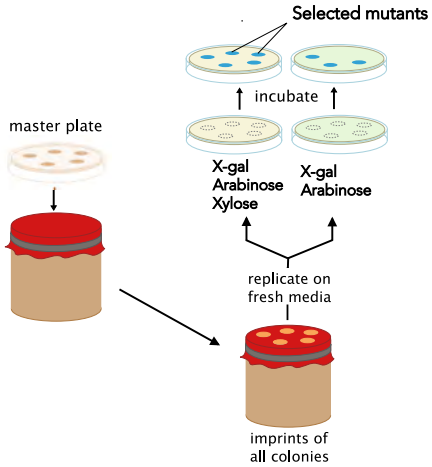


Figure 3. 13. *Replica plating method*

Screening conditions were set up with the wild type enzyme varying inducer concentration (L-arabinose, 0.2%-0.002%), xylose (10 mM - 100 mM), temperature (37ºC or 23ºC) and incubation time (6h - 14h). The addition of more inducer, higher temperature or longer incubation times generated colonies that turn blue too quickly and differences between plates would be difficult to detect. The finally selected conditions were the optimal ones according to laboratory availability.

Temperature increase caused faster activity and it was more difficult to detect changes in plates. Therefore, temperature was set up at 24ºC to ease color changes detection.

Arabinose concentration was evaluatued (Figure 3. 14) and, as expected, higher arabinose concentration increased protein expression. Final L-arabinose concentration was decided to be 0.2%.

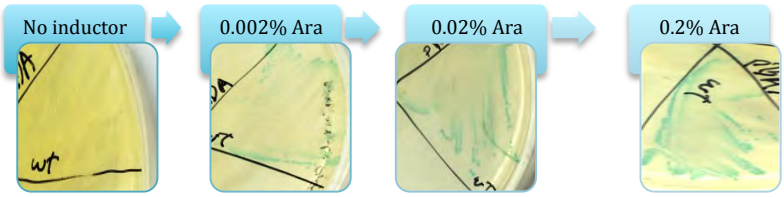


Figure 3. 14. *Diverse L-arabinose concentration effect on TOP10 cells carring pBAD vector with β-galactosidase gene from [redacted]. (Temperature incubation 24ºC, 12h, 0.1 mM X-gal)*

The effect of xylose concentration was also evaluated (Figure 3.15). Xylose at 5 mM had no detectable effect on activity, while 50 mM xylose decreased activity. The main drawback to finely tune xylose concentration was the lack of availability of a positive control; therefore, the finally selected xylose concentration (20 mM) was that which allowed detection of inhibition.

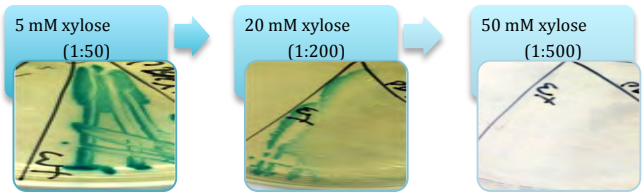


Figure 3.15. *Xylose effect evaluation on TOP10 cells carring pBAD vector with β-galactosidase gene from [redacted]. (Temperature incubation 24ºC, 8h, 0.1 mM X-gal, 0.2% L-arabinose)*

The wild type, an inactive mutant with hydrolytic residual activity (E439A) and an empty vector were evaluated in LB-Ap agar plates with inducer (L-arabinose), substrate (X-gal) and with or without xylose. The inactive mutant and empty vector showed no colour, whereas the wild type enzyme presented a soft inhibition detected by a slow reaction (pale blue) (Figure 3. 16).

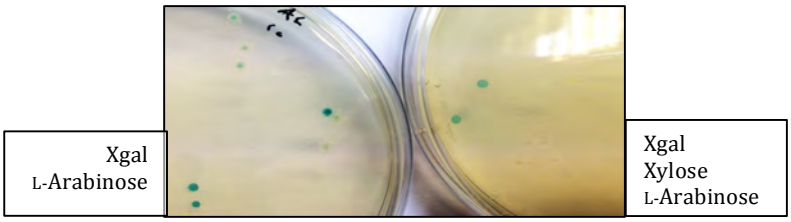


Figure 3. 16. *Colour of blue colonies (TOP10 cells carring pBAD vector with β-galactosidase gene from [redacted] in plates with X-gal, arabinose and with/without xylose*

The replica-plating assay was used for the screening of site-directed saturation mutagenesis libraries (*3.2.3.3*). Three libraries were built and more than 1,100 colonies were screened.

3.2.1.3.2 Digital imaging method (Prof. Charles Tellier laboratory, University of Nantes)

In collaboration with Prof. Charles Tellier group, their method was applied for the screening of random libraries of β-galactosidase from [redacted] (*3.2.3.3*).

Its main advantage, compared with the replica plating assay, is the use of a digital imaging-based system for quantification and comparison of enzymatic activities on agar plates.

The procedure is composed of three steps (Figure 3. 17). Firstly, transformed cells are spread onto a nitrocellulose membrane (0.45 µm, 21x21 cm^2). This membrane is then placed on LB-Ap agar plates and grown overnight at 37ºC. Secondly, the membrane, with colonies grown on it, is transferred onto a fresh agar plate of minimal medium (15 g/L agar, 100 mM phosphate, pH 7.0) containing *X*-Gal and L-arabinose for 1 hour at room temperature. This step allows the enzyme to be expressed and to hydrolyse the chromogenic *X*-Gal substrate, but not to grow. Thirdly, the nitrocellulose membrane is transferred on a new plate containing the donor substrate (*X*-Gal) and acceptor (xylose) (where transglycosidase takes place) and the first image is acquired (T_0). Finally, digital images are registered after 2-3 hours (T_f). Colony colour intensity at T_0 is related to it glycosidase activity, while colony intensity at T_f is connected with hydrolytic and transglycosylation reactions.

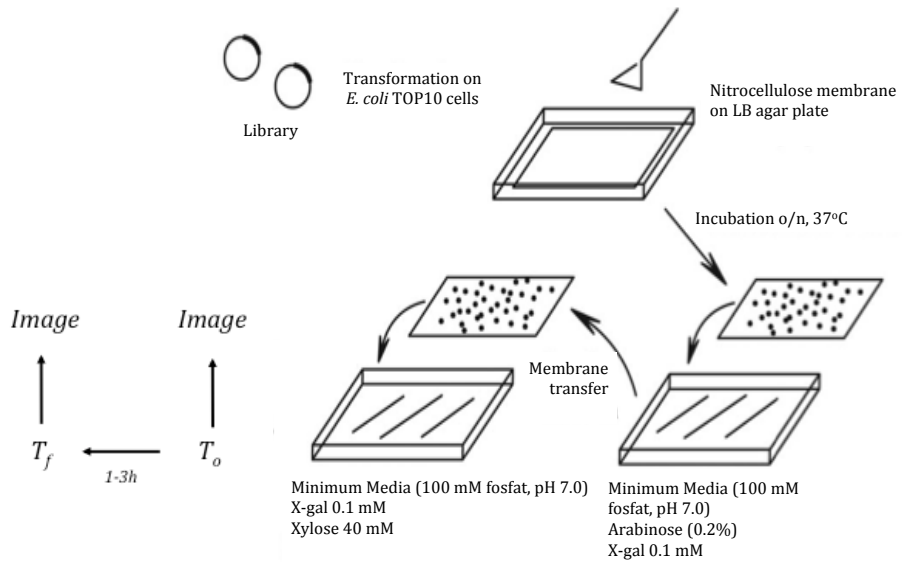


Figure 3. 17. *Methodology used for digital image procedure (Image from* [310]*)*

Digital image analyses involve the comparison of the intensity of the colonies at times T_f and T_0 and the identification of those colonies. The desired mutants are those with the highest change in intensity.

To illustrate the different outputs, the evolution of colour intensity of four selected mutants from the libraries is presented (Figure 3.18). Some of them show high colour intensity from the start and this intensity remains constant or slightly increases (mut1 and mut2, Figure 3.18). On the other hand, some colonies present lower initial intensity and a higher increase with time and xylose addition (mut3 and mut4, Figure 3. 18):these colonies are the selected ones.

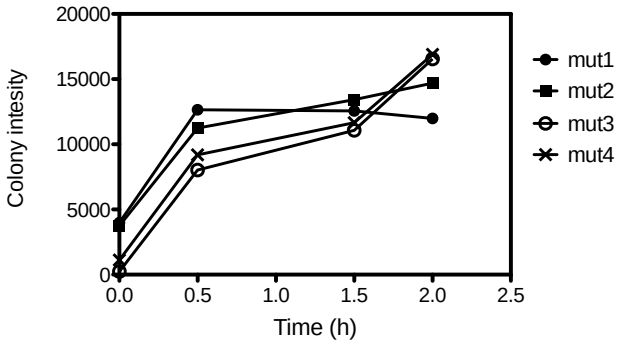


Figure 3. 18. *Colony intensity on time of some colonies of* E. coli *TOP10 expressing variants of* E. coli *β-galactosidase in the digital imaging method in LB-Ap-agar plates. CONDITIONS: incubation at 23-24ºC, plates with 0.1 mM X-gal, 0.2% L-arabinose, 40 mM xylose (100mM phosphate, pH 7.0)*

The digital imaging method was used to screen random mutagenesis libraries (*3.2.3.3*) and eight plates (21 x 21 cm^2) were analyzed (over 5,500 colonies).

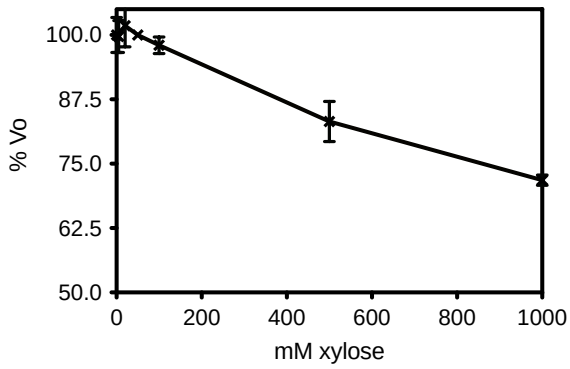
3.2.1.4 Liquid screening assay

The colonies selected as potential good mutants in solid agar assay were analysed in a second screening assay: a liquid assay. Liquid conditions allow quantification of the inhibition or the activation of the enzymatic activity by acceptor addition. Liquid screening was performed with *o*NPG as substrate, and the release of *o*NP was quantified spectrophotometrically (Abs 405 nm).

Assay set-up

Adjusting screening conditions involves the evaluation of the effect of xylose concentration on enzymatic release of *o*NP (Figure 3. 19). Assay set-up was carried out using the wild type enzyme. It was performed with purified protein and with crude extract of induced cell cultures; the same results were obtained in both cases.

Different concentrations of xylose (0-1 M) were added to reactions with constant amount of *o*NPG (10 mM). *o*NP release was monitorized and the initial rates were determined. Xylose did not induce more than a diminution of 5 % using up to 100 mM (Figure 3. 19): activity began to decrease with xylose 100 mM. Acceptor at 0.5 - 1 M



caused up to 30% inhibition. The final assay includes the use 0 - 300 mM of xylose (four concentrations) for mutants screening.

Figure 3. 19. *Xylose effect on wild type [redacted]-galactosidase activity (on* o*NPG). CONDITIONS:* 10 *mM* o*NPG and 0-1M xylose, 0.05 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$).* o*NP released determined by absorbance (410 nm)*

Xylose inhibits wild type enzyme activity, as described bibliographically. If transglycosylation is not high enough, inhibition effect masks transglycosylation, (Figure 3. 20).

It was checked that enzyme presented transglycosidase activity under those conditions (Figure 3. 20). Transglycosidase activity increased dramatically with the addition up to 100 mM xylose, then it increased at a lower rate. On the other hand, hydrolysis decreased with xylose concentration up to 100 mM xylose, when the hydrolytic activity was stable. The fact that both activities change their behaviours at xylose concentration higher than 100 mM xylose may indicate that above that concentration, xylose inhibits hydrolase and transglycosidase activities (in accordance with xylose effect on *o*NP release (Figure 3. 19) where inhibition in *o*NP release is detected from xylose 100 mM).

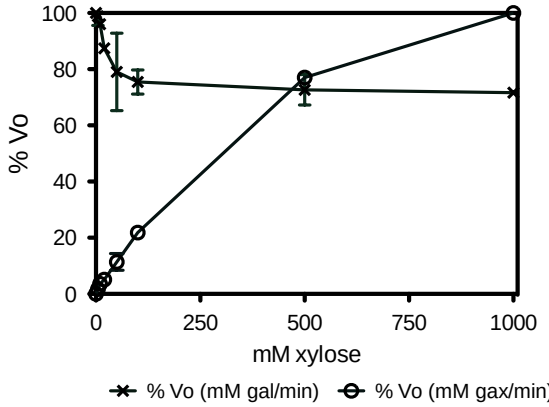


Figure 3. 20. *Effect of xylose in initial rate of gaxilose production and galactose release (trasnglycosylation reaction). CONDITIONS: 10 mM oNPG and 0-1 M xylose, 0.05 U of enzyme, 100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH7.0, 37ºC). Products were quantified by HPLC-MS*

Liquid screening assay was used as second step for the screening of site saturation mutagenesis libraries (*3.2.3.2*) and random mutagenesis libraries (*3.2.3.3*).

3.2.2 Protein structure

Protein engineering can be approached by rational design or random mutagenesis. In this work, both strategies were used. The rational approach requires previous knowledge of protein structure to design mutants in relevant positions within the enzyme.

3.2.2.1 Building a model

At the time when the rational design was attained, the 3D structure of [redacted] β-galactosidase was not known. For this reason, a structural model was built using *E. coli* β-galactosidase structure as template (homology modelling, (Box 3. 3)), while the [redacted] β-galactosidase structure was being solved in collaboration with the group of Dr Alberto Marina at Instituto de Biomedicina de Valencia (IBV-CSIC) (*3.2.2.2*).

Structural model was built using the structure from *E. coli* β-galactosidase as template. Both proteins are composed of five domains and they form a homotetramer. They share 65% of identity; and besides, the catalytic domain has up 85% sequence identity (Table 3. 6).

Box 3. 3 *Protein structure modeling*

Homology modelling plays a central role in determining protein structure; it is extensively used due lack of experimentally known 3D structures of some target proteins. Computational approaches are employed to bridge the gap between the number of known sequences and that of 3D models. A protein sequence with over 30% identity to a known structure can often be predicted with an accuracy equivalent to a low-resolution X-Ray structure [412]. Recent advances in homology modelling (e.g. detecting distant homologues, aligning sequences with template structures, modelling of loops and side chains, as well as detecting errors in a model) have contributed to a reliable prediction of protein structure [413].

Homology modelling usually consists of the following steps [414]: 1) identification of the homologue of known structure; 2) alignment of the query sequence to the template structure; 3) building of the model based on the alignment; 4) assessment and refinement of the model. It is established that when sequence identity is above 40%, the alignment may be straight forward.

Domain 1	72%	
Domain 2	41%	
Domain 3	85%	65%
Domain 4	49%	
Domain 5	57%	

Table 3. 6. *Percentatge of identity between sequences of β-galactosidases from* E. coli *and* [redacted]

Template

The *E. coli* β-galactosidase has two binding modes, as described in the introduction (*i.4.2.3.1*): the *shallow* and the *deep* mode. The main structural changes are a shift in a mobile loop (Gly794-Pro803) of 9 Å and a swing of an amino acid side chain (Phe601) (Figure 3. 21).

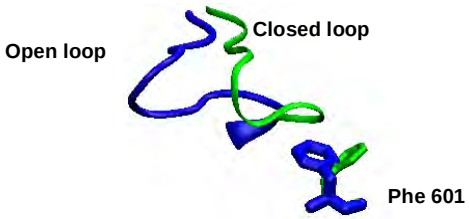


Figure 3. 21. *Structural changes in shallow and deep mode. Colours: shallow mode-blue; deep mode-green. Visualitzation by VMD software: loop represented by New Cartoon and residu by licorice*

There were fifty-nine structures of β-galactosidase from *E. coli* in PDB database (www.rcsb.org [185,186]) (as of April 2016); among them, there were ten with substrate bound in a *deep* binding mode. Structure in *deep* bindin*g* mode was selected to be employed as a template because it is the active form of the enzyme. Structure *4DUV* from PDB was selected to be used as a template (it is in *deep* binding mode and it has a substrate covalently bound, 2-deoxy-β-D-galactose, at 1.73 Å to the nucleophile residue).

Model

There are different servers to build a protein model, all of them based on the use of experimentally determined structures of related family members as templates (e.g. HHpred, I-Tasser, M4T, Modeller, Swiss-Model). SWISS-MODEL server was used to construct this model (www.swissmodel.expasy.org [311–314]).

Several tools allow the evaluation of the built model [315]. SWISS-MODEL server offers some quality evaluations, e.g. the Global Model Quality Estimation (GMQE). GMQE is a quality estimation which combines properties form the target-template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1: higher numbers indicate higher reliability. Our model had 0.83 and it was considered reliable for further evaluations.

The modelled structure of [redacted] β-galactosidase is composed of a central domain (*green*- Figure 3. 22), which is the catalytic domain, sorrounded by two domains: domain one (*blue*- Figure 3. 22) and domain five (*orange*- Figure 3. 22). Domain two (*red*- Figure 3. 22) seems to be a link between domains one and three, and domain four (*grey*) between three and five.

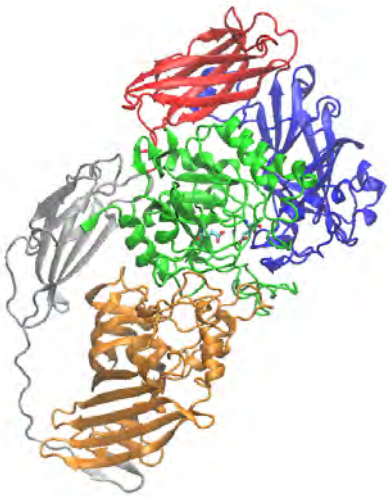


Figure 3. 22. *Protein structure model of the* [redacted] *ß-galactosidase (build with SWISS-PROT and 4DUV as template). Visualitzation by VMD software: representation by New Cartoon and each domain is in different colour (domain 1-blue; domain 2-red; domain 3- green; domain 4-grey; domain 5-orange*

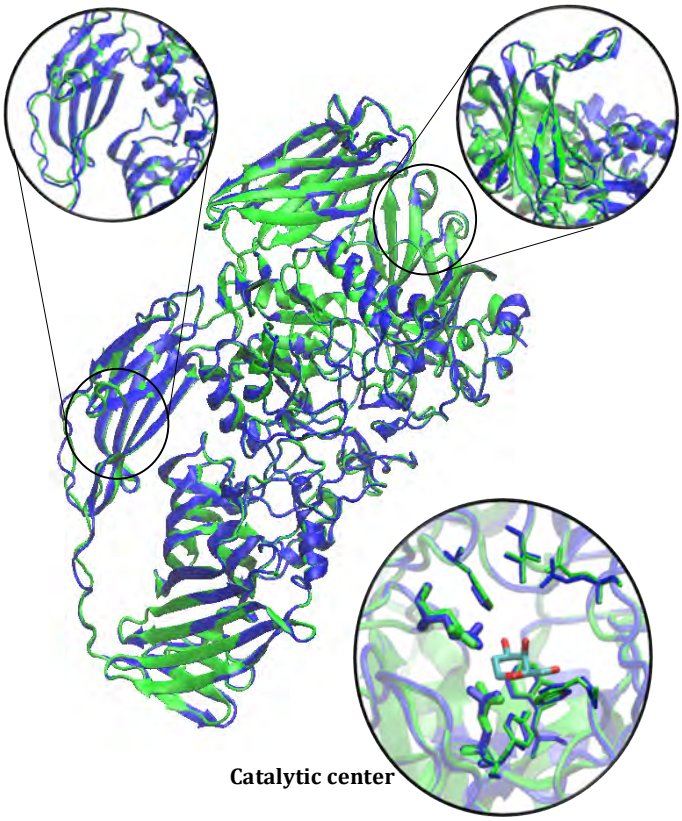


Figure 3. 23. *Structures of* E.coli *(blue) and* [redacted] *model (green) β-galactosidases superposed. Some divergences are shown (upper circles) and catalytic center of both proteins (down circle). Visualitzation by VMD software: representation by New Cartoon*

Protein structures superposition only let to detect small divergences (*up-circles*-Figure 3. 23). Catalytic centre (*down-right*-Figure 3. 23*)* is formed, nearly, by same amino acids (comparison at (Table 3. 7)) than the *E. coli* enzyme.

3.2.2.2 Protein structural analysis: crystallization trials, X-Ray diffraction and structure determination

The β-galactosidase from [redacted] had not a released 3D structure. In collaboration with research group of Dr. Alberto Marina at IBV-CSIC, the protein structure of [redacted] β-galactosidase was solved.

Twelve mg of inactive mutant E539A were used in several conditions for protein crystallization trials. E539A was used to obtain the crystal structure of the enzyme with entrapped product (gaxilose) to better understand product interactions and enzyme mechanistic details.

Large stable crystals were obtained and taken to ALBA Synchrotron (https://www.cells.es) for X-Ray diffraction. The diffraction process was successful and Dr. Alberto Marina's team could collect diffraction data.

The structure of [redacted]β-galactosidase was solved by molecular replacement at 2.1 Å resolution. Its structure was based on space group $P2_1$ and the unit cell was formed by a dimer of tetramers. Therefore, eight monomers composed the unit cell (each monomer had 1,029 amino acids).

Different structures were obtained with or without substrate. The analyses of these structures revealed only subtle differences between β-galactosidases from [redacted] and *E. coli*. (Note: at the time of the composition of this thesis, only structures with open loop where obtained (*shallow* mode)).

Crystal structure obtained was compared with the built model (*3.2.2.1*) (Figure 3. 24). The observed differences between both structures were:

- Mobile loop position: the model was built with *E. coli* β-galactosidase in deep binding mode while the crystal structure was obtained in shallow mode. Therefore, loop (Gly794-Pro803) and Phe603 have different positions in model and crystal structures.
- Non-structured sequence is found in the model (same part found in comparison between [redacted] β-galactosidase crystal structure and [redacted] model) on domain 4. It is distant from the catalytic centre and it has no interactions with other monomers.

Superposition of catalytic centre also exhibited that the only differences are in the loop position and Phe603 residue orientation.

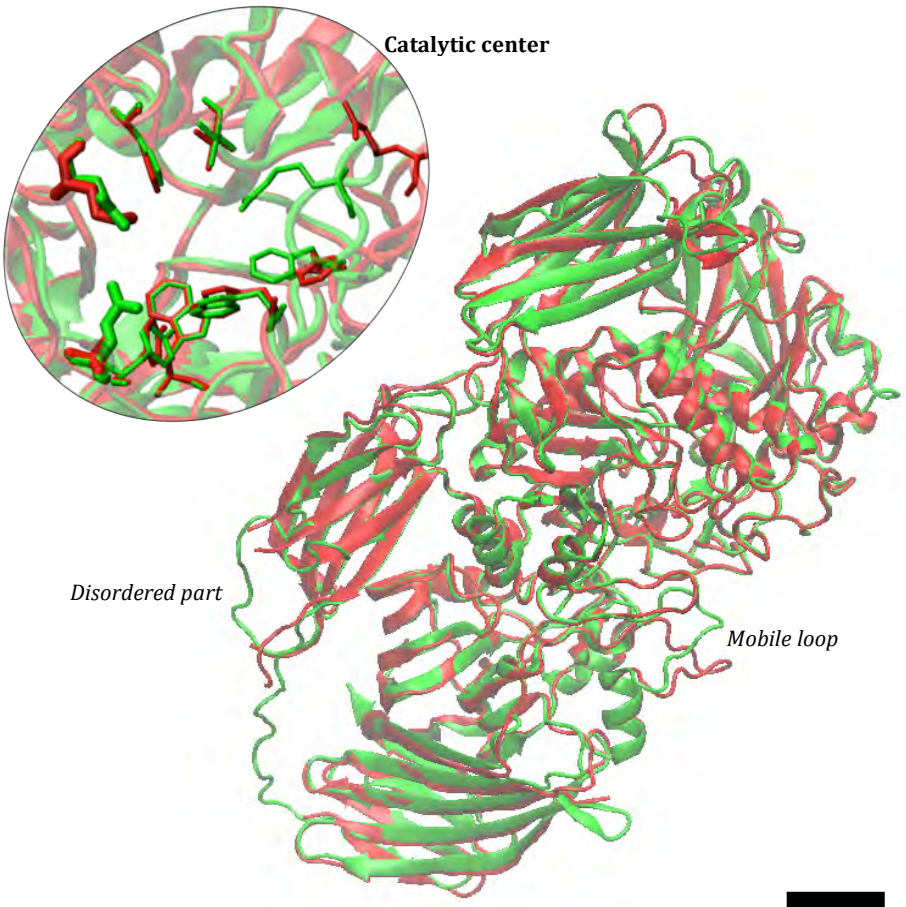


Figure 3. 24. *Superposition of crystal structure (red) and model built (green) of the* [redacted] *β-galactosidase. Up-circle: catalytic centre of both enzymes Visualitzation by VMD software: representation by New Cartoon*

3.2.3 Mutants design

Protein engineering on [redacted] β-galactosidases was performed using three methodologies:

RATIONAL APPROACH. [redacted] β-galactosidase structure was analysed searching for suitable positions to modify enzymatic activity.

SITE SATURATION MUTAGENESIS was carried out in positions previously identified in the rational approach.

RANDOM MUTAGENESIS in the catalytic domain.

3.2.3.1 Rational approach: point mutations

The first attempt to improve enzymatic activity was a rational approach. The first step was the analysis of the protein structure: catalytic site, including catalytic residues and residues involved in substrate binding.

β-Galactosidases from [redacted] and *E. coli* share 65% of identity: due to this high percentage of identity, some structural and experimental information reported for *E. coli* could be used to understand the new enzyme.

3.2.3.1.1 Selecting positions

Ligand dockings simulations

Xylose was employed in the docking (Box 3. 4) analyses to study acceptor interactions (+1 subsite) using the model previously constructed (*3.2.2.1*). Simulation was performed using a covalently-bound galactose in -1 subsite, to mimic an enzyme-substrate covalent intermediate. This would mimic xylose environment before the second step of the reaction (Figure 3. 25). The galactose molecule was added to the catalytic site of [redacted] enzyme by superposition of the model with *E. coli* structure that contains a covalently bound galactose moiety (2-deoxy-galactosyl-enzyme). Galactose was transferred from one structure to the other.

Box 3. 4 *Docking*

Docking procedures model physical interactions focusing on the final configuration of the complex. It is a useful tool for a first structural analysis of the active site of the enzyme.

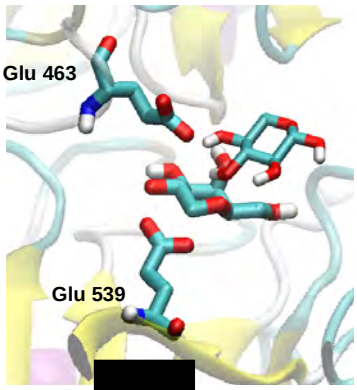


Figure 3. 25. *Catalytic pocket of the* [redacted] *β-galactosidase with galactose added by structure superposition and xylose added by docking. Visualitzation by VMD software: representation New Cartoon*

Relevant residues

More than twenty-five relevant residues linked to galactosidase activity and/or substrate(s) binding were proposed. Residues were selected because: A) they were described in the *E. coli* enzyme (65% identity) to have a role in activity or

substrate/product binding; B) they interact with xylose (found in dockings); C) they are close to the catalytic centre and could interact with solvent and facilitate water entrance.

All found residues are conserved residues in the *E. coli* enzyme, except for position 359. The 359 position is a histidine for *E. coli* and a leucine for [redacted] (Table 3. 7).

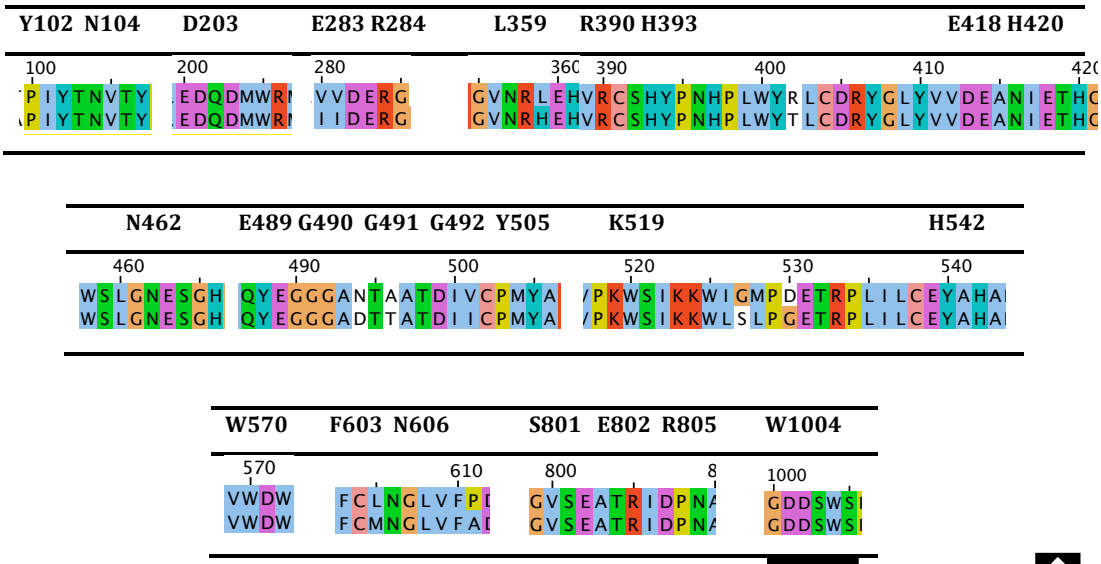
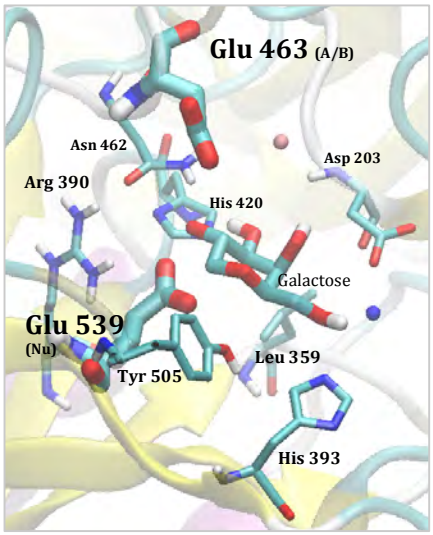


Table 3. 7. *Sequence aligment of important residues of β-galactosidases: Up-[redacted] and down-* E. coli *numbering)*

Selected residues can be classified in three groups based on their roles in enzyme mechanism: Subsite -1 (A), Subsite +1 (B) or other interactions (C):

A- Subsite -1

Asp203	Its carboxil group interacts with Na^+ and Tyr102 (C). Hydrogen bond is formed with hydroxyl group of C4 of the galactose.
Leu359	It is the only amino acid that differ in between catalytic centre of galactosidases from *E. coli* and [redacted] *E. coli* enzyme has a histidine that interacts with C3 hydroxyl of galactose through a water molecule.
Arg390	It interacts with nucleophile residue (E539).
His393	It interacts with C3 hydroxyl (hydroxyl that interacts with Mg^{2+} and two water molecules).
Asn462	It interacts with C2 hydroxyl and also with acid/base (E463).
Tyr505	It interacts with nucleophile residue and it

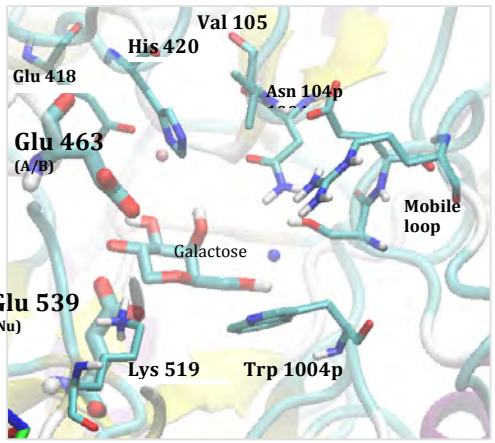


*Blue sphere: Na^+; Pink sphere: Mg^{2+}

	has no influence in binding.
His542	It interacts with galactose C6 hydroxyl and Na^{2+}.
Trp570	It is involved in galactose stacking.
Phe603	It shifts when substrate moves from *shallow* to *deep* mode. It also interacts with Na^{+} and with galactose hydroxymethyl.
Asn606	It interacts with hydroxyl group of C6.

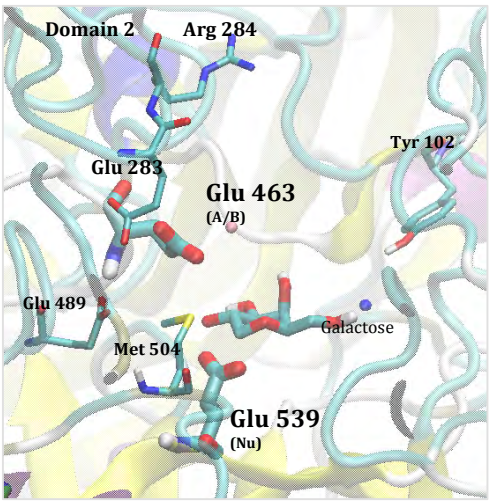
B- Subsite +1

Asn104	It interacts acceptor moiety.
Val105	It does not interact with the substrate but it is in the second shell of the catalytic subsite.
Glu418	It interacts with Mg^{2+}.
His420	It interacts with galactose C3 hydroxyl across a water molecule. It also interacts with Mg^{2+}
Lys519	It interacts with galatose C4 hydroxyl through a water molecule.
Ser801 **Glu802** **Arg805**	These residues belong to mobile loop and it interacts with substrate when loop is closed
Trp1004	It stacks glucose moiety (donor).



C- Other interactions

Tyr102	It interacts with Asp 203, which interacts with C4 hydroxyl of galactose (second shell residue).
Arg284	It belongs to loop of domain 2 and it interacts with catalytic center of other monomer.
Glu283	Belongs to loop of domain 2 and interacts with catalytic center of other monomer.
Glu489	It is 4.6 Å distance of acid/base residue. It is in the surface of catalytic cavity.
Gly 490-491-492	They form second sphere of catalytic subsite.
Met504	It is described as non-essential. It forms -1 subsite and it is involved in loop stabilisation.



Selected residues

Bibliographic search and docking results indicated that some amino acid residues were essential for enzymatic activity. For this reason, they were discarded for further mutagenesis because modifications were designed to modulate activity, or change water-substrate-enzyme interactions, but not to abolish enzymatic catalysis. Taking into account those considerations, eight of the twenty-five pre-selected amino acids were chosen for mutagenesis. The selected residues (column 1), their selected mutations (column 2) and the purpose of each mutation (column 3) are presented below:

Tyr505	**Phe**	It has the same structure but removing hydroxyl group. It has been described that the hydroxyl group helps nucleophile residue in second step in the reaction forming a hydrogen bond. Its substitution will decrease hydrolytic activity, but what about transglycosylation activity?
	Gln	It has the same length than Tyr and a functional group that can mantain H-bond. How will affect catalysis? And to T/H ratio?
Arg390	**Lys**	The change causes a length reduction but mantains functional group to keep interactions with nucleophile residue. It could change position of nucleophile, how will this affect activity?
	His	Structural change that will maintain interaction with nucleophile residue.
Leu359	**His**	It is the only different residue between catalytic centre of *E. coli* and [redacted] β-galactosidases. Will it behave as the *E. coli* enzyme? Will it change T/H ratio?
His420	**Trp**	This change will decrease size of +1 subsite pocket. It was expected a better fit (xylose is smaller than natural substrate, glucose).
Val105	**Trp**	It is a hydrophobic change and it decreases subsite size: xylose is smaller than natural substrate, glucose.
Asn104	**Leu**	It is a hydrophobic change in catalytic subsite entrance.
Phe603	**Leu**	Reduction size of this residue could move Trp1004, residue that stacks xylose. This movement could increase/modify xylose reactivity.
Asn606	**Ser**	This change will destabilize interactions with donor, galactose, and it could increase rate of second step of reaction favouring trasnglycosylase reaction.

3.2.3.1.2 Analysis of point mutations

Mutants glycosidase and transglycosidase activities were evaluated. Firstly, glycosidase activity on *o*NPG was studied (Table 3. 8). R390H mutant was the only non-active enzyme on *o*NPG. Mutants N104L, R390K, H420W, Y505F and Y505Q decreased activity by 200-2,000-fold. Decrease could be caused by: A) modification of essential amino acids; B) too radical modifications (*left*-Figure 3. 26). Mutations on R390 and Y505 could modify nucleophile reactivity, nucleophile position or even its pK_a value. On the other hand, positions H420 and N104 interact with acid/base residue and/or subsite +1. Trp for His420 substitution must be a too big size modification and it could be changing subsite cavity; or even losing Mg^{2+} stabilisation, which interacts

with A/B residue. Leu for Asp104 substitution could cause interaction losses that were needed to maintain cavity shape.

Mutants V105W, L359H, F603L and N606S reduced activity up to ten-fold. V105 and F603 are located in subsite +1, meanwhile L359 and N606 interact with the galactose moiety (*right*-Figure 3. 26). All of them are located at more than 6.8 Å from the catalytic residues.

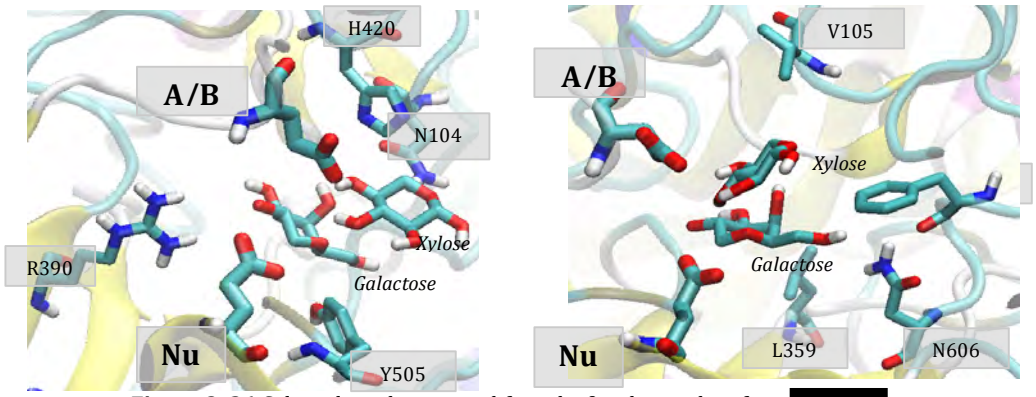


Figure 3. 26 *Selected residues to modify in the β-galactosidase from* [redacted]

		Glycosidase activity (13 mM oNPG)	Transglycosydase reaction		
			Analytic conditions (5 mM oNPG, 25 mM xylose)		Industrial conditions
		V_o(mutant)respect V_o(wt)	V_o(T) (mM gax /min/U)	V_o(Transg.)/ V_o(Hydro)	Max. β-1,4
Asn104	**Leu**	1/630	Not detected		Not detected
Val105	**Trp**	1/1.5	$(5.80\pm 0.16)\cdot10^{-3}$	0.01-0.02	8-9 mM
Leu359	**His**	1/3	$(1.25 \pm 0.08)\cdot10^{-4}$	0.01	0.8-1 mM
Arg390	**Lys**	1/2,600	Not detected		Not detected
	His	-	Not detected		Not detected
His420	**Trp**	1/430	Not detected		Not detected
Tyr505	**Phe**	1/600	minimum		minimum
	Gln	1/2,200	Not detected		Not detected
Phe603	**Leu**	1/2	$(1.24 \pm 0.33)\cdot10^{-3}$	0.01	2.5 mM
Asn606	**Ser**	1/12	$(1.09\pm 0.07)\cdot10^{-3}$	0.05	7-8 mM
Phe603 / Asn606	**Leu / Ser**	1/10	$(1.04 \pm 0.26)\cdot10^{-3}$	0.04-0.05	4 mM
Wild type			$(5.84 \pm 0.73)\cdot10^{-3}$	0.04-0.06	10 -12 mM

Table 3. 8. *Summary of point mutations evaluated. 1-Glycosidase activity (13 mM oNPG) expressed as a reduction of the activity respect to wild type enzyme. 2- Transglycosidase initial rate (mM gaxilose/min/U enzyme) under analytical conditions, 3- T/H ratio under analytical conditions and 4-max. gaxilose synthesised by mutants under industrial conditions (CONDITIONS: analytic: 5 mM oNPG and 25 mM xylose, 0.01-0.1 U of enzyme and industrial 0.1 M oNPG, 0.4M xylose, 1 U of enzyme), 100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH7.0, 37ºC). Products were quantified by HPLC-MS)*

Once glycosidase activity had been evaluated, **transglycosidase activities** at analytical and industrial conditions were analysed. Mutants with reduced glycosidase activity on *o*NPG did not exhibit transglycosidase activity at any condition assayed. Mutants that retain glycosidase activity revealed that transglycosidase initial rates, under analytical conditions (5 mM *o*NPG, 25 mM xylose), were of the same order of magnitude than wild type enzyme; except for L359H mutant that it showed less activity (one order of magnitude less). The T/H ratio was also evaluated and only two mutants (N606S and F603L/N606S) exhibited same ratio than that of the wild type enzyme.

Transglycosylase industrial conditions were also tested. All mutants that presented transglycosidase activity under analytical conditions, also exhibited it at industrial conditions. None of the mutants synthesised more gaxilose than wild type (*left*-Figure 3. 27). N606S and V105W mutants achieved similar yields than wild type, while L359H and F603L reached lower yields. Mutants' T/H ratios were not higher than that of the wild type (*right*-Figure 3. 27).

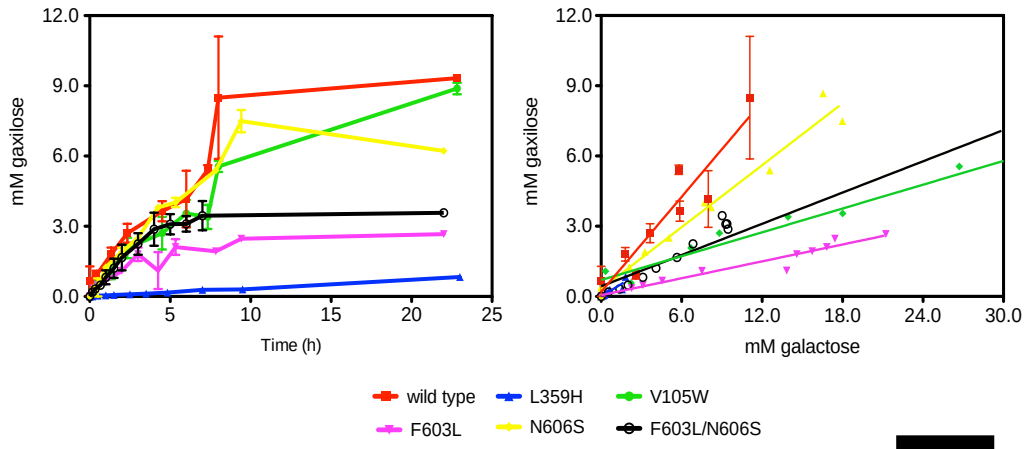


Figure 3. 27. *Analysis of gaxilose synthesis under transglycosidase industrial conditions by the* [redacted] *β-galactosidase wild type and mutants L359H, V105W, F603L, N606S and F603L/N606S.* Left: *gaxilose p on time by point mutants enzyme.* Right*: T/H ratio. CONDITIONS: 0.4 M oNPG, 0.1 M xylose, 1 U of enzyme (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0, 37ºC). Products were quantified by HPLC-MS*

Laboratory conditions were also analysed for a longer time course (20h) but the same behaviour was observed: the highest gaxilose yields and the highest T/H ratios belonged to the wild type enzyme (data not shown).

Point mutations did not render any improved enzyme. N606S and V105W maintained gaxilose production with lower T/H ratio. This outcome pointed at these positions as non-essential but with important effect on activity. Therefore, other amino acid substitutions in such positions may increase activity and/or the T/H ratio.

3.2.3.1.3 Site saturation mutagenesis

Point mutation analysis (*3.2.3.1*) allowed the identification of some relevant positions. Different enzymatic activities were obtained:

A) some mutants presented equivalent activity to the wild type enzyme (e.g. N606S, V105W)

B) other mutants left the enzyme with low or any activity:
   - because the mutated amino acid was essential for activity (direct interaction with catalytic residues, e.g. Y505 or R390)
   - or because the selected change was too radical (e.g. H420W)

Selected amino acids

Three positions were selected for site-directed saturation mutagenesis based on presented hypothesis.

1) V105 was selected because V105W mutant was able to maintain gaxilose synthesis, albeit at lower T/H ratio. It was hypothesised that another change could improve T/H ratio and consequently increase gaxilose synthesis.

2) L359 is the only amino acid residue in the active site of *E. coli* and [redacted] β-galactosidases that differs between both enzymes: *E. coli* has a histidine, while [redacted] has a leucine [redacted] L359H enzyme showed a decrease in activity and a decrease in T/H ratio. If the T/H ratio is compared between the three galactosidases (*E. coli,* [redacted] and L359H mutant from [redacted], L359H mutant and *E. coli* enzyme presented the same T/H ratio. Therefore, L359H mutant seemed to revert the effect on T/H ratio and behave as the *E. coli* enzyme (Figure 3. 28).

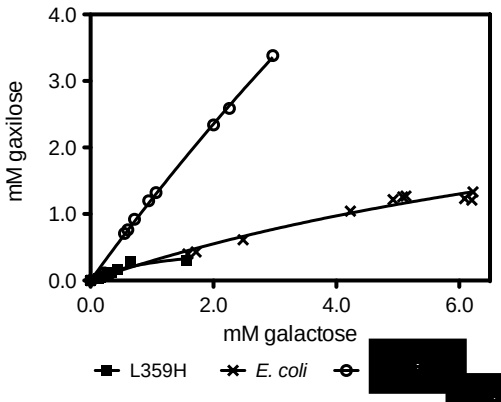


Figure 3. 28. *T/H ratio comparation between β-galactosidases from* E. coli, [redacted] *and mutant L359H of* [redacted] *under laboratory conditions (CONDITIONS: 50 mM oNPG, 250 mM xylose (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0, 37ºC).) Producst analysed by HPLC-MS)*

3) H420W mutant showed a huge decrease in activity. Position 420 interacts with galactose moiety through a water molecule and with the acid/base residue through

Mg^{2+}; it does not interact directly with the catalytic residues. Trp substitution for H420 causes a change that decrease activity (probably due to a loss of important interactions o a distorsion of the catalytic subsite); it was expected that another amino acid change would be able to improve activity.

Assembly of V105, H420 and L359 libraries

Site saturation mutagenesis was selected in order to find a better substitution in the chosen positions. Libraries of each position were composed of the twenty amino acids and one codon stop, obtained with degenerated primers (*NNK*) (Box 3. 5). The total number of codons was reduced to 32 and to cover 95% of all variants, 94 colonies needed to be screened.

Box 3. 5 *Library NNK*

Saturation mutagenesis is frequently used in deep mutational scanning. The amino acids at selected positions are replaced by the 20 amino acids. The randomization is obtained typically via degenerate primers containing a mixture of sequences at the chosen codons. To decrease the chances of introducing a premature stop codon, reduce codon sets are often used: NNB, NNS and NNK codons are popular choices that encode all twenty amino acids (N: A/C/G/C, B: C/G/T, S:C/G and K: G/T) [305,307,415,416]. NNS is equivalent to NNK. It is not immediately clear which of the two intermediates schemes, NNK or NNB, is better: NNB induces a lower probability for a stop codon (1/48 *vs* 1/32 in NNK), but NNK enjoys a lower ratio between the most common and the rarest amino acids. In addition, NNK library spans 32 codons and therefore, 94 colonies should be screened to cover 95% of all variants (NNB library has 48 codons and 142 colonies must be screened to cover 95% of sequence variants).
Other alternatives are more complicated like MAX, which assigns equal probabilities to each of the twenty amino acids without encoding stop codons [417].

NNK libraries were analysed with the HTS method described previously (*3.2.1*). Replica plating was used to find active mutations using the LB-plate screening methodology (*3.2.1.3.1*). Then, the liquid screening methodology was applied to refine the search and select few mutants to be analysed (*3.2.1.4*).

Firstly, it was tested that each library was degenerated. A pull of DNA plasmid sample obtained from the NNK library was sequenced. The goal was to detect if DNA bases had been randomized in desired position. Sequence analysis allowed to determine the variability in the selected positions (Figure 3. 29).

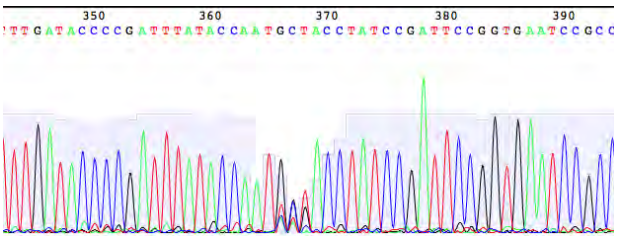


Figure 3. 29. *Example of chromatogram obtained sequencing site saturation position library (Sended by StabVida service)*

Screening of site saturation mutagenesis

E. coli TOP10 cell were transformed with the individual libraries and transformed cells were screened by means of the replica plating assay (*1st screening step*). From this analysis, several transformed cells were selected to move on to the *2nd screening step*. To select mutants in liquid screening analysis (effect on *o*NP release with addition of xylose), activity should accomplish two features: A) rates should be higher than spontaneous hydrolysis; B) *o*NP release should increase with xylose addition. If both conditions occurred, clones were selected as candidates for further sequencing and characterization.

Each library was examined through the two screening steps and those with better characteristics were sequenced (Table 3. 9).

Libraries	# of analysed colonies		Sequenced	Amino acids found as mutations
	1st screening step	2nd screening step		
H420	352	72	8	Leu (1), His (7)
L359*	403	72	12	Ala (1), Glu (2), Gly (4), Pro (3), Leu (2)
V105	379	65	16	Trp (1), Gly (8), Thr (2), Val (5)

Table 3. 9. *Sum up of mutants screened in libraries of site saturation mutagenesis (H420, L359 and V105). Columns: number of colonies screened on the first and second step. Colonies finally selected as a potential candidates and sended to sequencing; and finally amino acid mutations found*

Seven mutants were found in the screening cascade of the three libraries: V105G and V105T in library V105; L359A, L359E, L359G and L359P in library L359, and H420L in library H420.

Library V105: mutants V105G and V105T

Mutants V105G and V105T were analysed: glycosidase (13 mM *o*NPG) and transglycosidase (5 mM *o*NPG, 25 mM xylose) reactions were evaluated. The glycosidase activities of V105T and V105G were similar to the wild type activity, (Table 3. 10).

	Glycosidase specific activity, (Vo/[E],s^{-1})	Transglycosylase reaction (Analytical conditions)	
		Specific activity (*gaxilose* mM/min/U)	T/H ratio
Wild type	205.5 ± 29.9	$(4.99 \pm 0.73)\cdot10^{-3}$	0.1
V105G	172.2 ± 2.9	$(3.92 \pm 1.0)\cdot10^{-3}$	0.08-0.1
V105T	149.9 ± 16.0	$(1.84 \pm 0.25)\cdot10^{-3}$	0.1

Table 3. 10. *Kinetic parameters of glycosidase and transglycosidase activity in analytical conditions. CONDITIONS. Transglycosylase reaction 5 mM oNPG, 25 mM xylose, 0.01-0.1 U of enzyme, (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0, 37ºC).) Producst analysed by HPLC-MS. Glycosidase reaction:* 13 *mM* o*NPG, 0.01-0.1 U of enzyme, 37ºC, pH 7.0 (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$). o*NP released determined by absorbance (410 nm)*

Transglycosidase reaction was also evaluated. Firstly, analytical conditions were used to determine initial rates of transglycosidase (Table 3. 10). V105G and V105T exhibited similar initial rate than wild type and their T/H ratios did not show any significant difference compared to the wild type enzyme.

Longer reaction times were tested under analytical and industrial conditions, (Figure 3. 30). All analysed conditions proved that mutant V105T catalysed gaxilose production in higher yields than wild type (A-industrial conditions; B- analytical conditions with 0,3 U of enzyme; C- analytical conditions with 0,6 U of enzyme (Figure 3. 30)). V105T synthesised 20-30% more gaxilose than wild type under industrial conditions (0.1 g *o*NPG) (Table 3. 11). V105G mutant was also able to reach more or equal gaxilose yield than wild type.

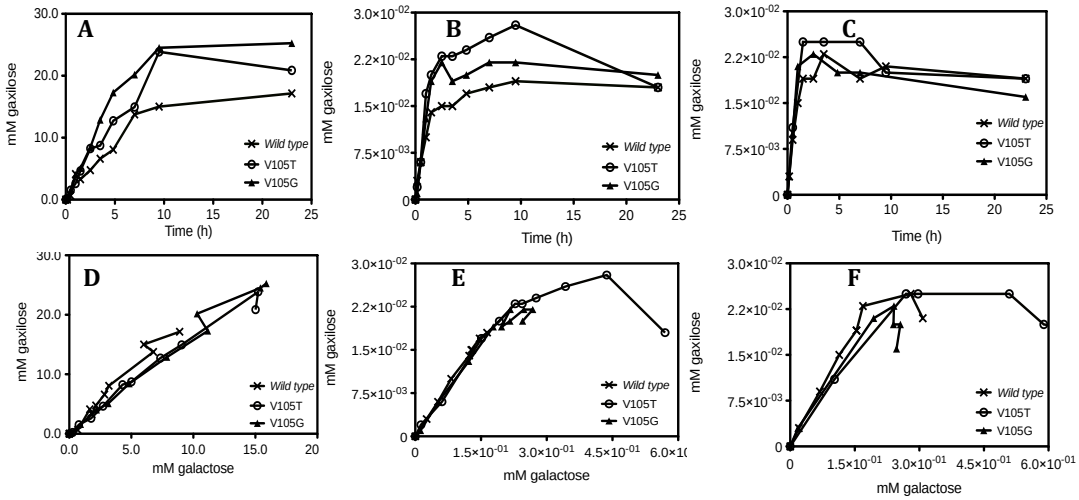


Figure 3. 30. *Kinetics of wild type enzyme and mutants V105T and V105G. A. Gaxilose production under industrial conditions and D its T/H ratio. B, Gaxilose production in analytical conditions and 0.3U of enzyme, and E its T/H ratio. C, Gaxilose production in analytical conditions and 0.6U of enzyme and F its T/H ratio CONDITIONS: 100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0, 37ºC; analytical conditions (5 mM* o*NPG and 25 mM xylose) and industrial conditions (0.1 M* o*NPG and 0.4 M xylose, 1U of enzyme)*

	Industrial conditions
	Maximum gaxilose
Wild type	17 - 19 mM
V105T	22 - 23 mM
V105G	24 - 25 mM

Table 3. 11. *Maximum gaxilose yield under industrial conditions (0.1 M oNPG, 0.4 M xylose) by [redacted] wild type enzyme, V105T and V105G mutants (link to Figure 3. 30)*

T/H ratio and transglycosidase initial rates of V105T and V105G mutants were the same than those of wild type. However, V105T conversions were higher and therefore it achieved higher gaxilose yields. It could be caused by: A) less product inhibition, B) less formation of other products; C) enhanced enzyme stability.

V105T provides a higher hydrophilic catalytic site (opposite to what was intended in the point mutations section above (*3.2.3.1*), where Trp was chosen to increase hydrophobicity and decrease subsite size). Threonine at position 105 could interact better with xylose and improve path for binding. On the other hand, it can be observed that Thr at position 105 could interact with E802, the latter residue belonging to the mobile loop described previously (*i. INTRODUCTION*) (Figure 3. 31). Threonine could establish an H-bond with E802 and the hydrogen bond formation could cause several effects: it could maintain the mobile loop closed (conformation needed for transglycosidase activity) or it would decrease/lose other interactions of E802 residue, such as interaction with hydroxyl C1 of glucose moiety (in our case, would be xylose), as observed in residues of *E. coli* β-galactosidase [165,175,176], or other interactions. Weaker interactions with acceptor could impede xylose reactivity, but it also could cause less interactions with galactose at subsite +1 (product that could inhibit galactosidase activity) and obtain higher conversion yields, as observed. (This could be checked by analysing protein inhibition by galactose). This residue is also in *E. coli* β-galactosidase (E797) and it is described that it interacts with the hydroxyl at C1 of the glucose moiety when the loop is closed and with R599 when it is open [165,175,176]. Changing its H-bond network, the loop could change its mobility and thus change enzyme activity.

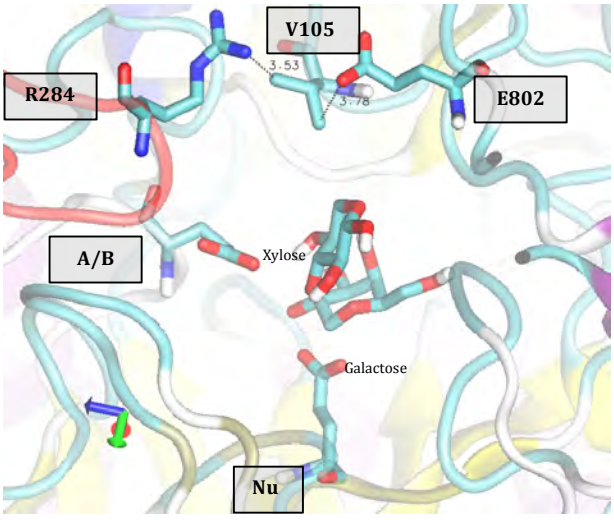


Figure 3. 31. *Catalytic center of the β-galactosidase from [redacted]. Possible interactions of the 105 residue with adjacent residues*

In addition, V105T could also add an interaction with R284 from domain 2 of the adjacent monomer. This interaction could make the catalytic site narrower and improve substrates binding.

Despite all the hypotheses that partly explain the behaviour of V105T, it would be interesting to carry out more experiments to substantiate and understand its features.

Libraries L359 and H420: mutants L359A, L359E, L359G, L359P and H420L

Firstly, specific activities of transglycosidase were determined (Table 3. 12). Transglycosidase initial rates were slightly lower than wild type, and this was also the case for theT/H ratio.

	V_0 (*gaxilose* mM·min^{-1}·U^{-1})	T/H ratio
Wild type	$(4.99 \pm 0.73)\cdot10^{-3}$	0.1
H420L	No data	
L359A	$(6.51 \pm 0.81)\cdot10^{-4}$	0.02
L359E	$(1.70 \pm 0.28)\cdot10^{-3}$	0.05
L359G	No data	
L359P	$(1.25 \pm 0.15)\cdot10^{-3}$	0.02

Table 3. 12. *Initial rates of transglycosidase activity (mM gaxilose/min/U) and T/H ratio of mutants of site saturation libraries. CONDITIONS: 5 mM oNPG, 25 mM xylose,1-8 nM enzyme (0.01-0.1 U), 37ºC, pH 7.0, 100 mM phosphate, 50 mM β-mercaptoetanol, 1 mM $MgCl_2$*

Transglycosylase reactions under industrial conditions were also evaluated (Figure 3. 32). Monitoring gaxilose concentration with time (*left*-Figure 3. 32) showed that L359P was able to synthesise nearly gaxilose with nearly the same yield than wild type. L359A and L359E synthesised half of the amount compared to the wild type, while L359G and H420L were not able to achieve significant yields of gaxilose. T/H ratio analysis (*right*-Figure 3. 32) showed that L359A and L359P mutants displayed nearly the same T/H ratio than the wild type protein. Surprisingly, those mutants showed better ratio than expected despite their gaxilose yield was lower than that of the wild type enzyme. For this reason, those mutants would need to be re-analysed in different conditions of substrate and enzyme amounts.

If L359A and L359P mutants are compared with wild type, the changes of leucine by alanine and proline, respectively, maintain hydrophobicity but the two mutations involve a reduction in size. In addition, proline confers rigidity to the polypeptide chain. In order to understand the effect of these mutations, more kinetics need to be performed to conclude if there is a positive effect or they are neutral mutations.

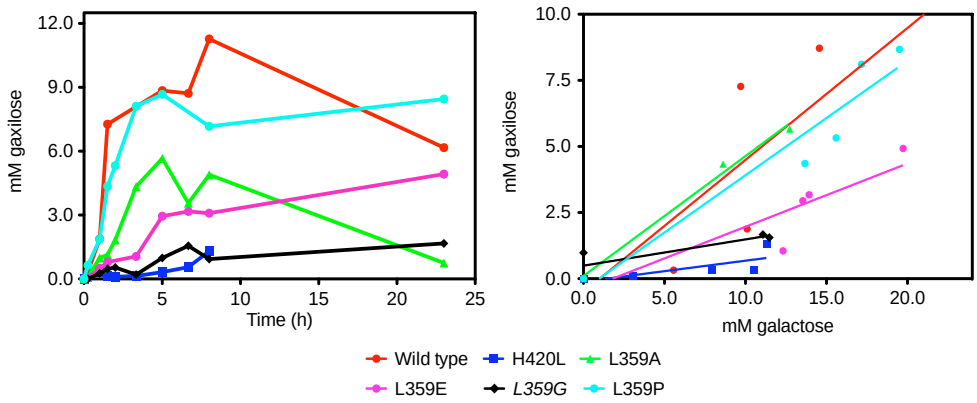


Figure 3. 32. *Gaxilose produced by different mutants (H420L, L359A, L359E, L359G, L359P) of the β-galactosidase from* [redacted] *under industrial conditions.* Left: mM gaxilose *on time;* Right: *ratio T/H. CONDITIONS: CONDITIONS: NPG, 0.4 M xylose, 1U of enzyme,100 mM phosphate, 50 mM β-mercaptoetanol, 1 mM* $MgCl_2$*, pH 7.0 at 37ºC*

3.2.3.2 Combinatorial approaches: bigger libraries

3.2.3.2.1 Libraries

Domain 3 random library

Random strategy was selected to find (unexpected) mutants with improved activity.

Among diverses methods (Box 3. 6), the standard method for random mutagenesis is the error-prone PCR (epPCR) method because of its robustness and simplicity.

Box 3. 6 *Random mutagenized libraries*

Several techniques have been developed and applied to build random mutagenized libraries. These methods can be classified into three main categories: 1) enzyme based methods, 2) chemical mutagenesis and 3) whole cell methods [307,316].

Enzyme based methods include known techniques such as error-prone PCR [125,317,318] and DNA shuffling [319,320]. There are a number of procedures based on PCR techniques such as: *POE-PCR* [321], *MegaWhop* or *Meg Anneal* [322,323], TRINS method or epRCA [324,325], *MCST* [326], SeSaM [327]. Additionally, *in vitro* recombination of DNA techniques (DNA shuffling) are also widely used: RACHIT [320], ITCHY [328], SCRATCH [329], SHIPREC [330], OE-PCR [331], SISDC [332].

The epPCR was the most suitable technique to be applied on the catalytic domain of [redacted] β-galactosidase (domain 3, 310 amino acids) for the sake of library size reduction (if epPCR was applied on whole gene sequence (1029 aa), the library would have been of 20^{2019} mutants; randomized only catalytic site helps to reduce library).

Synthesis of this random mutagenesis library on the catalytic domain was attempted with different methodologies but was only successful using a combination of *epPCR* and *CPEC* (Figure 3. 33).

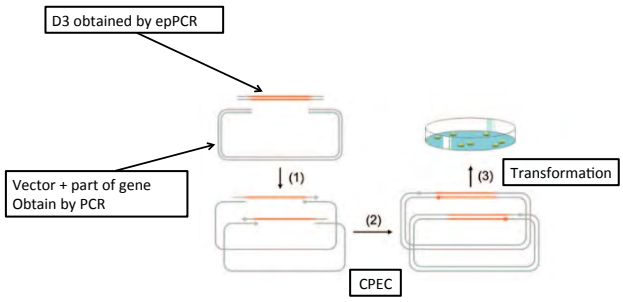


Figure 3. 33. *Diagram of the method used for assemble random library using epPCR and CPEC*

This method is based on the use of epPCR technique to amplify the catalytic domain (and introduced mutations) and its assembly it in a vector and amplification of the rest of the gene by circular polymerase extension cloning (CPEC) [333–336].

Multi Site-Directed mutagenesis

Combination of point mutations (3.2.3.1.3) led to a library of 144 mutants (when this library was built, positions F5603 and N606 were not taken into account). The *multi-site directed* library was composed by the combination of following mutants:

N104L, V105T, L359H, R390K, R390H, H420W, Y505F, Y505Q

Four positions had two possibilities, N104-V105-L359-H420 (mutant and wild type), and two positions had three possibilities, R390-Y505 (two mutants and wild type). Therefore, that library had 144 combinations (2^4 x 3^2). The main goal was to detect if one mutation could affect another mutation and increase transglycosidase activity and/or decrease hydrolysis.

This library was analysed by digital image methodology .

3.2.3.2.2 Digital screening methodology

This part of the thesis was carried out at Prof. Tellier's laboratory (Nantes, France).

Digital screening method developed previously by Prof. Charles Tellier group (*3.*) was applied in diverse libraries: one from *multi site-directed mutagenesis* and two libraries obtained from randomized domain 3.

In first screening step (agar plate screening) twenty colonies were selected as potential candidates (Table 3. 13) but promising mutants were not found in second screening assay.

		# of mutants	
		Multi site	**Domain 3**
# of analysed colonies	**1st screening step**	543 (2 plates)	5,027 (6 plates)
	2nd screening step	20	96
	Sent to IQS	-	24

Table 3. 13. *Number of mutants screened in each step for multi-site library and random library on domain 3*

Domain 3 library was thoroughly analysed. Six plates were analysed with 5,027 colonies spread on them (Table 3. 13). Digital image processing allowed the selection of ninety-six colonies for the second screening step. Liquid screening (2nd screening step) allowed to distinguish twenty-four promising colonies.

HPLC-MS analysis of the transglycosidase activity of these selected colonies is pending.

3.3 Industrial approach

Gaxilose production was implemented in Interquim S.A. after process optimization (*CHAPTER 1*), which included several approaches. Strategy was based on the optimization of reaction conditions using β-galactosidase from *E. coli*, that had allowed a yield of pure gaxilose of 21%.

Nevertheless, the main effort in this thesis has been the design of a new biocatalyst: searching and selecting another β-galactosidase able to improve current enzyme's performance, and the evaluation of this new enzyme. The analyses of optimal reaction conditions (pH, temperature) and evaluation of the enzyme at semi-industrial scale (10 g *o*NPG) allowed to achieve 35% of pure gaxilose. (*CHAPTER 3*). Moreover, protein engineering has been used to modify enzyme and improve transgalactosidase capacity. One mutant version of the selected enzyme achieves higher gaxilose yield under laboratory conditions (*CHAPTER 3*).

This section focuses on the protection of the newly developed technology and speculates about future steps in the project in the short and long term.

3.3.1 Patent application

Box 3. 7. *Previous patents on the project*

Previously, different patents had protected gaxilose synthesis. Spanish patent ES-P-9001680 (1996) describes a chemical method for preparing gaxilose with high number of steps that implies selective protection, glycosylation and deprotection reactions. This procedure had elevated costs and it had difficulties to be carried out on an industrial scale. Later on, Spanish patents ES-9502185 (1997) and ES-P-9701156 (1997) made it possible to obtain, in a single reaction, mixtures of 2-, 3- and 4-*O*-β-D-galactopyranosyl-D-xylose from lactose using transglycosylation activity. However, this process had difficulties from the industrial synthesis point of view.
Gaxilose synthesis procedure was optimized using *E. coli* or *Klyveromyces lactis* β-galactosidase and it was protected by a European patent, EP1408118 (2002). This patent provided a process of gaxilose preparation based on xylose transglycosylation. However, this process required a high number of enzymatic units of β-galactosidase *per* gram of β-galactopyranosyl donor to reach an appropriate efficiency of the reaction.

This project is part of an industrial project of Interquim S.A., in the framework of the scheme "Doctorats Industrials" promoted by the Generalitat de Catalunya (described in *i. INTRODUCTION*).

Previosly in the project different patents had protected gaxilose synthesis (Box 3. 7). The optimization of the industrial process of gaxilose production has been protected by the filling of a patent application "Enzymatic process for the preparation of 4-*O*-β-D-galactopyranosyl-D-xylose" (application number EP17382668.6 , priority date Oct 5th,

2017). For confidentially reasons, the full text of the patent application is not included in the thesis.

CLOSING REMARKS

This chapter includes the characterisation of the β-galactosidase of [redacted] previously selected because of its proficiency in gaxilose production compared to the *E. coli* enzyme. The results presented in this chapter can be summarised:

- Glycosidase specific activity of *E. coli* β-galactosidase was one order of magnitude higher than β-galactosidase from [redacted] Transglycosidase specific activity was 12-fold higher for *E. coli* β-galactosidase than for [redacted] β-galactosidase.
- Analysis of hydrolysis and transglycosidase capacity was also compared according to enzyme Units used (activity on *o*NPG). This comparison showed higher initial rates of transglycosidase for [redacted] β-galactosidase. Additionally, this enzyme showed higher T/H ratio.
- [redacted] β-galactosidase showed a pH optimum for transglycosidation around 6.5. Moreover, reactions performed at pH 6.5 presented the highest T/H ratio.
- Temperature evaluation showed that higher T/H ratio was achieved between 35-40ºC.
- *E. coli* β-galactosidase hydrolyses oNPG 1.3-fold faster than gaxilose and [redacted] β-galactosidase hydrolyses oNPG 8-fold faster than gaxilose.
- *E. coli* β-galactosidase hydrolyses *o*NPG 5.5-fold than [redacted] β-galactosidase, and 38-fold faster gaxilose. It remkars the lower T/H of the *E. coli* enzyme and the higher yields achieved by [redacted] β-galactosidase.
- In semi-industrial level reactions (10 g of oNPG), a yield of 35% was achieved with the [redacted] β-galactosidase, compared to the 21-23% when the *E. coli* enzyme was used. Therefore, a better biocatalyst was developed and thus one of the goals of project was completed.
- The amount of [redacted] β-galactosidase needed for industrial production of gaxilose is over 3-fold lower than that needed for the *E. coli* enzyme (5.5 U/g *o*NPG compared to 18 U/g *o*NPG).
- 3D structure of [redacted] β-galactosidase has been determined at 2.1 Å resolution. Its structure is based on space group $P2_1$ and the unit cell is formed by a dimer of tetramers. Therefore, eight monomers compose the unit cell (each monomer has 1,029 amino acids).
- N606S and V105W maintain gaxilose production with lower T/H ratio. This outcome pointed these positions as non-essential positions but with important effect on activity. Therefore, other amino acid substitutions in such positions could increase activity and/or T/H ratio.

- Screening method was set up for analysing libraries of the β-galactosidase from [redacted] Screening method included two steps: screening in agar plates and a liquid screening.
- Positions V105, H420 and L359 were selected as candidates to analyse the twenty amino acids. Site saturation mutagenesis (NNK libraries) were analysed with screening method set up previously.
- Analysed libraries showed an improved mutant: V105T. This mutant allowed reaching higher gaxilose yield: V105T synthesised 20-32% more gaxilose than wild type, but it did not improve T/H ratio. On the other hand, V105T achieved higher conversion. This feature could be the responsible for the higher gaxilose yield.

DISCUSSION

Glycosidases (GHs) have an intrinsic ability to form glycosidic bonds through transglycosylation (e.g. glucosidases [337–339] , galactosidases [151,221], chitinases [340]). This thesis focused on the use of a β-galactosidase enzyme as a tool for the synthesis of a disaccharide with pharmaceutical interest (gaxilose) and the search for improved biocatalysts.

The challenge of using glycosidases as synthetic tools to catalyzed glycoside bond formation is how to unbalance the native T/H ratio of the enzyme towards transglycosylation in its natural aqueous environment.

The β-galactosidase from [redacted] a good biocatalyst for the synthesis of gaxilose.

In this work, a new enzyme with a better performance than the current enzyme used in gaxilose production was found: the β-galactosidase from [redacted] It had been previously described that this enzyme had some transglycosidase activity with several substrates [230,231,291], albeit the products of those reactions were not characterised. The [redacted] β-galactosidase, from GH2 (CAZy), has high similarity with the currently used β-galactosidase from *Escherichia coli*. Both enzymes share overall 65% sequence similarity, and up to 85% in the catalytic domain. Both enzymes present the same domain organization and equal quaternary structure. Although the [redacted] β-galactosidase had not been so deeply studied as the *E. coli* β-galactosidase, their high degree of similarity allowed us to compare some characteristics: e.g. the transglycosidase capacity for allolactose synthesis. Intramolecular allolactose production is not a universal feature of β-galactosidases, but two characteristics would indicate that the [redacted] enzyme has it:

(1) The [redacted] β-galactosidase also has the allolactose synthesis motif residues described by Wheatley and co-workers [176] (KREVSEATRIDPNA, residues from the loop and residues involved in the loop movement [165,168,172]). Intramolecular allolactose synthesis requires the loop (from domain 5 [172]) as well as the residues involved in its opening and closing (residues that contribute to bind and orient the glucose moiety and stabilize the transition state) [172,341]. The rate of allolactose synthesis can be significant if glucose remains bound long enough for the isomerization reaction to take place [342].

(2) The [redacted] microorganism also contains the *lac* repressor [176].

Both features indicated the ability of the [redacted] enzyme to synthesise allolactose, and therefore the potential of exhibiting transglycosidase activity with other

substrates, as it has been the case. However, it was unpredictable that transglycosidation would be higher than that of the *E. coli* enzyme.

In this work, it has been shown that the [redacted] enzyme has a higher transglycosylation capacity (higher T/H), when compared to the *E. coli* enzyme under the same conditions, and it is able to synthesise gaxilose in higher yield with less units of enzyme. But why would this enzyme synthesise more gaxilose if it is so similar to the *E. coli* β-galactosidase? The answers can be found in small divergences between enzymes, which some of them have been identified (e.g. H357 from *E. coli* and L359 from [redacted] deeply analysed later).

[redacted]. The distinct niches of *E. coli* and [redacted] could have influenced differently the evolution of both β-galactosidase allolactose-depending expression systems and therefore change the enzyme activity. It could be hypothesised that higher transglycosidase activity of the [redacted] β-galactosidase, that leads to a higher amount of allolactose, could be due to (1) the need of to produce allolactose (the operon natural inducer) in an environment where lactose is not abundant, i.e. the [redacted] β-galactosidase produces same amount of allolactose than the *E. coli* enzyme with less amount of lactose. It is tempting to speculate that [redacted] B5 habitat may have less amount of lactose (if compared with intestinal habitat of *E. coli*). In that condition, the lower hydrolytic activity of [redacted] would not render useful compounds as carbon sources, as it would be the case for the *E. coli* enzyme (hydrolysis of lactose would render glucose and galactose). Conversely, higher transglycosilation would allow the bacterium to build up more complex carbohydrates that may be used as carbon and energy sources. In this sense, [redacted] may use its β-galactosidase with anabolic purposes, profiting from the rich plethora of catabolic compounds generated by other microscopic prokaryotes or eukaryotes (i.e. algae, fungi, yeasts) in the ecologic niche. Another possibilities could be that (2) the β-

galactosidase needs to produce more inducer because the repressors of *E. coli* and [redacted] are different and the one from [redacted] requires more allolactose to be deactivated or (3) more inducer is produced because more protein is needed due to the [redacted] enzyme slower velocity (hydrolase activity *per* mg of protein) compared with the *E. coli* enzyme.

E. coli and [redacted] enzymes exhibit transglycosidation for allolactose synthesis at low lactose concentration, needed for β–galactosidase expression. Their high transglycosidase activities in nature make them good candidates for being employed as synthetic tools. Search for enzymes with the *need* of transglycosylation would be a good/feasible feature for the search of new biocatalysts.

Protein engineering of the β-galactosidase from [redacted]

The use of GHs for glycoside bond formation is linked to their intrinsic hydrolase activity, as previously mentioned. Different strategies to modify hydrolase activity and increase the T/H ratio have been approached by protein engineering. One of the most successful and general strategies to remove hydrolysis is the so called glycosynthase technology, developed by Planas and co-workers (1998) [114] at the Biochemistry laboratory of IQS School of Engineering in Barcelona, and by the Withers' team in the Chemistry Department at the University of British Columbia (UBC) in Vancouver (1998) [113]. This approach uses highly reactive glycoside donors and hydrolytically inactive mutant enzymes (mutation of the catalytic nucleophile to a non-nucleophilic amino acid, see INTRODUCTION). There are numerous examples of success in the literature following this strategy ([113–116,118–121,346–349]). Nevertheless, this approach could not be applied to the synthesis of gaxilose due to the difficulty in obtaining activated glycosyl donors (e.g. glycosyl fluorides) at appropriate scale and price for industrial production.

Beyond the glycosynthase technology, there are no general rules to increase transglycosylation in GHs (increasing transglycosidase activity and/or diminishing hydrolysis). Structural information does not give clear clues for understanding the effect of mutation on the T/H ratio. Highly conserved amino acids near the active site are involved in stabilization of the transition states (TS) and most mutations will have an unpredictable effect since they will impact on both the glycosylation and deglycosylation steps [124,350–352]. However, there are some strategies that have been successfully implemented [157,190,290]. These approaches are mainly based on (1) engineering the negative subsites of the enzyme with the aim of affecting donor binding

and transition state stabilization [129,289,351,353–356], (2) engineering the positive subsites of the enzyme to introduce new or improved interactions with the acceptor substrate with the aim of increasing acceptor affinity [129,131,272,310,357,358] and (3) modulating the access of water to the active site to reduce hydrolysis [134,136,359–361]. Relevant examples supporting these concepts are discussed in the following paragraphs.

With the goal of evolving an α-L-arabinofuranosidase into a transglycosylase, Fauré and co-workers performed an iterative protein engineering approach [129]. First, random mutagenesis starting from a previously reported single mutant (L352M) known to have a moderate transglycosylation yield (18%), a point mutation (R69H) at a position highly conserved in clan GH-A was obtained. It displayed slightly increased transglycosylation activity and severely reduced hydrolase activity. Then, other mutations were sequentially introduced using semi-rational and *in silico* approaches. Two positions on positive subsites (+1 and +2) were selected with the goal of introducing "new or improved interactions with the acceptor substrate", and a combinatorial library was prepared. Remarkably, a triple mutant with 80% transglycosylation yield was obtained. The effect of each mutation was rationalized by careful kinetic analyses, concluding that the T/H ratio can be modified by generating a hydrolytically crippled mutant that can be achieved by directly mutating residues that are likely to form hydrogen bonds with the catalytic nucleophile in the substrate donor site and/or highly conserved around the donor subsite (−1). These could be combined with direct mutation(s) that increase the affinity or the positioning of the acceptor, as predicted at hot-spots using *in silico* approaches.

Another example of transglycosidase improvement was performed by Tellier and Dion's team. Their work was focused on the β-glycosidase of *Thermus thermophilus*. First, one step of random mutagenesis and *in vitro* recombination allowed to reduce primary and secondary hydrolysis, and therefore transglycosylation could reach yields of 60-75% [130,310]. The found mutations (F401S and N282T) were located just opposite subsite -1 and they suggested that a repositioning in subsite -1 together with a better fit of the acceptor in subsite +1 might favour the attack of the acceptor at the expense of water. Later on, they identified four mutants able to increase the T/H ratio selecting conserved amino acids around the active site [133]. One mutant stabilized transition state and supported the excess of positive charge by adding a hydroxyl group (Y284F). The other three mutants (R75A, W120C and N163A) belonged to a network of residues in close contact with the catalytic residues. Mutations of these positions induced lower stabilization of the transition states of the reaction, which affected hydrolysis more

than transglycosylation. This work showed how subsite +1 modification and catalytic network alteration could improve transglycosylation.

Other examples exist in the bibliography that use a single mutation to increase transglycosidation. Random mutagenesis was used in the β-galactosidase from *Geobacillus stearothermophilus* and a single mutant (R390K) with increase synthetic capacity was found [289]. The mutated residue disturbed the interaction network with the galactose moiety and change of the hydrolytic capacity of the enzyme. Substrate binding modification was also used to modify a β-glucosidase from *Pyroccoccus furiosus* [359] (the mutation, F426Y, decreases the rates of the hydrolytic reaction while the rates of transglycosylation reactions were less influenced).

In this work, some of these strategies have been used to design the mutants. Rational design on identified hot spots has been combined with site saturation mutagenesis to find improved mutants.

1) Impairing TS stabilization by mutations at -1 subsite

Four positions located in subsite -1 (L359, R390, Y505 and N606) were mutated expecting an effect on the first transition state of the reaction (glycosidation step), (1). Mutation of L359 and N606 may alter substrate interactions, while mutations of R390 and Y505 may have an effect on the catalytic nucleophile reactivity.

Position **359** is highly conserved on β-galactosidases from family 2. The sequences alignment of 75 characterised β-galactosidases from family GH2 [3] reveals that histidine is a universally conserved amino acid in this position (*Appendix A.4.1)*. Structural studies of the β-galactosidase from *E. coli* with transition state analogue inhibitors showed that the nitrogen of that histidine (H357) interacts, via water molecules, with the C3 hydroxyl of the transition state analogue [168,362]. These data indicate that H357 interacts weakly with the ground state. The [redacted] β-galactosidase contains a leucine instead of a histidine in this position, therefore L359 (non-polar residue) can not form H-bonds interactions with galactose moiety trough water molecules (Figure D. 1). Moreover, this change could modify substrate network

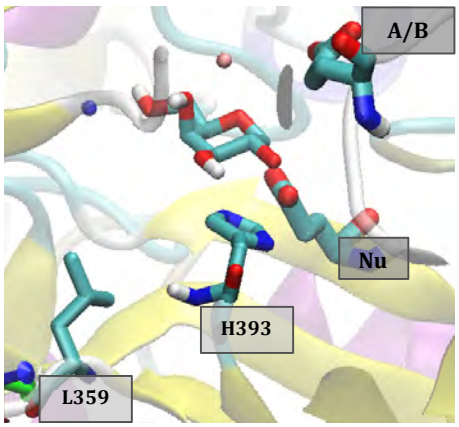


Figure D. 1 *Catalytic centre of β-galatcosidase from [redacted] It is shown the catalytic resid[redacted]59 and H393 residues involved in galactosemoiety binding (galactose bound)*

[3] 75 β-galactosidases characterised from family GH2 were aligned used Clustal program (Jalview)

(e.g. alter the interaction of the H393 residue, equivalent H391 in *E. coli* β-galactosidase, which is described to stabilize the transition state, primarily affecting the galactosylation step [363]). This change could modify transition state energy and favor transglycosylation over hydrolysis (modify substrate network has shown to improve transglycosidase in several studies [357,364,365]), which is a hypothesis that seems to be proved by our experimental data showing the higher T/H rate of the [redacted] enzyme. Moreover, the change in substrates' interactions has also an effect on the hydrolysis of the product too [redacted] enzyme hydrolyses 36-fold less gaxilose than *E. coli* enzyme). The study of the L359H mutant allowed to determine that the T/H ratio was reverted and the mutant behaves as the *E. coli* enzyme. Thereby, this position has a clear effect on T/H ratio. The screening of a site saturation library on this position ended up with the selection of four candidates: Ala, Glu, Gly and Pro. Altough clear results were not obtained, L359P and L359A seem to have a neutral effect on the enzyme (which could be explained because proline and alanine substitutions maintained a non polar and neutral side chain, probably needed to maintain hydrophobicity and to not contribute on the substrate network).

R390 is also a conserved residue in β-galactosidases from family 2 (*Appendix A.4.2)*. It forms direct interactions with the nucleophile residue (E539) and other conserved residues, such as H393 and N462 (also highly conserved [168]) (Figure D. 2). The equivalent asparagine to N462 in *E. coli* β-galactosidase (N460 in the *E. coli* enzyme) has been described to stabilize the transition state while destabilizing the bound substrate [366]. Wheatley and co-workers showed that N460 interacts with the O3 hydroxyl of the galactosyl moiety in the shallow mode and with the O2 hydroxyl in the deep mode, contributing to stabilize the TS. It is well known that H-bond interactions between the 2-hydroxy group and the catalytic nucleophile stabilizes the development of positive charge at the oxocarbenium-like TS [124,352,367–370]. Diverse studies show that mutated residues from the first and second shell of the subsite -1 can cause a reorientation of the amino acids that are in contact with the substrate and result in a better docking energy (α-L-fucosidase from *Thermotoga maritima* [365] or α-L-fucosidase from *Bifidobacterium longum* subsp. *Infantis* [364]). In this work, R390K and R390H mutants

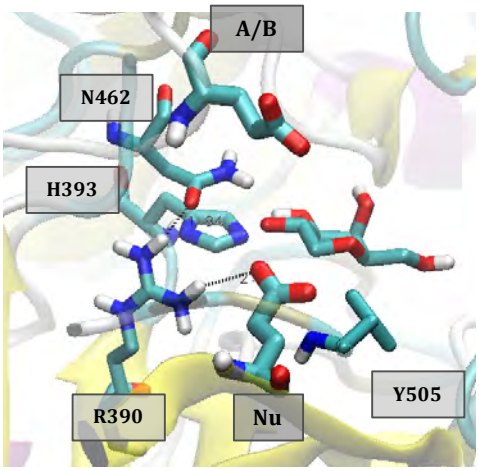


Figure D. 2 *Catalytic centre of β-galatcosidase from* [redacted] *with residues involved in gaac* [redacted] *g: N462, H392, R390, Y505 and catlaytic residues (galactose bound)*

reduced hydrolase and but also transglycosidase activity relative to the wild type enzyme, suggesting that the function of the nucleophile residue has been impaired by the local changes.

The **Y505** residue is also highly conserved in all the analysed β-galactosidases from family GH2 (*Appendix A.4.3*) (Figure D. 3). Mutation of the equivalent tyrosine in the β-galactosidase from *E. coli* resulted in a large decrease of the hydrolase activity [371–373]. Later on, it was proven that this residue may be important for the degalactosylation step for its ability to transfer protons and to facilitate cleavage of the transient covalent bond between galactose and the nucleophile residue [374]. The tyrosine residue between subsite -1 and +1 has shown to have an impact on T/H ratio on other enzymes, e.g. mutation F401S on β–glycosidase from *Thermus thermophilus* (located opposite subsite -1) induced a significant shift of the product within the binding site at subsite +1 and it favored transglycosylation (60-75% yield) [353]. Amino acid mutations near the catalytic residues can change active site electrostatics that can lead to a modification of the T/H ratio and increase transglycosylation as shown in a chitinase form *Serratia marcescens* (D142N) [357]. In this work, the Y505F and Y505Q mutants showed a reduction of glycosidase activity of about 600-fold and 2,000-fold, respectively. But this reduction in hydrolase activity was not reflected by an increase of the T/H ratio since Y505F exhibited very low transglycosidase activity whereas the Y505Q mutant completely lost detectable transglycosylation activity. In the light of the results, Y505Q and Y505F mutations may alter subsite -1 interaction (1) diminishing hydrolase activity but also the transglycosidation, without increasing T/H ratio.

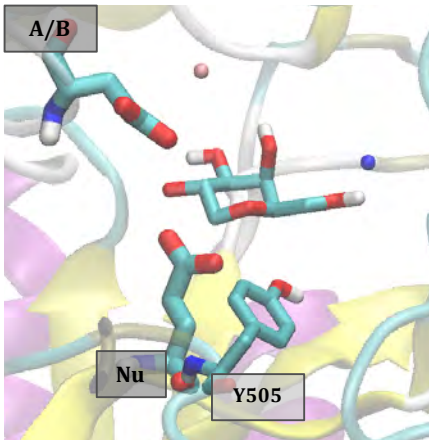


Figure D. 3 *Catalytic centre of β-galatcosidase from [redacted] representation of ca[redacted]lues and Y505 amino acid (galactose bound)*

The **N606** residue shows some sequence divergences among GH2 β-galactosidases, with Asp at the equivalent position in some cases (*Appendix A.4.1*) (Figure D. 4). The *E. coli* β-galactosidase also has an asparagine residue (N604) and it was shown that this residue is involved in the coordination of the Na^+ ion [366]. This monovalent cation is coordinated with O6 hydroxyl of galactose (or a water molecule when substrate is not bound) [165,167]; Na^+ affects the affinity of substrates, the stability of the transition states and the stability of the covalent intermediate [375]. Another example is *Thermotoga maritima* β-galactosidase, used to synthesize galacto-oligosaccharides. This enzyme

also has an asparagine (N574) at the same position interacting with the galactosyl moiety. The *T. maritima* β-galactosidase was studied and it was hypothesized that a disruption of the hydrogen bond between galactose and the asparagine would cause a decrease in activity due to loss of stabilizing interactions that would diminish the affinity of the enzyme to the substrate [157]. Even though asparagine mutants showed a decrease of glycosidase activity, the authors detected an increase of transglycosidase activity (around 30%-40%). They suggested that it was probably due to a higher rotational freedom of the galactosyl moiety in the covalent complex. In the present work, the N606S mutant reduced glycosidase activity more than 10-fold under analytic conditions. On the other hand, the N606S transglycosidase activity and the T/H ratio under industrial conditions showed similar activity as the wild type enzyme. The asparagine to serine mutation could maintain similar interactions and therefore be able to maintain the same T/H ratio than the wild type enzyme and only show a decrease on glycosidase activity. More experiments would be needed (e.g. modify enzyme concentration, kinetic parameters) to deeply understand this mutant.

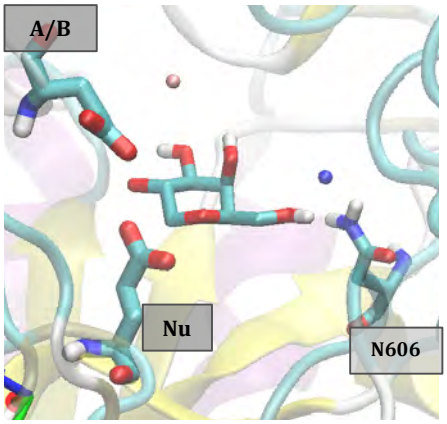


Figure D. 4 *Catalytic centre of β-galatcosidase from [redacted] and role of N606 on galactose binding (galactose bound). Catalytic residues are also represented*

2) Improving acceptor binding by at +1 subsite

N104, V105, H420 and F603 positions were selected for mutagenesis with the aim of improving acceptor binding (2) (residues H420 and F603) or modifying subsite entrance path (3) (residues N104 and V105).

The **N104** residue is not highly conserved among the GH2 family (*Appendix A.4.5*). It is located in a loop in domain 1 and it is part of the catalytic cavity. In this work, this residue was mutated to leucine. Addition of hydrophobic amino acids to decrease the access of water to the glycosyl-enzyme intermediate had been performed in other studies to decrease hydrolase activity, e.g. a mutation to modify the positioning of the nucleophilic water molecule position and thereby increase the possibility of other substrates to act as acceptors (H497M), as in the case of an α-galactosidase from *Bifidobacterium adolescentis* [361]. In this work, the N104L substitution reduced the glycosidase activity by more than 600-fold, but no transglycosidase activity was detected. The N104L mutant could disrupt needed interactions (H-bond) with (a)

acceptor molecule or (b) other amino acids such as the S801 residue (amino acid from the mobile loop that interacts with N104 when the loop is closed).

The **V105** position in β-galactosidases varies among different amino acids (*Appendix A.4.5*); most of them have a valine, a threonine or an isoleucine. This residue is located at the entrance path of the catalytic site and V105 does not interact directly with bound substrates (from docking analysis). Mutations located in the outer regions of the catalytic site have shown to lower the rates of hydrolysis reactions, while maintaining primary catalytic function [376]. Moreover, adding hydrophobicity could reduce hydrolysis and increase transglycosylation, as performed in other enzymes (e.g. neopullanase from *Bacillus stearothermophilus* improved the transglycosylation activity by replacing amino acids with more hydrophobic residues [134]). In this work, the V105W mutant maintained glycosidase and transglycosidase activity at similar levels of those of the wild type enzyme, however the T/H ratio was lower. This position seemed important for T/H ratio but not essential for catalysis. An analysis of a site saturation library on the 105 position found a new promising mutant: V105T. This mutant was able to synthesise gaxilose with higher yield (approx. 20% higher) without affecting the T/H ratio. V105T achieved higher conversions and therefore gaxilose yields were higher. This improvement could be caused by different reasons: less product inhibition (some β-galactosidases are inhibited by galactose [377–379]), less by-products ([220,380–382]), or less product hydrolysis or enhancement of enzyme stability. V105T could alter the network of binding interactions with substrate at subsite +1 (e.g. reorganitzation of the substrate network by modifying residues on the second shell can improve transglycosylation, as shown for a chitinase mutant from *Serratia marcescens* (W167A) [131]). Threonine hydroxyl group could add some new interactions, e.g. with E802. This new H-bond could (Figure D. 5):

1) Favor the closed conformation of the loop and it could help substrate movement to the shallow binding mode (needed for activity). The loop mainly in the closed conformation did not increase transglycosidase activity on *E. coli* β-galactosidase (see *1.4.1*), however soft

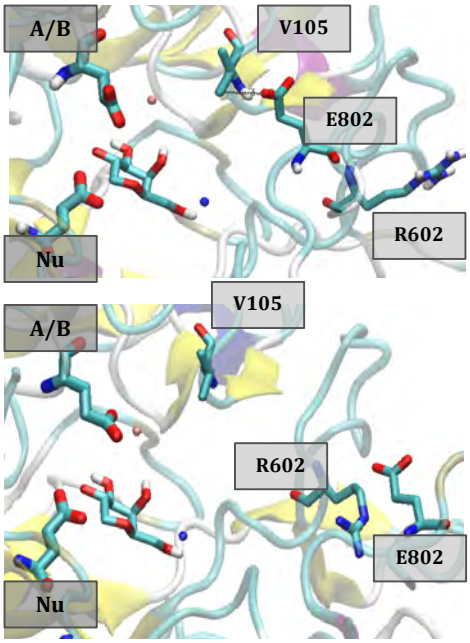


Figure D. 5 *Catalytic centre of β-galatcosidase from [redacted] and residues interacting with V105 when loop is closed (up) or when loop is open*

stabilization of the closed loop may help to increase transglycosidation.

2) Another possibility is that this new interaction modifies E802 network. It is described that the E802 equivalent residue in the *E. coli* β-galactosidase (E798) interacts with an arginine (R599) when the loop is open (weak electrostatic interactions) [165,175] and with glucose when loop is closed [366]. If E802 interacts with threonine of position 105, it could change interactions with the substrate (affecting xylose binding, galactose inhibition or gaxilose binding) and modify the transition state.

The **H420** residue is present in many β-galactosidases from family 2 (*Appendix A.4.2*). It has been described that the equivalent histidine in the *E. coli* β-galactosidase is important for acceptor binding: it forms water-bridged interactions with glucose C2 and C3 hydroxyl groups via water molecules [176]. This histidine also interacts with Mg^{2+} and together modulate the central role of the acid/base catalytic residue [383]. The importance of aromatic residues in determining the T/H ratio had been observed in several GHs: e.g. Lundemo and co-workers showed an improvement of transglycosylation by addition of a third platform for aglycone binding site (N220F) and G222F) [384]. However, in the current work, the H420W mutant showed a strong reduction of glycosidase activity and a suppression of transglycosidase activity. This loss of activity could be due to a distortion in the subsite cavity caused by a size modification (histidine to tryptophan) and a loss of essential interactions. Screening of a site saturation library in position 420 ended up with the selection of the H420L mutant. Leucine amino acid would modify interactions with the acceptor moiety and it could change the TS, as shown in other enzymes [364,365]: Matsui and co-workers had shown that mutations to non-aromatics residues could cause slight changes in the binding topology of substrates to favor transglycosylation over hydrolysis [355]. This mutant deserves further thorough analyses.

F603 is quite conserved among GH2 β-galactosidases (*Appendix A.4.4*), and it is reported to have relevant roles in *shallow* and *deep* binding modes. *E. coli* β-galactosidase studies have described that this residue interacts with a serine from the mobile loop and with C6 of the galactosyl moiety in the shallow binding mode [175]. In deep mode, it interacts with an arginine (599 for *E. coli* and 601 for the [redacted] β-galactosidase) via the π electrons of a phenylalanine and the guanidinium cation of the arginine; all movements are guided by a methionine (M542 for *E. coli* and M544 for [redacted] β-galactosidase) [177,178]. This position has also been studied for the β-galactosidase from *Thermotoga maritima* [157], where a substitution of this position (F571L) improved transglycosylation (around 40%). Docking and modeling studies

suggested that phenylalanine to leucine substitution may allow the rotation of a tryptophan, which is a platform for binding the acceptor and highly conserved in GHs: they suggested that higher flexibility could be favoring acceptor positioning. It has been revealed that modifying aromatic residues in the active center could modify substrate binding topology and increase transglycosylation [355,385]. In this work, F603L showed a reduction of glycosidase activity of more than 10-fold, as well as a decrease of the transglycosidase activity and the T/H ratio too. The F603L mutant would be losing important interactions needed for catalysis, as well as substrate interactions needed to help the substrate to move from shallow to deep binding mode [168].

Difficulties in finding mutants with improved transglycosidase activity: the screening method

Successful results have shown that highly conserved amino acids around the active site are good targets for obtaining very efficient mutants for transglycosidation ([129,136,354,356]). However, in this work, the most interesting result has been a mutation of a non-conserved position located on the outer region of the catalytic site: V105 (a residue selected by docking analysis). The combination of site directed mutagenesis and site saturation mutagenesis has allowed obtaining an improved mutant (V105T). V105T achieves higher conversions yields, although more experiments would need to be performed in order to understand the structural and mechanistic reasons for of this mutant's behavior.

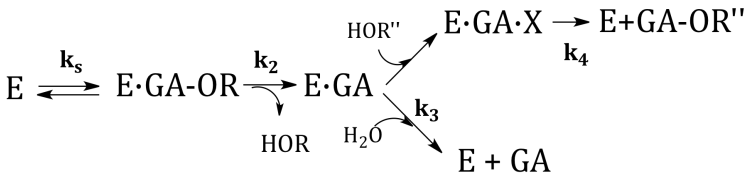
The screening method

The screening methodology set up in this work has been previously used in several works [39,43,46] and it is based on the search of enzymes with a high T/H ratio. In the current work, it has been applied to three libraries of site saturation mutagenesis and one of random mutagenesis. Site saturation libraries were analysed and one mutant was found to exhibit higher gaxilose yields than the wild type enzyme: V105T. This mutant presented higher gaxilose yields but it did not increase the T/H ratio. So, why was it selected in the screening procedures?

A) The site saturation library in position 105 may have not any mutant with a higher T/H ratio than the wild type enzyme. Therefore, colonies selected were those with similar T/H ratio than the wild type enzyme. For future assays, an increase of xylose amount could be employed in order to increase inhibition and not confuse it with wild type phenotype.

B) Other possible explanation falls on the rate limiting step. Although in this work it has not been analysed, some hypothesis can be proposed.

- If k_2 is the rate-limiting step and $k_4 > k_3$, the reaction rate (determined by aglycone release, *o*NP or 5-5'-dibromo-4,4'-dichloro-indigo) would be independent of the acceptor concentration [57].
- If k_3 or k_4 are rate-limiting, the reaction rate would increase as a function of acceptor.



It has been discussed, for the β-galactosidase from *E. coli,* that galactosylation step could be the rate limiting step for leaving groups with pK_a values above 8, and as the *o*NPG pK_a is 7.2, degalactosylation step would be the rate limiting [32]. However, it is not known which is the rate limiting step for [redacted] (moreover, it has to be taken into account that the main difference between *E. coli* and [redacted] enzymes is on subsite -1, L359 an H357, and interactions with donor moiety, directly affecting galactosylation). Therefore, it would be interesting analyse this feature in further experiments to clarify if the screening method used for the [redacted] β-galactosidase was searching for mutants with increased T/H ratio or mutants with high transglycosidase activity (if the transglycosylation is high enough, the enzyme could recover inhibition effect with the addition of xylose, but it would not increase the rate achieved without xylose: i.e. mutants with low transglycosidase activity would show a higher inhibition, than those with transglycosylation). Both cases would allow to find interesting mutants, but second case would have false mutants (wild type). One solution would be increase xylose concentration to increase inhibition effect; therefore, only mutants with higher transglycosidase activity than the wild type enzyme would be selected because they would show "less inhibition", but not an increase.

Modifying the enzyme

The wild type [redacted] enzyme has already transglycosidase activity for allolactose synthesis, and therefore it seems that the catalytic center has been already optimized by nature for transglycosidase activity. Modifications on the second shell and the outer regions (L359, V105) seem good choices for the improvement of its transglycosidase activity, e.g. on the role of the mobile loop. For this reason, generation and analysis of

mutations were also attempted in this domain. Nevertheless, the large size of the enzyme (1024 amino acids) makes the screening of such a big number of mutations very challenging. The screening in solid medium that was implemented in this work, albeit proving to be useful for the generation of mutants within domain 3, is still a time-consuming and manual method. Further improvements in the method for colonies spreading and identification and more precise and potent software for image analyses may be paramount to be able to screen and detect mutants with improved transglycosylation ability.

An analytical method allowing regioselective product quantification

The analytical method developed in this work allowed to analyse samples in less than ten minutes and without any previous sample treatment. This short time allowed a high to medium throughput, making the method compatible with the analyses of a large number of samples, especially because the instrument used allowed automatic sampling and autonomous overnight use. This feature is especially relevant when long-time reactions were carried out (kinetic analyses). Regarding the simplicity of the sample preparation, in comparison to those methods that need tedious, time-consuming sample manipulations (e.g. derivatization for gas-liquid chromatography analysis and complete drying of samples [181,183]) the implemented method was very convenient. The LoQ, achieved is valid for analysed samples (concentrations of substrate and products are quite high). One of the fundamental pros of the method is that it allows the identification of the three isomers in the same analysis, allowing the calculation of transformation and regioselectivity concomitantly.

Oligosaccharides are often analysed by High-Performance Anion Exchange Chromatography with pulsed amperometric detection (HPAEC-PAD) [129,235,242,276,386,387]. In this work, HPLC-MS allowed to quantify hydrolase and transglycosidase activities at the same time, in a very short period of time, and therefore determine T/H ratio easily. The HPLC-MS method developed is suitable for the purpose of the work, which is to evaluate gaxilose synthesis and determine the T/H ratio. Despite the efforts made in order to include *o*NP and *o*NPG detection in the same analysis, no conditions were found so that this product could be quantified in the same method. It is worth mentioning that the synthesis of other oligosaccharides by-products catalyzed by [redacted] β-galactosidase cannot be ruled out and that our method would not detect compounds with masses different than that of gaxilose (e.g. disaccharides such as Gal-Gal or trisaccharides such as Gal-Gal-Xyl).

The industrial production of gaxilose

The process for the industrial production of gaxilose has been improved compared to the procedure described in EP1408118 patent. The developed process uses higher concentrations of donor and acceptor (4.5-fold and 2-fold, respectively) in 4-fold less volume than the original method. This feature is important from the industrial point of view because it reduces the need for very large reactors and thus increases reactor productivity. The reduction of reaction volumes, due to a higher reaction concentration, makes downstream processing of gaxilose less energy demanding because less amount of aqueous solution needs to be evaporated obtain the final dried product. Consequently, gaxilose production is a more environmentally friendly process.

E. coli β-galactosidase has been extensively studied at industrial level in this project. On the other hand, β-galactosidase from [redacted] was only analysed at semi-industrial level (10 g and 50 g *o*NPG) when this thesis was composed. Project progress has focused in [redacted] β-galactosidase evaluation at industrial level. This enzyme was to be produced from biomass grown at 200 L reactor scale.

The amount of enzyme needed in the industrial production of gaxilosa, β-galactosidase from [redacted] is more efficient than the *E. coli* enzyme because using 3.3-fold less amount of enzymatic units per gram of *o*NPG, it produces 1.5-fold more gaxilose.

	E. coli	[redacted]
U/g o*NPG*	18	5.5
Gaxilose yield %	22	35.4

Table D. 1 *Improvements of galactosidase from* [redacted] *and compare with* E. coli *galactosidase. Comparison of U* per *g of* o*NPG and gaxilose yield achieved*

The increase in the yield of the product (from 20% to 35 %) has a big impact in the cost of the productive process of gaxilose. Regioselectivity was increased in more than 30% (85:15 for *E. coli* enzyme to 95:5 for [redacted] enzyme, (β-1,4: (β-1,3+ β-1,2)), making it a more selective enzyme for the production of gaxilose and also enhancing the purification of the desired isomer from the reactions mixture. This Additionally, the stability of gaxilose is higher in the reaction medium because is not the preferred substrate for the hydrolytic activity of the [redacted] β-galactosidase. This feature allows diminishing the risk of product hydrolysis if reaction is faster than expected or check points analyses in the process control are somehow delayed. Altogether, it results in a more efficient methodology for industrial scalability.

To further reduce the cost of goods in the biocatalytic reacion, the immobilization of the enzyme was pursued in order to be able to reuse it in consecutive reaction rounds. Unfortunately, this endeavor was not successful.

CONCLUSIONS

1. An HPLC-MS method for the quantitative analyses of transglycosylation reaction products using oNPG and xylose as donor and acceptor, respectively, has been developed. It allows the quantification of 2-*O*-β-D-galactopyranosyl-D-xylose, 3-*O*-β-D-galactopyranosyl-D-xylose, 4-*O*-β-D-galactopyranosyl-D-xylose (or gaxilose) and galactose.
2. The optimal conditions for the industrial production of gaxilose using *Escherichia coli* β-galactosidase have been determined. At the industrial scale (i.e. 12 kg of *o*NPG, 19 kg of xylose and 216 kU of enzyme in a 180 L reaction), 2.61 kg of pure gaxilose, have been obtained.
3. In a quest for a better biocatalyst, and using the CAZy database, a total of 123 classified and characterised as β-galactosidases enzymes were selected. Among them, 45 β-galactosidases with reported transglycosylation activity were further screened. Bibliographic search and phylogenetic trees analyses allowed final selection of nine enzymes to be expressed, purified and characterised: *Bacillus circulans, Bifidobacterium longum subsp. infantis,* [redacted] *Geobacillus kaustophilus, Lactobacillus delbrueckii subsp. bulgaricus, Paenibacillus thiaminolyticus, Saccharopolyspora rectivirgula, Sulfolobus solfataricus* and *Thermotoga neapolitana. E. coli* β-galactosidase was included in all the enzyme's evaluation as a standard: it is considered the threshold for the new enzyme selection.
4. Under laboratory conditions, six candidates were discarded (i.e. *B. circulans, B. longum subsp. infantis, G. kaustophilus, P. thiaminolyticus, S. solfataricuss* and *T. neapolitana*). The other three candidates, i.e. [redacted], *L. delbrueckii subsp. Bulgaricus* and *S. rectivirugula*, were further screened under analytical and industrial conditions. [redacted] β-galactosidase was selected as the new biocatalyst.
5. Characterization of [redacted] β-galactosidase has shown that the enzyme has high T/H ratios and little secondary hydrolysis which makes it a better biocatalyst for the synthesis of gaxilose compared to the *E. coli* enzyme. Temperature and pH effect on [redacted] β-galactosidase showed that, under laboratory conditions, 37ºC and pH 6.5 were the optimal conditions.
6. In an attempt to increase the ability of [redacted] β-galactosidase as biocatalyst for the synthesis of gaxilose, eight point mutations were analysed. After that, three different libraries of mutants were constructed and the appropriate screening methods were set up. An improved mutant has been found: V105T.

This mutant achieves higher conversion rates and reaches higher gaxilose yields (20-32% more gaxilose than wild type) with the same T/H ratio.

7. 3D structure of [redacted] β-galactosidase has been determined at 2.1Å resolution. Its structure was based on space group $P2_1$ and the unit cell was formed by a dimer of tetramers.
8. In semi-industrial level reactions (10 g of oNPG), a yield of 35% of pure gaxilose has been achieved with the [redacted] β-galactosidase, compared with the 21-23% when the *E. coli* enzyme is used. Moreover, [redacted] β-galactosidase reduces 3-fold the amount of enzyme Units needed for industrial production Therefore, a better biocatalyst has been developed.

METHODS AND MATERIALS

ii. CHAPTER 1

ii.1.1 Protein expression and purification

ii.1.1.1 Bacterial strain

E. coli DH5α (Thermofisher, catalogue #265017) was used for cloning and DNA plasmid production; meanwhile *E. coli* BL21 star (DE3) (Thermofisher, catalogue #C601003) for protein overexpression.

E. coli transformation was performed with $CaCl_2$ method.

ii.1.1.2 Expression and purification

E. coli β-galactosidase gene had been subcloned in expression vector pRSF-1b (Milipore, cat. #71363-3) previously.

Protein expression trials

Protein expression procedure was evaluated at two temperatures (24ºC-37ºC) and different OD_{600} induction with 1 mM IPTG as inducer.

Samples were centrifuged and frozen for analysis. After sample resuspension (1 mM of protease inhibitor phenylmethane sulfonyl fluoride, PMSF from Sigma, cat. #93482-250mL, 300 mM NaCl, 50 mM phosphate), cells were lysated with sonicator (50% amplitude; 5 min. of 10s ON and 20s OFF, 0ºC) *Sonics Vibre-cell*. Then samples were centrifuged to finally analyse supernatant by SDS-Page (8-12%).

Expression and purification

Transformed *E. coli* BL21(DE3) cells, carrying the respective recombinant plasmids, were grown in 2-3 mL LB medium (ampicillin 0.1 mg/mL) overnight at 37ºC and 250 rpm. Those cultures were used as inoculum the next day for 2 L Erlenmeyer flasks containing 500 mL of LB medium (Sigma, cat. #L3022-1kg) with ampicillin (0.1 mg/mL). Inoculated media were incubated at 37ºC, 250 rpm. When cultures reached optical density (OD_{600}) of 2, expression of β-galactosidase protein was induced by adding 1 mM isopropyl-β-thiogalactopyranoside (IPTG, Anatrace, cat. #GEN-S-02122-5-5G). Cells were incubated overnight at 37ºC at 250 rpm.

Then, cells were harvested by centrifugation (5,000 rpm for 20 min at 4ºC) and resuspended in 50 mL of washing buffer (300 mM NaCl, 50 mM phosphate, 10 mM imidazole, pH 8.0) with 1 mM of PMSF. Cells were lysed with sonication (50% amplitude; 5 min. of 10s ON and 20s OFF, 0ºC) with sonicator *Sonics Vibre-cell*. Cell

debris was removed by centrifugation (12,000 rpm, 1h, 4ºC) and filtrated through 0.45 µm filters (Teknokroma, cat. #TR-200106). Supernatants were loaded onto a HisTrap® 5 mL column (GE Healthcare, catalogue #71-5027-68 AF) previously equilibrated with washing buffer (300 mM NaCl, 50 mM phosphate, 10 mM imidazole, pH 8.0). The column was washed to remove unbound proteins. One step of 50% of elution buffer (300 mM NaCl, 50 mM phosphate, 500 mM imidazole, pH 8.0), 255 mM imidazole, allowed protein recovering. Previously in the project, purification was also performed in gradient mode; however protein eluted was too diluted.

Eluted proteins were dialyzed against 50 mM phosphate (pH 7.0) to extract imidazole. Finally, purified proteins were stored at 4ºC.

Enzymes were analysed by SDS-PAGE to determine their purity that in all cases was >95%. Protein concentration was determined by the bicinchoninic acid (BCA) method (Thermo Scientific ™, cat. #23227) using bovine serum albumin (BSA) as standard.

ii.1.2 Analytical

ii.1.2.1 HPLC-MS

Specficic conditions of the mass detector are (Figure ii. 1):

	Galactose		Gaxilose	
Adduct	Na^+	NH_4^+	Na^+	NH_4^+
m/z	203	198	355	330
Fragmentor	50	50	50	50
Actual Dwell	69	69	69	69

Figure ii. 1. *Conditions of the detector for the detection of galactose and gaxilose adducts (HPLC-MS method)*

During the analysis by HPLC-MS diverse controls were performed.

- Analysis always began with blank injections to stabilize MS detector.
- Standard curve of galactose, 4- and 3-*O*-β-D-galactopyranosyl-D-xylose were analysed two or three times during analysis depending on lenght of time analysis. It is known that factor response can change on time (mainly in the first time injection). The objective was to control MS detector response and use exact factor response.

ii.1.2.2 Analytical method of Interquim S.A.

Interquim S.A facilities has its own analytic department and samples of reactions carried out on Interquim were analysed by another method. They used a HPLC method with an evaporative light scattering detector (HPLC-ELSD). The chromatopgraphic column is *XBridge Amida* (100x4.6mm, 3.5µm). Two columns were connected in serie

at 35ºC. Elution of anlaytes were carried out in a gradient mode with phase A (water) and phase B (CAN with NH_4OH 0.1%). The chromatopgrap was programmed as follows: T_0 5%-A, 95%-B; 10min: 20%-A, 80%-B; 40min 20%-A, 80%-B; 42 min 5%-A, 95%-B (column flow 0.5 mL/min, injection volume of 5µL). The main peak has a retention time around 22.4 min.

Detector parameters were: evaporator temperature at 80ºC, nebulizer temperature at 50ºC, gas flow at 1SLM, date rate of 40Hz, smoothing 50 and PMT gain 1.

This methodology allows to calculate the gaxilose content as area per cent and calculate the impurities content as area per cent.

ii.1.3 Kinetics

All reactions have been performed with duplicates or triplicates.

ii.1.3.1 Glycosidase: Activity on oNPG

Glycosidase activity on *o*NPG (13 mM), hydrolysis, was performed in reaction buffer (100 mM phosphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0) with volume of 200 µL. Amount of enzyme was 6.4 nM-66 nM for wild type enzyme and 6,650 - 650 nM for analysed mutants.

Rates of reactions were determined by incubating the substrates in reaction buffer in a thermostated microplate and initiated by addition of enzyme to the preincubated mixture. Activities were determined following changes in UV absorbance at 405 nm due to the release of *ortho*-nitrofenol (*o*NP). Extinction coefficient was determined under same experimental conditions. Initial rates were calculated from linear part of the progress curve. One unit hydrolyzes 1.0 µmole of o-nitrophenyl β-D-galactoside to *o*-nitrophenol and D-galactose per min at pH 7.3 at 37 °C.

ii.1.3.2 Transglycosidase activity

ii.1.3.2.1 Specific activity

Specific activity was carried out in 200 µL volume with 5 mM *o*NPG and 25 mM xylose in reaction buffer (100 mM phoshphate, 50 mM β-mercaptoethanol, 1 mM $MgCl_2$, pH 7.0). Reactions were performed in thermostated microplates using Bravo Automated Liquid Handling Platform (Agilent). Microplates were incubated at 37ºC for 10 minutes; then, enzyme addition initiated reactions. Samples were withdrawn and diluted in stop solution (1:1, isopropanol: water). After reaction stop, microplates were sealed and analysed by HPLC-MS.

ii.1.3.2.2 Long time reactions

All reactions were carried out in reaction buffer (1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol pH 7.0) at 37ºC. Different conditions were evaluated with different proposals.

Laboratory conditions: allow simulatig industrial conditions but with initial substrate totally dissolved: 250 mM xylose, 50 mM *o*NPG and 0.5-2 U of enzyme.

Analytic conditions to calculate kinetic parameters and evaluate reaction in homogeneous conditions: 0.01-0.1 U of enzyme, 25 mM xylose and 5 mM *o*NPG.

Industrial conditions: essential to evaluate real conditions for the enzymes: 2,5-5U of enzyme; 0.1 mM *o*NPG and 0.4 mM xylose

All reactions were analysed by HPLC-MS (CHAPTER 1, see 1.2) and different dilution factors were employed according concentration used.

ii.1.4 Industrial approach

-

ii. CHAPTER 2

ii.2.1 Molecular biology

ii.2.1.1 Bacterial strains

Same strains and methodology of *CHAPTER 1.*

ii.2.1.2 Gene design and subcloning

β–gal genes were supplied by GeneArt® Life Tecnologies (sequences in Appendix). The sequences included the open reading frames of every gene in cloning vector *pMK* or *pMA*. For genes with sequences longer than 3 kbp, two parts per gene were ordered. Genes were designed including restriction sites to subclone in plasmid *pET-22b* (Millipore, cat. # 697744) (Figure ii. 2). *NdeI* (New England Biolabs, cat. #R0111S) was added in N-terminal and *Hind III* (New England BioLabs, cat. #R3104T) in C-terminal. Both restriction sites were designed to subclone genes in frame and to add a His-tag on C-terminal (tag included in *pET-22b(+)* for further protein purification).

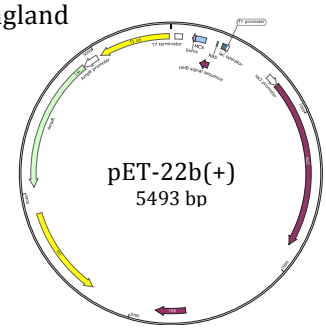


Figure ii. 2 *Map of the expression vector pET-22b(+)*

According to the length of the gene different strategies were used to build the constructs *(*Figure ii. 3*)*. If genes were less than 3 kbp, normal procedure was carried out: digestion of gene and vector with corresponding restriction enzymes, fragments purification (gel extraction kit from Sigma, cat. #NA1111-1KT), ligation (10 min, room temperature, *T4 DNA ligase*, ThermoFisher Scientific, cat. #EL0014) and final transformation in *E. coli* DH5α competent cells ($CaCl_2$).

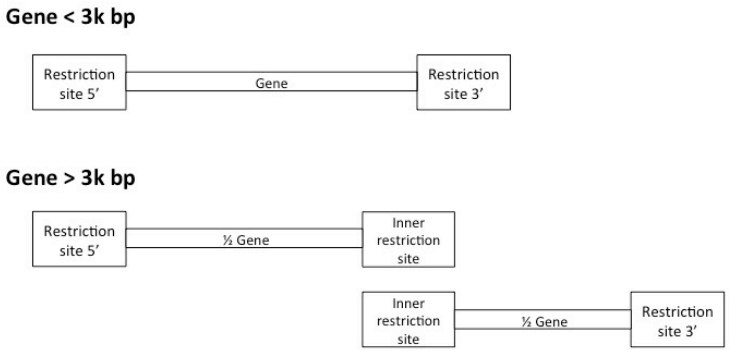


Figure ii. 3. *Strategies used to assemble the genes and subclone in* pET-22b according gene length

GENE < 3 kbp

This strategy was used for β-galactosidase genes of *Sulfolobus solfataricus* (GH1), *Paenibacillus thiaminolyticus* (GH35), *Bifidobacterium longum subsp. infantis* (GH42), *Geobacillus kaustophilus* (GH42) and *Thermotoga neapolitana* (GH42).

Sequences of all genes were checked by agarose gel electrophoresis (Figure ii. 4) and subsequent automated DNA sequencing.

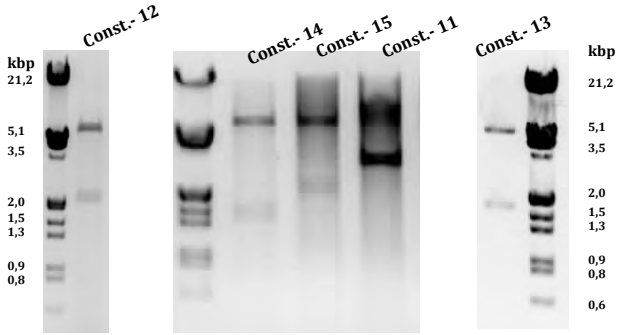


Figure ii. 4. *Agarose gel (0.8%) to check presence of insert in plasmid pet22-b for genes < 3 kbp. Lenghts: pet22b-5.5 kbp; const-12: 2.1 kbp; const-14:1.5 kbp ; const-15: 2.0 kbp ; const-11: 2,2 kbp; const-13: 1.8 kbp*

GENE >3 kbp

For genes higher than 3 kbp, an inner restriction site was needed to assemble the whole gene. Each gene had a different restriction site according to their sequence (Table ii. 1*).*

β–galactosidase	**Restriction site**
B. circulans	*Age I* (ACCGGT)
[redacted]	*Aci I* (AACGTT)
L. delbrueckii	*Pvui I* (CGATCG)
S. rectivirgula	*Vsp I (Ase I)* (ATTAAT)

Table ii. 1. *Inner restriction site selected for each gene*

In this strategy, three DNA fragments should be ligated (vector + N-terminal and C-terminal fragments of the gene). Different options were used depending on the gene:

Option A) Digest and purify each fragment (from agarose gel and gel extraction kit following the manufacturer's direction) and ligate three fragments in one step: 10 min, room temperature, *T4 DNA ligase* (ThermoFisher Scientific, cat. # EL0014). Used for the *L. delbrueckii* galactosidase gene.

Option B) Digestion and ligation of the two gene fragments (1 and 2, Figure ii. 5). PCR amplification of the complete gene to increase the amount of ligated DNA (3, Figure ii. 5) followed by digestion with external restriction

sites (4, Figure ii. 5). Finally, ligation with the vector, previously digested (5, Figure ii. 5), was performed before a standard transformation with *E. coli* DH5α competent cells ($CaCl_2$). This methodology was used for galactosidase genes from *B. circulans*, [redacted] and *S. rectivirgula.*

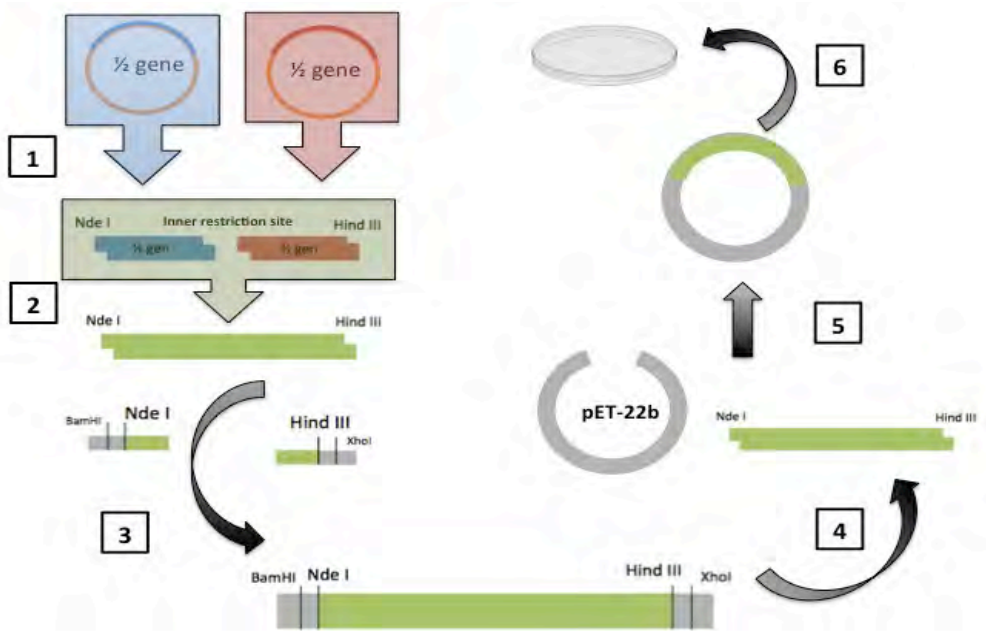
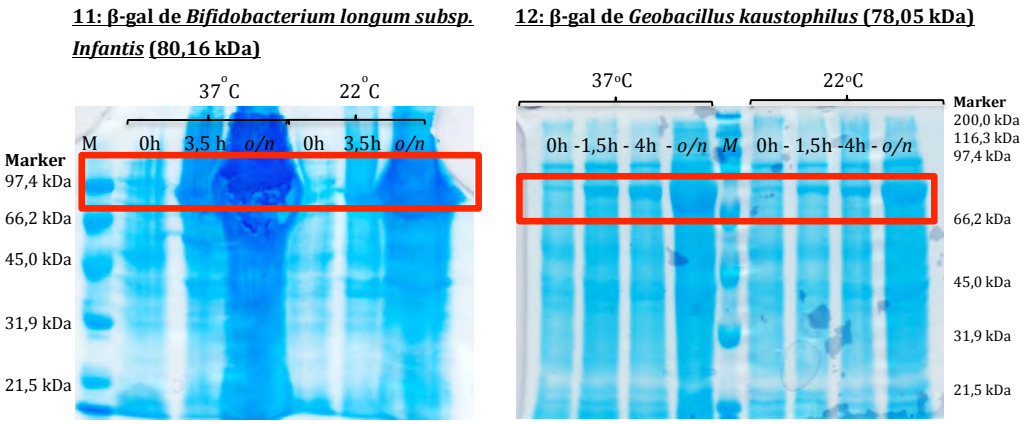


Figure ii. 5. *Subcloning methodology used to assemble three fragments (Gene > 3kbp)*

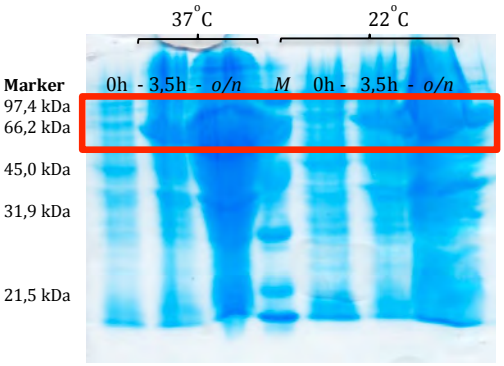
ii.2.1.3 Expression and purification: pET22b(+)

Protein expression trials

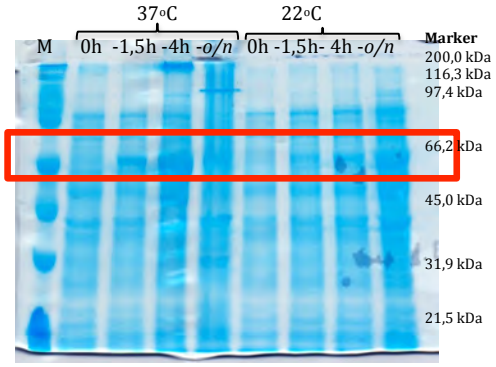
Protein expression from pET-22b clones was evaluated in protein expression trials at different temperatures (22 ºC -24ºC and 37 ºC) for each clone in *E. coli* BL21(D3) transformed cells using IPTG as inducer (Figure ii. 6).



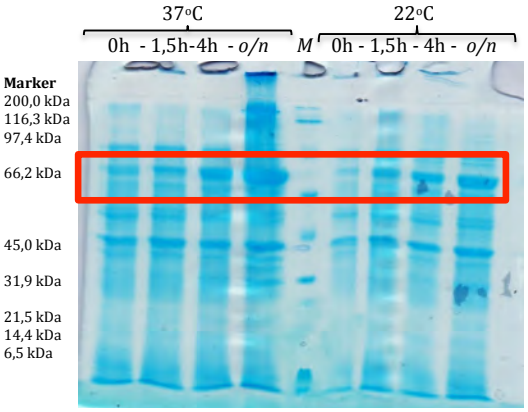
13: β-gal de *Paenibacillus thiaminolyticus* (66,16 kDa)



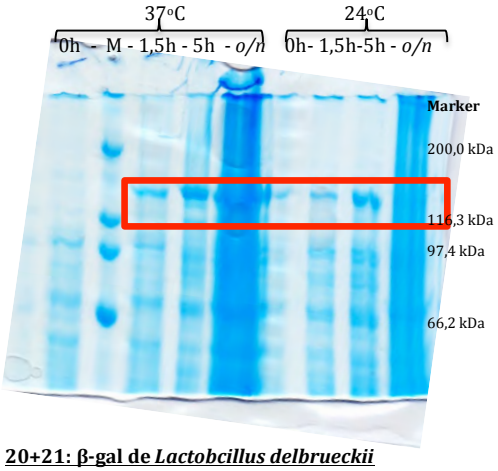
14: β-gal de *Sulfolobus solfataricus* (56,91 kDa)



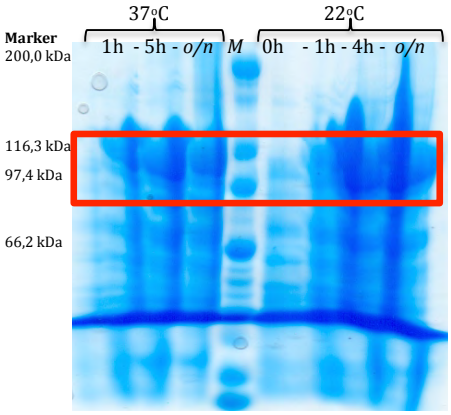
15: β-gal de *Thermotoga neopolitana* (75,09 kDa)



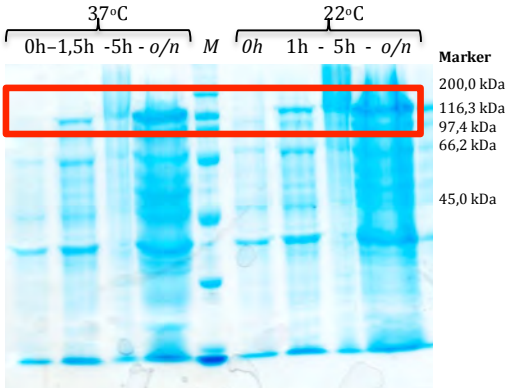
16+17: β-gal de *Bacillus circulans* (135 kDa)



18+19: β-gal de [redacted] (116 kDa)



20+21: β-gal de *Lactobcillus delbrueckii* (114 kDa)



22+23: β-gal de *Sacharopolyspora rectivirgula* (140 kDa)

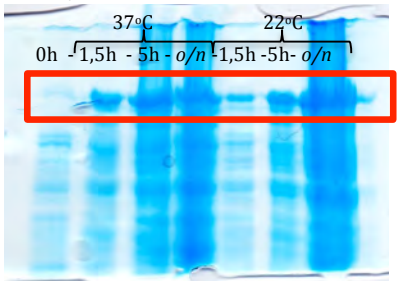


Figure ii. 6. *SDS-Page analysis of cell extracts to evaluate protein expression at different temperatures and different times of induction*

Expression and purification

Transformed *E. coli* BL21(DE3) cells, carrying the respective recombinant plasmids, were grown in 2-3 LB medium (ampicillin 0,1mg/mL) o/n. Those cultures were used as inoculum the next day of Erlenmeyer flasks containing 500 mL of LB medium (Sigma, catalogue #L3022-1kg) with ampicillin (0,1 mg/mL). Inoculated media were incubated at 37ºC, 250 rpm. When cultures reached an optical density (OD_{600}) of 1-2, expression of *β*–galactosidase genes were induced by adding 1 mM isopropyil-*β*-thiogalactopyranoside (IPTG, Anatrace, cat. #GEN-S-02122-5-5G). Cells were incubated overnight at 37ºC, 250 rpm, except *β*–gal for *E. clocacae* that was induced at 23ºC. Then, cells were harvested by centrifugation (5,000 rpm for 20 min at 4ºC) and resuspended in 50 mL of washing buffer (300 mM NaCl, 50 mM phosphate, 10 mM imidazole, pH 8) with 1 mM phenylmethanesulfonyl fluoride (PMSF, protease inhibitor, Sigma, cat. #93482-250mL). Cells were lysed with sonication (50% amplitude; 5 min of 10s ON and 20s OFF) with sonicator *Sonics VIBRA-cell*. Cell debris was removed by centrifugation (12,000 rpm, 1h, 4ºC) and filtrated through 0,45 µm filters (#TR-200106, Teknokroma). Supernatants were loaded onto a HisTrap 1mL (or 5mL) column (GE Healthcare, cat. #17-5248-01) previously equilibrated with washing buffer (300 mM NaCl, 50 mM phosphate, 10 mM imidazole, pH 8.0). The column was washed to remove unbound proteins. One step of 50% of elution buffer (300 mM NaCl, 50 mM phosphate, 500 mM imidazole, pH 8.0), 255 mM imidazole, allowed protein recovering. Eluted proteins were dialyzed against 50 mM phosphate (pH 7.0) to extract imidazole. Finally, purified proteins were stored at 4ºC.

Enzymes were analysed by SDS-PAGE to determine their purity that in all cases were >95% (Figure ii. 7). Concentration was determined by the bicinchoninic acid (BCA) method Thermo Scientific™, cat. #23227,) using bovine serum albumin (BSA) as standard.

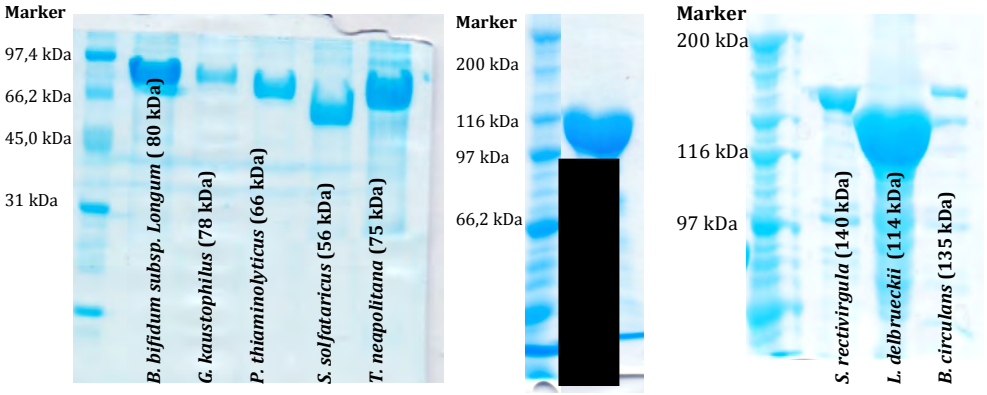


Figure ii. 7. *SDS-PAGE of purified enzymes*

ii.2.2 Kinetics

ii.2.2.1 Glycosidase and transglycosidase assays

Glycosidase activity was determined as explained in *METHODS AND MATERIALS* of *CHAPTER 1 (ii.1.2).*

ii.2.2.2 Glycosidase pH profile

pH dependece of *glycosidase* activity was determined on k_{cat}/K_M in the range 5 to 9. At each pH donow was varied (0.1 to 12 mM, 11 different concentration) and acceptor kept constant at 25 mM. Initial rates were fitted to the Michaelis-Menten equation to calculate the kinetic parameters by non-linear regression. k_{cat}/K_M *versus* pH data were fitted to a single ionization curve (Eq ii.1).

$$\frac{k_{cat}}{K_M} = \frac{A+B\cdot 10^{(pH-pK_{a2})}}{1+10^{(pK_{a1}-pH)}+10^{(pH-pK_{a2})}}; \qquad \textbf{Eq ii 1}$$

Reactions were done as described in glycosidase activity (absorbance at 405nm by releas of *o*NP group) (see *METHODS AND MATERIALS* of *CHAPTER 1, ii.1.2*).

ii.CHAPTER 3

ii.3.1 Kinetics

ii.3.1.1 Glycosidase and transglycosylase reaction

ii.3.1.1.1 Specific activity and long-time reactions

Specific activity of hydrolysis and trangslycosylase were performed as described in *METHODS AND MATERIALS* of *CHAPTER 1*

ii.3.1.1.2 pH profile

pH profile of *glycosidase* was done as described in *METHODS AND MATERIALS* of ii.*CHAPTER 2, ii.2.2.*

pH depende of the *transglycosidase* activity was determined on k_{cat}/K_M in the range 5 to 9 (9 points). At each pH, donor was varied (0.1 to 1mM, 4 different concentrations) and acceptor kept constant at 25 mM. Initial rates were fitted to straight lone corresponding to the lineal part of Michaelis-Menten equation when $[S]>K_M$ (Eq ii. 1).

$$\frac{V_i}{[E]} = \frac{k_{cat} \cdot [S]}{K_M + [S]}$$

$$if\ [S] \ll K_M, then;\ \frac{V_i}{[E]} = \frac{k_{cat} \cdot [S]}{K_M + \cancel{[S]}} = \frac{k_{cat}}{K_M}[S] \qquad \textbf{Eq ii. 1}$$

Initial rates were determined as explained in METHODS and MATERIALS, ii.CHAPTER 1.

ii.3.1.2 Hydrolytic assay by HPLC-MS

Hydrolytic assay was set up to determine primary and secondary hydrolysis. Reactions were performed in 200 µL volume with *o*NPG or gaxilose at 5 mM (dissolve on DMSO) in reaction buffer (1 mM $MgCl_2$, 100 mM phosphate, 50 mM *β*-mercaptoethanol pH 7.0). Amount of enzyme was between 0,1 nM-1 nM.

All reactions were performed as done in specific activity (*METHODS AND MATERIALS, 1.*)

ii.3.2 Screenig methodology

ii.3.2.1 Screening in agar plates

ii.3.2.1.1 Screening with digital screening method

Digital screening methodology was applied in two libraries: random mutagenesis in domain 3 and site directed mutagenesis at different sites simultaneously.

Methodology

Three plates were used: 1) growing plate (LB-agar and ampicillin 100 μg/mL); 2) hydrolysis analysis plate (100 mM phosphate, 15 g/L agar, 0.1 mM X-gal, 0.2% arabinose); 3) transglycosylase analysis plate (100 mM phosphate, g/L agar, 0.1 mM X-gal, 20 mM xylose).

Cells were spread above nitrocellulose membrane and it was used to transfer cells from one plate to another. Cells grew over night in plate one (37ºC) (*left*-Figure ii. 8). Once colonies were large enough, membrane were moved to plate two, where cells expressed protein (L-arabinose induction) and hydrolyse X-gal, blue colour appearance indicated hydrolysis activity (room temperature). As soon as different kind of soft blues appeared (1h), membrane was moved to plate three, where transglycosidase activity could take place. When membrane was moved first image was captured (T_o). Different images were captured during the two or three following hours (*right*-Figure ii. 8).

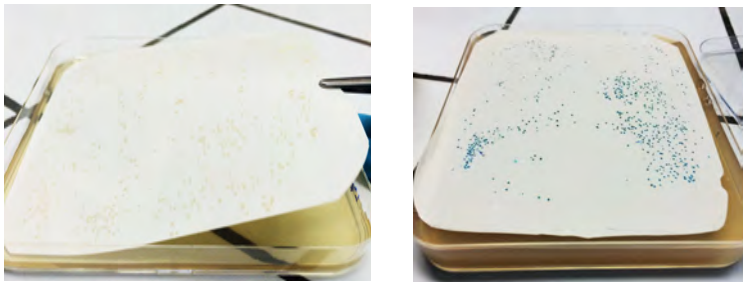


Figure ii. 8. *Nitrocellulose membrane used with colonies grown on it.* Left*: after growth in LB-plate.* Right*: after incubation with substrates (X-gal 0.1 mM, xylose 40 mM, L-arabinose 0.2%, 100mM phosphate, incubated at 24ºC)*

Image treatment

Image treatment was performed with VISIOLOG 6.3 software (Noesis). Different steps were carried out:

1) Image superimposition to be able to compare colonies at same position. Images must be of the same measure and quality to be superimposed (TIFF format, resolution of 600 dpi, 16 bits).
2) Grey calibration of each image (*left*-Figure ii.8).
3) Colony identification and creation of a binary image (*center*- Figure ii.8).
4) Count of all colonies (right- Figure ii.8).

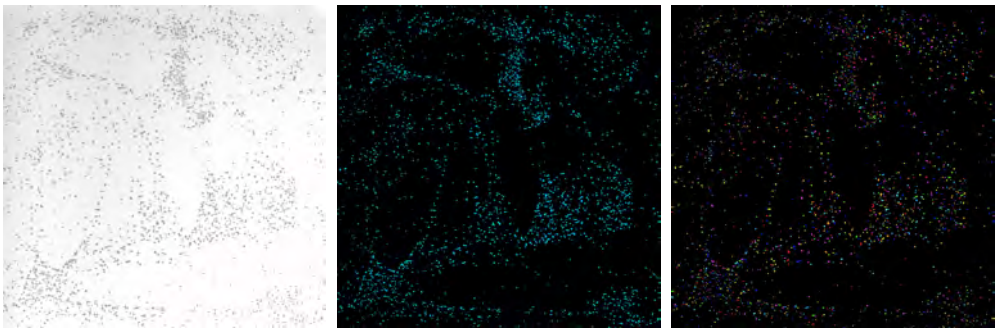


Figure ii. 9. *Image processing steps carry out with Visiolog software.* Left- *grey calibration;* Center- *colonies identification;* Right- *colonies numbered*

5) Superimposition of image of colonies numbered with images previously calibrated (good superimposition was verified) and grey intensity was determined (Figure ii. 10).

Figure ii. 10. *Final superposition of colonies image (black and grey) and binary image with colonies numbered*

6) Intensity of each colony was evaluated taking into account intensities of T_o and T_x (Eq ii. 2).

$$\frac{(Intensity\ T_f - Intensity\ T_o)}{Intensity\ T_o}$$ **Eq ii. 2**.

7) Results of calculations were ordered and plotted to easy choose the best mutants (Figure ii. 11).

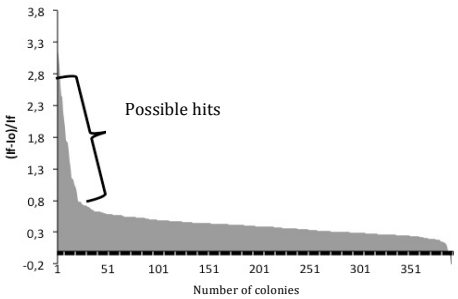


Figure ii. 11. *Mutants intensity calculation*

ii.3.2.1.2 Replica plating methodology

Replica plating methodology was used to evaluate small libraries without quantification.

Cells transformed (TOP10 strain) with those libraries were spread in LB-Ap agar plates (Ap, ampicillin 100 µg/mL) and incubated overnight at 37ºC. Replica plating was done and plates were incubated at 21ºC-25ºC. Plates used were:

- AXA: ampicillin (100 µg/mL), X-Gal (0.1 mM) and L-arabinose (0.2%)
- AXAX: ampicillin (100 µg/mL), X-Gal (0.1 mM), L-arabinose (0.2%) and xylose (20 mM)

After 14 h colour changes appeared and colour was compared between plates. After 6 hours, differences were difficult to detect (Figure ii. 12).

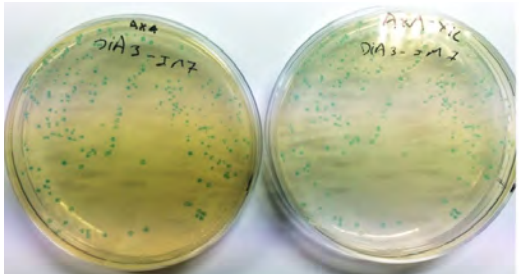


Figure ii. 12. *Replica plating results.* Left: *Arabinose, X-gal and ampicillin.* Left: *Arabinose, X-gal, xylose and ampicillin*

ii.3.2.2 Screening at liquid conditions

Second step of screening was carried out at liquid conditions. 1.5 mL cultures were grown and **protein expression** was induced in 2 mL microplates (0.2% L-arabinose, OD_{600} 0.5, 24ºC, *overnight)*.

Altough it was checked that substrate could enter cells, cells were lysed for this screening to reduce possible interferences. Cultures were harvested by centrifugation (5,000rpm, 30min at 4ºC) and resuspended in 200 µL of commercial buffer for lysis step (BugBuster™ Protein Extraction Reagent; Millipore, cat. #70584-4). Lysis buffer was incubated with cells during 15 min at room temperature. Insoluble cell debris were removed by centrifugation (1h, 4ºC, 5,000 rpm) and supernatant was transferred to a new plate using Bravo Automated Liquid Handling Platform (Agilent).

Reactions were performed 200 µL in reaction buffer (1 mM $MgCl_2$, 100 mM phosphate, 50 mM β-mercaptoethanol pH 7.0), 10 mM *o*NPG and four different concentrations of xylose (0 – 0.3 M).

Initial rates of reaction were determined as done in glycosidase reactions.

ii.3.3 Molecular biology and protein expression

ii.3.3.1 Subcloning in pBAD

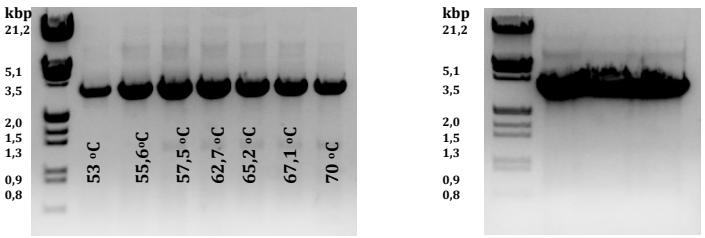
β-galactosidase gene from [redacted] was extracted from previous plasmid construction (*pET-22b(+)*) obtained in *CHAPTER 2*. Gene was amplified with corresponding primers including restriction sites (*fw*: GGAATTAACCATGGATTGGGAAAATCCGGG; *rv* from T7*:* GCTAGTTATTGCTCA) by PCR protocol (Table ii. 2).

PCR PROTOCOL				PCR VOLUMS		
1	start	98ºC	5min	V_{TOTAL}	70 µL	
	Denature	98ºC	20 s	I*Proof mix* (#172-5310, Biorad)	35 µL	
X 35	Annealing	variable	20 s			
	Extension	72ºC	2,5 min	DNA	4 µL	4 ng
	Final extension	72ºC	10 min	Primers (*fw+rv*)	3,5µL each	0.5 µM each
	Kept at	0ºC		Water	24 µL	

Tabl

e ii. 2. *PCR protocol and volumes used for amplification of β-galactosidase gene form*[redacted]

Firstly, it was verified which annealing temperature amplifies better, in a higher amount and more specificly the desired gene ($T_{annealing}$ 53ºC to 70ºC). After agarose gel analysis (*left*- Figure ii. 13) it was seen that $T_{annealing}$ of 60 ºC provided higher yields. Using same protocol (Table ii. 2) and selected temperature (60ºC) gene amplification was performed (*right*-Figure ii. 13). Then a kit was used to extract and purify gene band (Sigma, cat. #NA1111-1KT).



	T (ºC)	Time
	98°C	30 s
	98°C	10 s
X 30	59°C	2 min
	72°C	2 min
	72°C	10 min
	4 °C	∞

Figure ii. 13. *Agaros[redacted]rify DNA amplification for gene of the β-galactosidase from[redacted]e sublconing.*Left*; agarose gel for temperature analysis of $T_{annealing}$ in gene ampliciation by PCR.* Right*: agarose gel of gene amplified by PCR step*

Table ii. 3. *PCR conditions for the amplification of the β-galactosidase gene from*[redacted]

PCR band and vector (among possibilities offer by pBAD vector, *pBAD-A* version was chosen) were digested with corresponding restriction enzymes: *XhoI* and *HindIII* (New

England BioLab, cat. #R0146L and #R3104T,). Then, they were purified with cleaning kit (#NA1020-1KT, Sigma) and ligated with ligase (10 min, room temperature, *T4 DNA ligase,* ThermoFisher Scientific, cat. #EL0014,). Finally, ligation product was transformed into $CaCl_2$ competent *E. coli* TOP10 cells. After colonies grew in liquid medium and corresponding plasmid DNA extraction, it was checked digestion (*Hind III and XhoI*) and by agarose gel Figure ii. 14) and by sequencing (STABVIDA).

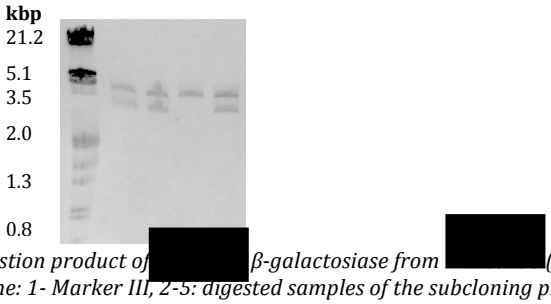


Figure ii. 14. *Agarose gel (0.8%) of digestion product of* [redacted] *β-galactosiase from* [redacted] *(3051 bp) subcloned in pBAD vector (4100 bp). Lane: 1- Marker III, 2-5: digested samples of the subcloning procedure with Hind III and XhoI*

ii.3.3.2 Protein expression: TOP10 cells and pBAD vector

Protein expression trials were performed under different conditions: different concentrations of inducer (L-arabinose, 0.2%-0.0002%), different temperature of induction (24ºC or 37ºC) and different initial OD_{600} before induction (0.4-2.0). It was concluded that optimum conditions for protein expression in TOP10 cells transformed with recombinant pBAD plasmids was: OD_{600} 0.4-0.6, with 0.2% of L-arabinose at 24ºC.

ii.3.3.3 Dockings

Molecular docking was performed with AUTODOCK (http://autodock.scripps.edu [388–390]). Protein and ligand structures were parameterized with Autodock 4.2. All rotable bonds of the ligand were considered free during the docking calculations, while protein was kept fixed. The search space was defined as a cubic box of 68 x 68 x 44 Å centred in the active site of the enzyme.

200 dockings were analysed and 10 were situated inside catalytic site.

ii.3.3.4 Point mutations

Site directed mutagenesis and site saturation mutagenesis were carried out with *Quick Change*® protocol optimized in lab.

PCR was performed to obtained DNA mutated (summary of primer sequences and temperatures used are in *Appendix 4.6.1*) (Table ii. 4).

PCR PROTOCOL				PCR VOLUMS		
1	start	98ºC	3 min	V_{TOTAL}	40 µL	
	Denature	98ºC	30 s	I*Proof mix* (#172-5310, Biorad)	20 µL	
X 35	Annealing	variable	30 s			
	Extension	72ºC	2,5 min	DNA	x µL	4 ng
	Final extension	72ºC	10 min	Primers (*fw+rv*)	2 µL each	0,5 µM each
	Kept at	0ºC		Water	40-x µL	

Table ii. 4. *PCR protocol and volumes used for QuickChange (site directed mutagenesis)*

After PCR, samples were cleaned-up of primers, polymerase, *d*NTPs a kit for PCR samples purification (Sigma, cat. #NA1020-1KT). Digestion with *DpnI* endonuclease (New England Biolabs, cat. #R176L) was used to digest methylated DNA template (*wild type* template). Then, presence and quality of DNA were checked before transforming cells by agarose gel.

Finally, DNA samples were used to transform TOP10 $CaCl_2$ competent cells.

ii.3.2.5 Random mutagenesis

Random library on domain 3 (catalytic) was performed in several steps:

1) epPCR on domain3 (Table ii. 5).

		Mg^{2+}	Mn^{2+}	dNTPs
1	Control +	2 mM	0 mM	0.8 mM
2	Standard protocol	6 mM	0.5 mM	1 mM dCTP 1 mM dTTP 0.2 mM dATP 0.2 mM dGTP
3	2, with more Mg^{2+}	8 mM		
4	3, with more Mg^{2+}	12 mM		

Table ii. 5. *epPCR conditions for construction of de randomize library on domain 3*

2) *DpnI* endonuclease digestion (New England Biolabs, cat. #R176L) of PCR product to remove wild type template, followed by thermal inactivation (20 min, 80ºC).
3) PCR of epPCR product to amplify mutated DNA. Purification of product by agarose gel and kit of DNA extraction (Qiagen, cat. #28704).
4) PCR of cloning vector plus mutated gene fragment. Purification of products by agarose gel and kit of DNA extraction.

5) Circular Polymerase Extension Cloning (CPEC) to assemble both fragments (Table ii. 6).

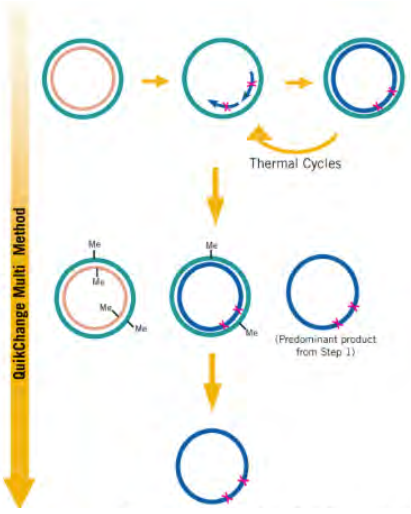
	Volumes
Phusion HF buffer (5x)	4 µL
dNTP mix (10mM)	1,6µL (0,8mM)
Vector	50-100 ng
Insert	variable
H_2O	X µL
Phusion HF DNA polymerase (2U/µL)	0,2 µL
Total volume	20 µL

Table ii. 6. *CPEC volumes used for the assemble of the construct (part of the gene randomize, external parts of the gene and vector pBAD)*

6) Competent cells transformation: commercial ultracompetent cells (XL10-Gold ultracomptents cells provided by kit of *QuickChange® Multi Sire- Directed Mutagenesis*, Agilent cat. #200514).
7) Growth of transformed competent cells with the pool of mutant plasmid in liquid LB-Ap over night.
8) Extraction of pool of plasmidic DNA.
9) Transformation of $CaCl_2$ competent *E. coli* TOP 10 cells.
10) Screening of transformed cells in LB-Ap agar plates.

ii.3.3.6 QuickChange® Multi Site-Directed Mutagenesis kit

This kit offered a rapid and reliable method for site-directed mutagenesis of plasmid DNA at different sites simultaneously (Agilent, cat. #200514) (Figure ii. 15).



1) PCR using primers that binds at one strand and ligation of nicks with the QuickChange® Multi enzyme blend.
2) *Dpn I* digestion of template.
3) Transform mutated ssDNA into XL10 Gold ultracompetent cells.
4) Miniprerp of transformed cells pulls.
5) Transformation in TOP10 competent cells, needed for screening (Δlac cells).
6) Screening

Figure ii. 15. *QuickChange® Multi Site-Directed Mutagenesis method*

ii.3.3 3D structure

ii.3.3.1 Model

Model building was performed with SWISS MODEL server (www.model.expasy.org [311–314]). High homology sequence between [redacted] and *E. coli* β-galactosidase, identity 65%, allowed the use of *E. coli* β-galactosidase structure as feasible templates. Different β-galactosidase structures from *E. coli* were used as templates to build the [redacted] β-galactosidase structure model. However, 4DUV structure of *E. coli* β-galactosidase was chosen for building acquisition due to 4DUV presented close conformation (active conformation).

SWISS MODEL server evaluated models achieved and all of them had acceptable values (Figure ii. 16). Local Quality plot shows for each residue (x-axis) the expected similarity to the native structure (y-axis): values below 0.6 are expected to be low quality. It is highlighted that nearly all residues were above 0.6 value: around amino acid 750 was a lower quality probably due that this part belongs to domain 4: non-catalytic domain, hugely variable in β-galactosidases from GH2 and not present in β-galactosidase from families GH35 and GH42.

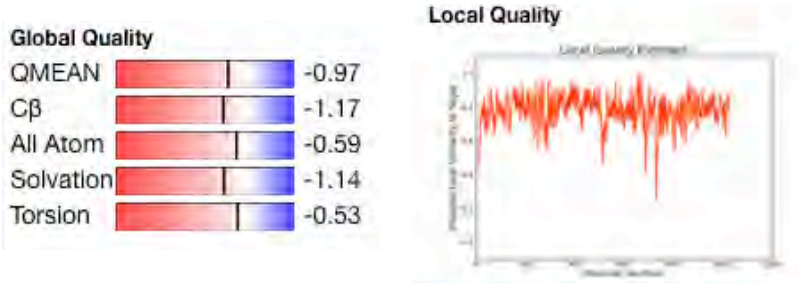


Figure ii. 16. *Swiss model server evaluation*

ii.3.3.1 Crystal trials and X-Ray diffraction experiments

3D protein structure determination of [redacted] β-galactosidase was performed in collaboration with Dr Alberto Marina from the IBV-CSIC in Valencia (Spain). Inactive mutant E539A was purified at Biochemistry laboratory of IQS School of Engineering from Δ lacZ TOP10 cells overexpressing the protein cloned in L-arabinose pBAD vector. Two litres of cell culture (4x0.5L) were used to produce enough protein (induction with 0.2% arabinose, at OD_{600} 0.5 and 24ºC).

Purification step, after lysis procedure explained in *CHAPTER 2*, included two columns: affinity column (HisTrap; GE Healthcare, catalogue #71-5027-68 AF) and gel filtration (Superdex 200, 16/60). First step purification inclued two steps of cleaning: one with 10 mM imidazole (*Peak 1*) and another with 59 mM imidazole (*Peak 2*). 255

mM imidazole was used for protein elution (*Peak 3*) (*left*-Figure ii. 17) (buffer 300 mM NaCl, 50 mM phosphate, pH 8.0). Second step purification allowed rejecting aggregates (*right*-Figure ii. 17) (buffer 50 mM TRIS, 100 mM NaCl, pH 8.0). After protein elution, protein concentration was performed (Amicon 10K, 15 mL, Millopre, cat. #UFC901024) and sent to Dr Marina's laboratory.

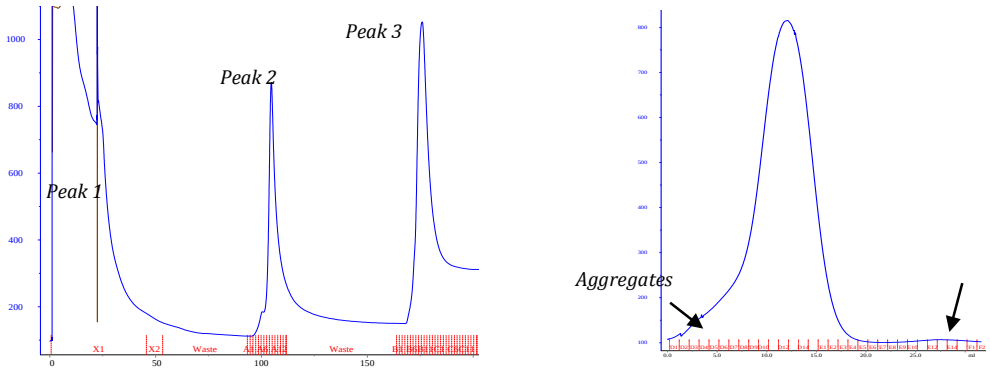


Figure ii. 17. *Purification chromatograms.* Left: *affinity purification chromatogram.Peak1: 10 mM imidazole, 59 mM imidazole, 255 mM imidazole*. Right: *gel filtration chromatogram*

Protein crystals were obtained at different conditions. Crystals analysis were performed at ALBA SYNCHROTRON (Cerdanyola del Vallès, Barcelona, Spain) and SYNCHROTRON DIAMOND (Oxford, UK).

Data processing was performed with PHENIX software (Python-based Hierarchical Environment for Integrated Xtallography is a software suitable for the automated determination of molecular structure using X-ray crystallography and other methods [391,392]).

COOT software (Crystallographic Object-Oriented Toolkit [393]) was used to display and manipulate models get. This program is focused on building and validating atomic models into three-dimensional electron density maps, obtained by X-ray crystallography or electron microscopy. COOT allows model manipulation such as real space refinement, manual rotation/translation, rigid-body fitting ligand search, solvation, mutations, rotamers, among other tools.

APPENDIX

A. 1 Summary of candidates

Bacteria	GH	MW kDa	Gene	UNIPROT	Quaternary structure	pH	T	Hydrolysis Product(s)	Transglycosylation Product(s) (substrate used)	Prot. Eng	Ref.
Arthrobacter psychrolactophilus strain F2	GH2	111.7	bglA	Q08KN3		8.0	10[0]C (cold active)	Lactose (K_M 42.1 mM; k_{cat} 3.02s^{-1}) *o*NPGal (K_M 2.7 mM; k_{cat} 12.07s^{-1}) pNPGal	Trisaccharides		211,212
Arthrobacter sp. C2-2		110.8	lacZ	Q8KRF6 (1YQ2)	Homo-tetrameric	7.5	40[0]C (thermolabile)	Lactose (K_M 53.1 mM; k_{cat} 1,106 s^{-1})	34 mM trisaccharides 6 mM tetrasaccharides (680 mM lactose)		213
Arthrobacter sp. 32cB		109.7	N/A	A0A023UGN9	dimer	8.0	28[0]C	Lactose (K_M 1.52 mM; k_{cat} 30.55 s^{-1}) *o*NPGal (K_M 16.56 mM; k_{cat} 31.84 s^{-1})	Trisaccharides, tetrasaccharides, low amount pentasaccharides (292-582 mM lactose) Aklyl glucosides (2-propyl, 1-butyl, 1-hexyl, cyclohexyl) Heterooligosaccharides (fructose, xylose, arabinose)		214
Bacillus circulans		192	bgaD	E5RWQ2	monomer			Lactose (V_{max} 158.5 U·min^{-1}·mg^{-1}) *o*NPGal (K_M 45.9 mM; V_{max} 277.7 U·min^{-1}·mg^{-1}	63 % yield (β-1,4)		216–219,294
									65% (β-1,3 and β-1,4) changing product synthesise	R484S	220
Bacillus megaterium		120	bgaBM	A3FFK4	monomer	7.5-8.0	55[0]C	Lactose (K_M 12.6 mM; k_{cat} 54.4 mM/min) *o*NPGal (K_M 9.5 mM; V_{max} 16.6 mM/min)	Heterooligosaccharides: fructose, glucose, sorbitose, sucrose, cellobiose, sorbitol, mannitol, alkyl alcohol among others (with *o*NPG as donor)		221
Bifidobacterium bifidum (DSM 20215)		188	BIF3	Q9F4D5	monomer	6.0	40[0]C[4]	*o*NPGal	GOS		222,223
Bifidobacterium bifidum (NCIMB 41171)		124	BbgIV	Q0ZII7	dimer	6.4-6.8	65[0]C[5]	*o*NPGal	59% di- and trisaccharides (43% w/w lactose)		109,224–227
[redacted]		116.8	lacZ	Q2XQU3	tetramer	6.5-10.5	35[0]C	*o*NPGal (K_M 0.01 mM; V_{max} 2 mM/min) Lactose (K_M 0,30 mM; V_{max} 0,21 mM/min)	49% GOS (275 g/L lactose) Heterooligosaccharides: galactose,, glucose, sucrose, trehalose, sorbitol, salicin(with *o*NPG as donor)		229–231
Klebisella pneumoniae		117	lacZ	F2X2W7 (P06219)		7.5	40[0]C	*o*NPGal (K_M 0.72 mM; k_{cat} 121 s^{-1})	43.2 g/L 1-lactulose (400 g/L lactose and 200 g/L		232–234

[4] Optimum temperatura not determined
[5] Optimum temperatura not determined

									fructose)		
Klyveromyces lactis		117	Lac4	P00723 (3OB8)	tetramer	6.8	38^{0}C		Lactulose 14.4 g/L; allo-lactulose 47.5 g/L (200 g/L lactose and 20 g/L fructose)	Immobilized	235–238
Lactobacillus acidophilus R22		72 + 35	LacL, LacM (R22)	B2LWG2	Heterodimer	6.5	55^{0}C	*o*NPGal (K_M 0.73 mM; V_{max} 28.8 µM/min/mg prot) Lactose (K_M 4.04 mM; V_{max} 361 µM/min/mg prot)	39% GOS: 16% disaccharides, 19% trisaccharides, 4% tetrasaccharides		239
Lactobacillus delbrueckii sbsps. Bulgaricus DSM 20081		115	LacZ	P0C1Y0	Homodimer	6.0 - 9.0	50^{0}C	Lactose (K_M 19.2 mM; k_{cat} 234 s^{-1}) *o*NPGal (K_M 0.92 mM; k_{cat} 603 s^{-1})	GOS (102 g/L): disaccharides and trisaccharides (β-1,6> β-1,3)		240
Lactobacillus plantarum WCFS1		35 + 72	lacLM	Q938Q6	Heterodimer	7.0-7.5	55-60^{0}C	Lactose (K_M 29 mM; k_{cat} 98 s^{-1}) *o*NPGal (K_M 0.9 mM; k_{cat} 92 s^{-1})	41% GOS (w/w); β-1,6> β-1,3 (600 mM lactose)		241
Lactobacillus sakei		72.5 + 35	lacLM	Q48846	Heterodimer	6.5	55^{0}C	Lactose (K_M 20 mM; k_{cat} 43s^{-1}) *o*NPGal (K_M 0.6 mM; k_{cat} 180 s^{-1})	41% GalOS (lactose 215 g/L): disaccharides and trisaccharides (β-1,6> β-1,3)		242
Enterobacter agglomerans		116.4	lacZ	A3FEW8	Homodimer	7.5-8.0	37-40 ^{0}C	Lactose (K_M 114 mM; k_{cat} 43 s^{-1}) *o*NPGal (K_M 0.06 mM; k_{cat} 180 s^{-1})	38% GalOS (13%, w/v, lactose) (*o*NPGal as donor and difernts acceptors inlcudincs hexose, pentose, alchohol)		228
Pseudoalteromonas sp 22b		117.5	lacZ	Q6RGS1	Tetramer	6.0 – 8.0	40 ^{0}C	Lactose (k_{cat} /K_M 47.5 mM^{-1}·s^{-1}; k_{cat} 157 s^{-1}) *o*NPGal (K_M 0.06 mM; k_{cat} 180 s^{-1})	Alkyl galactosides: 32 g/L 2-propyl galactopyranoside, 30 g/L 1-butyl galactopiyranoside (lactose and alcohols), 30% water	Immobilized on glutarldehydre-treated chitosan beads	243–245
Saccharopolyspora rectivirgula		145		Q9ZBD1	Monomeric	6.5-7.2	Thermophilic	Lactose (K_M 0.75 mM; k_{cat} 63.1 s^{-1}) *o*NPGal (K_M 0.03 mM; k_{cat} 47.6 s^{-1}) *p*NPGal (K_M 0.04 mM; k_{cat} 55.8 s^{-1})	41% (w/w) (1.75 M lactose)		246–250
Streptococcus thermophilus		117	LacZ	P23989		7.1		Lactose (K_M 6.90 mM) *o*NPGal (K_M 0.98 mM)	40% oligosaccharides: di- and trisaccharides (70% lactose)		251–253
Thermotoga maritima		129.5	lacZ	Q56307		6.5	80-85^{0}C	Lactose (K_M 1.6 mM; V_{max} 63.3 µmol·min^{-1}·mg^{-1}) *o*NPGal (K_M 0.33 mM; V_{max} 79.6 µmol·min^{-1}·mg^{-1})	97g/L GOS: tri- and tetrasaccharides , β-1.3 (500 g/L lactose)	F571L/N574S	156,157,254,255

		120	lacZ	Q56307		6.5	80-85^{0}C	*o*NPGal (K_M 0.33 mM; V_{max} 79.6 µmol $min^{-1}mg^{-1}$) Lactose (K_M 11.6 mM; V_{max} 27.8 µmol $min^{-1}mg^{-1}$)	GOS: Two di- and two tri-saccharides (30%, w/v, lactose)		254,256,257
Pyrococcus furiosus	GH1	54.7	celB	Q51723 (3APG)		5.0	90^{0}C	Lactose (K_M 186 mM; k_{cat} 3,800 s^{-1}) *o*NPGal (K_M 5.3 mM; k_{cat} 10,000 s^{-1})	40% tri- and tetrasaccharide: β-1,3> β-1,6 (70%, w/v, lactose)	F426Y/M424K	153,155,189–194
Pyrococcus woesei		61.7		O52629	dimer	5.4	95^{0}C	*o*NPGal (K_M 1.2 mM; k_{cat} s^{-1})	Galβ(1,4)GlcNAc-linker-tBoc (also β-1,6) (lactose as donor, GlcNac-link-tBoc as acceptor)	8 point mutations	195–199
Sulfolobus acidocaldarius		57	bgaS	P14288	dimer	5.5	90^{0}C	*o*NPFuc (K_M 0.12 mM; k_{cat} 26.4 s^{-1}) *p*NPFuc (K_M 0.14 mM; k_{cat} 18.6 s^{-1}) *o*NPGlu (K_M 0.14 mM; k_{cat} 16.7 s^{-1}) *o*NPGal (K_M 3.19 mM; k_{cat} 31.5 s^{-1})	25-30 g/L tri- and terasaccharides (pNPGal as a donor and Gal as a acceptor or lactose)		200
Sulfolobus solfataricus		56.7	lacS	G8GCT7	tetramer	6.0	85^{0}C	*o*NPGal (K_M 1.1 mM; k_{cat} 1,300 s^{-1}) Lactose (K_M 196 mM; k_{cat} 1,500 s^{-1})	53% GOS: tri- and tetrasaccharides (600 g/L lactose) (major product: β -D-Galp-(1–6)-β-D-Galp-(1–4)-β-D-Glcp) (58% F359Q; 62% F441Y)	F359Q/F441Y	153,154,201–205
Sporobolomyces singularis (Bullera singularos)		53	bglA (KCTC 7534)	Q564N5	dimer	5.0	50^{0}C		90 g/L GalOS (180 g/L)		206
		47.9	bglA (ATCC 24193)			3.5	45^{0}C	*o*NPGal (K_M 5.40 mM; V_{max} 3.07 µmol $min^{-1}mg^{-1}$) *p*NPGlc (K_M 1.96 mM; V_{max} 2.30 µmol $min^{-1}mg^{-1}$)	55% oligosaccharides. Tetra saccharides or higher GlcOS (cellobiose) > di-, tri- and tetra- GalOS (300 g/L lactose)		207–210
Thermus termophilus								Thermostable β-galactosidases for the synthesis of human milk oligosaccharides			
Bacillus circulans	GH35	66	bgaC	O31341	monomer	6.0		*p*NPGal (V_{max} 48.7 U·mg^{-1}) Lactose (V_{max} 61.0 U·mg^{-1}) *o*NPGal (V_{max} 12.7 U·mg^{-1})	Trisaccharides, β-1,3 (lactose)		264
Carnobacterium maltaromaticum (Carnobacterium piscicola)			bgaC	Q93DW7			40^{0}C (coldactive)	*o*NPGal (K_M 1,. mM V_{max} 450 mol· mim^{-1}·mg^{-1})			265,266

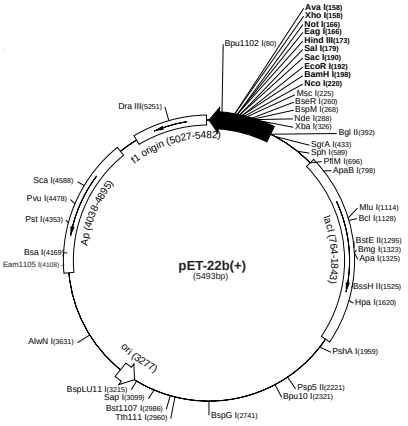
Aspergillus candidus (GCMCC3.2919)		109	LacB	Q8TFE6		6.5	40⁰C		32% GOS (400 g/L lactose)		259
Aspergillus niger		109	lacA	P29853		4.0			Mainly disaccharides and low amount of trisaccharides (lactose)		258
Aspergillus oryzae		109	lacA	Q2UCU3 (AIUG)		2.5-5.5	40⁰C		27% GOS: trisaccharides (50% lactose) Galactosyl-polyols (*o*NPG as donor)	Immobilized	260–263
Paenibacillus thiaminolyticus		68.6	galF	C7U109	monomer	5.5	65	*p*NPFuc (K_M 1.18 mM; V 1, µmol·min^{-1}·mg^{-1}) *p*NPGal (K_M 250 mM; V 270 µmol·min^{-1}·mg^{-1}) *p*NPGc (K_M 77 mM; V 0.1 µmol·min^{-1}·mg^{-1})		*p*NPFuc as donor with different acceptors: 69% (*p*NPMan; β-1,3), 42% (*p*NPGal; β-1,6), 94% (methanol), 72% (ethanol)		268
Penicillium expansum		110	lacA	B8QGZ3			25⁰C		44 % GOS (100 g/L lactose)	Immobilized cells	269
Penicillum sp.		109	lacA	Q700S9					70% Gal-β-1,6-Gal (AcGal as substrate)		270
**Hypocrea jecorina (Trichoderma reese)*		110	bgal	Q70SY0 (3OG2)	monomeric	5.0	60⁰C	*o*NPGal (K_M 0.36 mM; k_{cat} 17.31 s^{-1}) Galactobiose (K_M 9.06 mM; k_{cat} 5.68 s^{-1})	Disaccharides		267
Alicylcobacillus acidocladarius	GH42	77.9	bglY	C8WV58	trimer	5.5	65⁰C	wt: *p*NPGal (K_M 4 mM; k_{cat} 2,400s^{-1}) Lactose (K_M 60.8 mM; k_{cat} 5,212 s^{-1})			271,272
								E157G: *p*NPGal (K_M 0.1mM; k_{cat} 0.30 s^{-1}) Lactose (K_M 0.35 mM; k_{cat} 16.8 s^{-1})	93% di-, tri- and tetrasaccharides, β-1,3 and β-1,4, (*p*NPGal as donor and *o*NPGlu or *o*NPXyl as acceptors)		
Arthrobacter sp.		75.9	galB	C7ASJ5		6.0	20⁰C		Lactulose (40% lactose, 20% fructose , w/v)		273
Bacillus circulans		77	bgaA	Q45092	monomer	6.0	40⁰C	*p*NPGal (V_{max} 169.8 U·mg^{-1}) Lactose (V_{max} 44.3 U·mg^{-1}) *o*NPGal (V_{max} 47.2 U·mg^{-1})	Tri- and tetrasaccharides (lactose)		264,274,275
		78	bgaB	Q45093	monomer			*p*NPGal (V_{max} 46.4 U·mg^{-1}) Lactose (V_{max} 45.3 U·mg^{-1}) *o*NPGal (V_{max} 11.8 U·mg^{-1})	Tri- and tetrasaccharides (lactose)		
Bifidobacterium adolescentis (strain ATCC 15703/DSM 20083/ NCTC 11814 /E194a)		78	bgaB	Q5J883	tetramer	6.0	35⁰C	*p*NPGal (K_M 2.2 mM; V_{ma} 5 U mL^{-1}) β-D-Galp(1,4)- β-D-Lactose (K_M 2.2 mM; V_{ma} 93 U·mL^{-1}) β-D-Galp(1,4)- β-DGal (K_M 3.7 mM; V_{max} 95.3 U·mL^{-1})	From di- up to octamers		276,277
Bifidobacterium adolescentis (strain ANB-7)				A1A399		7.5	37⁰C	*o*NPGal (V_{ma} 23.81 U·mL^{-1}) Lactose (V_{max} 133.01 U·mL^{-1})	43% Gos (30%, wt/wt, lactose)		278

Bifidobacterium bifidum NCIMB 41171		77.3	BbgII	A5A2I5		5.4-5.8	40^{0}C		20% (w/w) GOS: 25% di-, 35% tri. 25% tetra- and 15% pentasaccharides(50%, wt/wt, lactose)		224,226,227,279,280
Bifidobacterim longum sbsp. infantis		80	Blon 2123	B7GUN8		6.0	40^{0}C		48% oligosaccharides whole cell Differents GOS with TLC		222,278,281,282
Bifidobacterim longum sbsp. longum	*BRC1 5708*	77	lacA1	B3DTA3	dimer	6.8	45^{0}C		33% GOS (40% lactose)		283
	JCM 7052					7.0	50-55^{0}C	*p*NPGal (K_M 0.42 mM; V_{max} 5.21 µmol·min^{-1}·mg prot) *o*NPGal (K_M 0.8 mM; V_{max} 2.16 µmol·min^{-1}·mg prot)	(*p*NPGal as donor; Gal or Mel as acceptor)		284
Caldicellulosiruptor saccharolyticus (stran ATCC 43494/ DSM 8903/ Tp8T 6331)		79.2	Csac_1018	A4XI97	trimer	6.0	80^{0}C	*p*NPGal (K_M 0.11 mM; k_{cat} 32 s^{-1}) *o*NPGal (K_M 1.21 mM; k_{cat} 149 s^{-1}) Galactobiose (K_M 30.0 mM; k_{cat} 42 s^{-1})	0.171 g/L lactulose (1 mM *p*NPGal, 50 mM Gal)	F349S (reduce galactose inhibition)	285,286
						7.5			408 g/L lactulose , yield 74% (700 g/L lactose)		287
Geobacillus kaustophilus (Geobacillus starothermophilus)		78	bgaB	P19668		6.5	37^{0}C	Lactose (K_M 114 mM; V_{max} 0.05 U/mg) Lactulose (K_M 160 mM; V_{max} 0.03 U/mg)	82 mM (23% wt/wt) Gal-β-1,3-Lac ; 30% GOS (500 mM lactose)	R109W	288,289
Klebisella pneumoniae				F2X2W6		8.0	50^{0}C	*o*NPGal (K_M 0.82 mM; k_{cat} 48 s^{-1})	23.8 g/L of 1-lactulose (400 g/L lactose and 200 g/L fructose)		232–234
Thermus neapilitana		75		B9K9C2 = O85248 (1KWG)					17%-39% GOS (cellobiose and lactose as donors. Different acceptors)		

A.2 Sequences

A.2.1 Genes subcloned into *pET-22b(+) vector*

All genes were sublconed on pET-22b(+) vector and they were subcloned using *NdeI* and *HindIII* restriction sites. Map and MCS are shown below.



All genes included *NdeI* and *BamHI* on 5' and *HindIII* and *XhoI* on 3' extreme. Genes higher that 3kbp also include a restriction site in the middle of the sequence.

Bacillus circulans

GGATCCACATATGGGTAATAGCGTTAGCTATGATGGTGAACGTCGTGTGAATTTTAACGAAAATTGGCGTTTTCAGCGC
GAAACCAATGGTAGCATTGCCGGTGCACAGAATCCGGGTTTTGATGATAGCAGCTGGCGTAAACTGAATCTGCCGCATG
ATTGGAGCATTGAACTGGATTTTAACAAAAACAGCCTGGCAACCCATGAAGGTGGTTATCTGGATGGTGGTATTGGTT
GGTATCGTAAAACCTTTACCATTCCGGAAAGCATGAAAGGTAAACGTATTAGCCTGGATTTTGATGGCGTGTATATGA
ATAGCACCACCTATCTGAATGGTGAAGTTCTGGGCACCTATCCGTTTGGTTATAATGCCTTTTCCTATGATATCAGCGA
TAAACTGTATAAAGATGGCCGTGCAAATGTTCTGGTTGTGAAAGTGAATAATACCCAGCCGAGCAGCCGTTGGTATAGC
GGTAGCGGTATTTATCGTAATGTTTATCTGACCGTTACCGATCCGATTCATGTTGCACGTTATGGCACCTTTGTTACCA
CCCCGAATCTGGAAAAAAGCATTAAAGAAGATCGCGCAGACGTCAACATCAAAACGAAAATCAGCAATGATGCAGCAG
AAGCCAAACAGGTTAAAATCAAAAGCACCATTTATGATGGCGCAGGTAATACCGTTCAGACCGTTGAAACCGAAGAAA
AAACCGCAGCAGCAGGCACCGTTACCCCGTTTGAACAGAATACCGTTATTAAACAGCCGAAACTGTGGTCCATTGATAA
ACCGTATCGTTATAATCTGGTGACCGAAGTTATTGTTGGTGGTCAGACCGTGGATACCTATGAAACAAAATTTGGTGTG
CGCTACTTCAAATTCGATGAAAATGAAGGCTTTAGCCTGAACGGCGAATACATGAAACTGCATGGTGTTAGCATGCATC
ATGATCTGGGTGCACTGGGTGCCGCAACCAATGCACGTGGTGTTGAACGTCAGATGCAGATTATGAAAGATATGGGTGT
TAATGCCATTCGCGTTACCCATAATCCGGCAAGTCCGGAACTGCTGGAAGCAGCAAATAAACTGGGTCTGTTTATTATC
GAAGAAGCCTTTGATAGCTGGGCACAGAGCAAAAAACCGTATGATTATGGTCGCTTTTTTAACGCATGGGCAGAACACG
ATATCAAAGAAATGGTTGATCGCGGTAAAAATGAACCGGCAATTATCATGTGGTCAATCGGCAACGAAATTTATGATA
CCACAAATGCAGCCGGTGTGGAAACCGCACGTAATCTGGTTGGTTGGGTTAAAGAAATTGATACCACCCGTCCGACCAC
CATTGGTGAAGATAAAACCCGTGGTGATAAAGTTAATGTGACCCCGATTAACAGCTACATCAAAGAGATCTTCAACATC
GTTGATGTGGTGGGTCTGAATTACAGCGAAAACAACTATGATGGCTACCATAAACAGAACCCGAGCTGGAAACTGTAT
GGTAGCGAAACCAGCAGCGCAACCCGTAGCCGTGGTGTTTATACCCATCCGTATCAGTATAATCAGAGCACCAAATATG
CCGATCTGCAGCAGAGCAGCTATGATAATGATTATGTGGGTTGGGGTCGTACAGCCGAAGATGCATGGAAATATGATC
GTGATCTGAAACATATTGCCGGTCAGTTTATCTGGACAGGCTTTGATTATATTGGTGAACCGACCCCGTATTATAACAG
CTATCCGGCAAAAAGCAGCTATTTTGGTGCAGTTGATACCGCAGGTTTTCCGAAAGATATTTTCTATTATTACCAGAGC
CAGTGGAAAAAAGAACCGATGGTTCATCTGCTGCCGCATTGGAATTGGAAAGAAGGTGAAAAAGTTCGCGTTCTGGCA
TATACCAATGCAAGCAAAGTTGAACTGGTTCTGAATGGCGAAAGCCTGGGTGAGAAAAACTATGATAACAAACAAACC
AGCTGGGGTGCACCGTATAAAGAAACCAAAGATGGTAAAACGTATCTGGAATGGGCAGTTCCGTTTAAACCGGGTAAA
CTGGAAGCCGTTGCAAAAGATGAAAACGGTAAAGTTATTGCCCGTGATCAGGTTGTTACAGCAGGCGAACCGGCATCAG
TTCGTCTGACCGCAGATCGTAAAGTTGTTAAAGCAGATGGCACCGATCTGAGCTTTATTACCGCAGATATTGTTGATAG

CAAAGGTATTGTTGTTCCGGATGCAGATCATCTGATTACCTTTAATGTGACCGGTCAGGGTGAACTGGCAGGCGTTGAT
AATGGTAATGCAAGCAGCGTTGAACGCTACAAAGATAACAAACGTAAAGCCTTTAGCGGTAAAGCACTGGCAATTGTT
CAGAGCAGCAAACTGAGCGGTAAAATTACCGTTCATGCAAGCGTTGCAGGTCTGAGCAGCGATAGCACCAGCGTTTTTA
CCGTTACACCGGCAGATCATGACAAAAAAATCGTTGCCGGTATTGATGATGTTAACCTGACCGTTGATGTTAATGAAGC
ACCGAAACTGCCGAGCGAAATCAAAGTGTATTATAGTGATGAAAGCGCAGCAGCCAAAAATGTTACCTGGGATGAAGT
TGATCCGAAACAGTATAGCACCGTTGGTGAATTTACCGTTGAAGGTAGCGTTGAAGGCACCAGCCTGAAAGCAAAAGCA
TTTGTTATTGTGAAAGGTATCGTTGCGGTTAAACCGTATAGTACCGCAACCAAAGTTGGTGTTCAGCCGGTTCTGCCGG
AAAAAGCAACCCTGCTGTATTCAGATGGCACCACCAAAGGTGCAACCGTGACATGGGATGAAATTCCGGAAGATAAACT
GGCAAAAGAAGGTCGTTTTACAGTCGAAGGTTCAGTGGAAGGCACCGATCTGAAAGCCAATGTTTATGTTCGTGTTACC
AACGAAGTGAAAAGCGTGAATATTATGCTGCAAGAACAGGGTAGCGCATATCCGAAACTGGAAGCAACCTTTACCAAT
CCGGCAGATAATCTGCAGCATCTGAATGATGGTATCAAAAGCTATACCAATAATCCGGTTAATCGCTGGACCAATTGGA
CCCGTACACCGCGTGATGCCGGTGATAGCATTACCGTTAATTTTGGCAAAAAACACGTGATCAATAACCTGGACCTGTT
TGTGTTTACCGATAGCGGCACCGTTGTTCCAGAAAAAGCCGAAGTTCAGTATTGGGATGGCACCGCATGGAAAGATGTT
GAAAATCTGACCCAGCCGAGCCCGTATGTTGTTGAAAAAAATGAACTGACCTTTGATGCAGTGGCAACCGAAAAACTGA
AATTCCATCTGACCCCGAGCGTTAAAGGCAAATTTCTGGCACTGACCGAAGCGGAAGTTTATGCAGATCAGATTGTTAT
GGGTGAAACCGCAAAACTGCAGAGTATTACCGTGAATGGTAAAGCCCTGGAAGGTTTTGATCATGCGAAAAAAAACTA
TGAACTGGTGCTGCCGTATGGTAGCGAACTGCCGAAAATTGAAGCAGCAGCCGCAGATAATGCAACCGTTACCATTCTG
CCTGCATTTAGCTATCCGGGTACAGCAAAACTGTTTGTGACCAGCGAAGATGGTAAAGTTACCACCGAATATAGCATTG
GTGTTAGCACCGAAGAACCTAAACTGGTTAGCAAGCTTTAACTCGAG

B. longum

GGATCCACATATGCGTGCACGTCGTGATTTTGCATGGCCGAAACTGCTGACCGCAGATGGTCGTGGTATTGCATTTGGT
GGTGATTATAACCCGGATCAGTGGCCTGAAGATGTTTGGGATGATGATATTCGTCTGATGGGTCAGGCAGGCGTTAAT
ACCGTTGCACTGGCAATTTTTAGCTGGGATCGTCTGCAGCCGACCGAAGATCATTGGAATTTTGATTGGCTGGATCGCA
TCATTGATAAACTGGGCCAGGCAGGTATTGCAGTTGATCTGGCAAGCGCAACCGCCACCGCACCGCTGTGGCTGTATGA
AAATCATCCGGAAGTTCTGCCTCGTGATAAATATGGTCATCCGGTTAATGCAGGTAGCCGTCAGAGCTGGTCACCGACC
AGTCCGGTTTTTAAAGAATATGCACTGACCCTGTGTCGTAAACTGGCAGAACGTTATGGCACCAATCCGTATGTTACCG
CATGGCACATGGGTAATGAATATGGTTGGAATAACCGTGATGATTATAGCGATAATGCACTGGAAGCATTTCGTGCAT
GGTGTCGTCGTAAATATGGCACAATTGATGCACTGAATCAGGCATGGGGCACCACCTTTTGGGGTCAAGAAATGACCGG
TTTTGATGAAGTTCTGATTCCGCGTTTTATGGGTGCAGATAGCATGGTTAATCCGGGTCAGAAACTGGATTTTGAACGC
TTTGGTAATGATATGCTGCTGGATTTCTATAAAGCAGAGCGTGATGCAATTGCCGAAATTTGTCCGGATAAACCGTTTA
CCACCAACTTTATGATTAGCACCGATCAGTGTTGTATGGATTATGCAGCATGGGCAGAAGAAGTGAATTTTGTTAGCAA
CGATCACTATTTTCACGAAGGTGAAAGCCATCTGGATGAACTGGCATGTTCAGATGCACTGGTTGATAGCCTGGCCCTG
GGTAAACCGTGGTATGTTATGGAACATAGCACCAGCGCAGTTCAGTGGAAACCGCTGAATACCCGTAAACGTAATGGTG
AAACCGTTCGTGATAGTCTGGCACATGTTGCAATGGGTGCCGATGCAATTAACTTTTTTCAGTGGCGTGCAAGCGCCTT
TGGTGCAGAAGCCTTTCATAGCGCAATGGTTCCGCATGCCGGTGAAAATACCAAACTGTTTCGTCAGGTTTGTGAACTG
GGTGCAACCCTGCAGGCACTGGCAGATGCCGGTGTTCAGGGTAGCGAACTGGCCCATGCAGATACCGCAATTCTGTTTA
GCGCAGAAAGCGAATGGGCAACCCGTAGCGAAACCCTGCCGAGCATGAAACTGAATCATTGGCATGATGTTCGTGATTG
GTATCGTGCATTTCTGAATGCCGGTGCACGTGCCGATATTGTTCCGCTGGCCTATGATTGGAGCAGCTATAAAACCATT
GTTCTGCCGACCGTTCTGATGCTGAGTGATGCCGATACCCGTCGTCTGGCAGGTTTTGTTCAGGATGGTGGTCGTGTTG
TTGTTGGTTATGCCACCGGTCTGCTGGATGAGCGTTTTCATACCTGGCTGGGTGGCTATCCGGGTGCCGGTGATGGTCT
GCTGCGTAGCATGCTGGGTGTTCGTGGTGAAGAATTTAACATTCTGGGCACCGAAACCGAAGATGAACCGAGCGAAATT
CGTCTGGCCAGCACCGGTGATAGCCCGACCATGGATGGTGCAGTTACCCGTCTGTGGCAGAATGATGTTACCGTTGCCG
GTCCGCATGTGCAGGTTCTGGCAGCCTATGCGGGTGAAGAGGCAAATGAATGGGAACTGGATGGCACCGCAGCAATTAC
CCGTAATACCTATGGCGAAGGCGAAGCATATTTTCTGGGTTGCGATCTGGGTGTTAGCGATCTGACCCGTTTTGTTGGT
GGTTGGCTGGCAGCACGTCCGCAGGATGGCCGTCAGCCGGAAGCAAATCTGCGTAGTCCGGCAAGCGGTGTTACCAGTG
ATGTTCTGCATACCGTTCGTCAGAGTGATGATGCCATCTTTGATTTTTATCTGACGCGTGGTAAATCCGATGTTGAACT
GCGTGATATTGCAGGCGAACCGATTGTGCTGTTTCGTGCAGAACGTGGTAGTGATGGTGGTGCATATACCGTTCATCGT
AATGGCGTTCTGGTTATGAAACGTCCGAATCCGAGCGTTAAGCTTTAACTCGAG

[redacted]

G. kaustophilus

GGATCCACATATGAATGTTCTGAGCAGCATTTGTTATGGTGGTGATTATAACCCGGAACAGTGGCCTGAAGAAATTTG
GTATGAAGATGCAAAACTGATGCAGAAAGCCGGTGTTAATCTGGTTAGCCTGGGTATTTTTAGCTGGTCAAAAATTGA
ACCGAGTGATGGCGTGTTTGATTTTGAATGGCTGGATAAAGTGATCGACATCCTGTATGATCATGGCGTGTATATTAAC
CTGGGCACCGCAACCGCAACCACACCGGCATGGTTTGTGAAAAAATACCCGGATAGCCTGCCGATTGATGAAAGCGGTG
TGATTCTGAGCTTTGGTAGCCGTCAGCATTATTGTCCGAATCATCCGCAGCTGATTACCCATATTAAACGTCTGGTTCG
TGCCATTGCAGAACGCTATAAAAACCATCCGGCACTGAAAATGTGGCATGTGAATAATGAATATGCCTGCCATGTGAGC
AAATGCTTTTGCGAAAATTGTGCAGTTGCCTTTCGTAAATGGCTGAAAGAGCGTTATAAAACCATCGATGAACTGAAT
GAACGTTGGGGCACCAATTTTTGGGGTCAGCGCTATAATCATTGGGATGAAATTAATCCGCCTCGTAAAGCACCGACCT
TTATTAATCCGAGCCAAGAGCTGGATTATTACCGCTTTATGAATGACAGCATCCTGAAACTGTTTCTGACCGAAAAAGA
AATTCTGCGTGAAGTGACACCGGATATTCCGGTTAGTACGAATTTTATGGGTAGCTTTAAACCGCTGAACTATTTCCAG
TGGGCACAGCATGTTGATATTGTTACCTGGGATAGTTATCCTGATCCGCGTGAAGGTCTGCCGATTCAGCATGCAATGA
TGAATGATCTGATGCGTAGCCTGCGTAAAGGTCAGCCGTTTATTCTGATGGAACAGGTTACCAGCCATGTTAATTGGCG
TGATATCAATGTTCCGAAACCGCCTGGTGTTATGCGTCTGTGGTCTTATGCAACCATTGCACGTGGTGCAGATGGCATT
ATGTTTTTTCAGTGGCGTCAGAGCCGTGCCGGTGCAGAAAAATTTCATGGTGCCATGGTTCCGCATTTTCTGAATGAAA
ACAATCGCATTTATCGCGAAGTTACACAGCTGGGTCAAGAGCTGAAAAAACTGGATTGTCTGGTGGGTAGCCGTATTAA
AGCAGAAGTTGCAATTATCTTCGATTGGGAAAACTGGTGGGCAGTTGAACTGAGCAGTAAACCGCATAATAAACTGCG
CTATATTCCGATTGTGGAAGCCTATTATCGTGAGCTGTATAAACGTAATATCGCCGTTGATTTTGTTCGTCCGTCTGAT
GATCTGACCAAATACAAAGTTGTTATTGCACCGATGCTGTATATGGTGAAAGAAGGTGAGGACGAAAATCTGCGTCAG
TTTGTTGCAAATGGTGGCACCCTGATTGTTAGCTTTTTTAGCGGTATTGTGGATGAGAATGATCGTGTTCATCTGGGTG
GTTATCCGGGTCCGCTGCGTGATATTCTGGGCATTTTTGTTGAAGAATTTGTGCCGTATCCGGAAACCAAAGTCAACAA
AATCTATAGCAACGATGGCGAGTATGATTGTACCACCTGGGCAGATATTATTCGTCTGGAAGGTGCAGAACCGCTGGCA
ACATTTAAAGGTGATTGGTATGCAGGTCTGCCTGCAGTTACCCGTAATTGCTATGGTAAAGGTGAAGGTATTTATGTT
GGCACCTATCCGGATTCAAATTATCTGGGTCGTCTGCTGGAACAAGTTTTTGCCAAACATCATATTAACCCGATTCTGG
AAGTTGCCGAAAATGTTGAAGTTCAGCAGCGTGAAACCGATGAATGGAAATATCTGATCATCATCAACCACAACGATT
ATGAAGTTACCCTGAGCCTGCCAGAGGACAAAATCTACCAGAATATGATTGATGGCAAATGTTTTCGTGGTGGTGAACT
GCGTATTCAGGGTGTTGATGTTGCAGTGCTGCGTGAACATGATGAAGCAGGTAAAGTTAAGCTTTAACTCGAG

L. delbrueckii

GGATCCACATATGAGCAACAAACTGGTGAAAGAAAAACGTGTTGATCAGGCAGATCTGGCATGGCTGACCGATCCGGA
AGTTTATGAAGTTAATACCATTCCGCCTCATAGCGATCATGAAAGTTTTCAGAGCCAAGAAGAACTGGAAGAAGGTAA
AAGCAGCCTGGTTCAGAGCCTGGATGGTGATTGGCTGATTGATTATGCAGAAAATGGTCAGGGTCCGGTTAACTTTTAT
GCCGAAGATTTTGATGATAGCAACTTTAAAAGCGTGAAAGTTCCGGGTAATCTGGAACTGCAGGGTTTTGGTCAGCCGC
AGTATGTTAATGTTCAGTATCCGTGGGATGGTAGCGAAGAAATTTTTCCTCCGCAGATTCCGAGCAAAAATCCGCTGGC
AAGCTATGTTCGTTATTTTGATCTGGATGAAGCCTTCTGGGATAAAGAAGTTAGCCTGAAATTTGATGGTGCAGCAACC
GCCATTTATGTTTGGCTGAATGGTCATTTTGTGGGCTATGGTGAAGATAGCTTTACCCCGAGCGAATTTATGGTTACCA
AATTTCTGAAAAAAGAGAACAATCGTCTGGCAGTTGCCCTGTATAAATACAGCAGCGCAAGCTGGCTGGAAGATCAGG
ATTTTTGGCGTATGAGCGGTCTGTTTCGTAGCGTTACCCTGCAGGCAAAACCGCGTCTGCATCTGGAAGATCTGAAACT
GACCGCAAGCCTGACCGATAACTATCAGAAAGGTAAACTGGAAGTGGAAGCCAATATTGCATATCGTCTGCCGAATGCA
AGCTTCAAACTGGAAGTTCGTGATAGCGAAGGTGATCTGGTTGCAGAAAAACTGGGTCCGATTCGTAGCGAACAGCTG
GAATTTACCCTGGCCGATCTGCCGGTTGCAGCATGGTCAGCGGAAAAACCGAATCTGTATCAGGTTCGTCTGTATCTGT
ACCAGGCAGGTAGCCTGCTGGAAGTGAGCCGTCAAGAAGTTGGTTTTCGTAATTTTGAACTGAAAGATGGCATCATGTA
TCTGAATGGCCAGCGCATTGTTTTTAAAGGTGCAAATCGTCACGAGTTCGATAGCAAACTGGGACGTGCAATTACCGAA
GAAGATATGATTTGGGACATCAAAACCATGAAACGCAGCAATATTAACGCAGTTCGTTGTAGCCATTATCCGAATCAG
AGTCTGTTTTATCGTCTGTGCGATAAATATGGCCTGTATGTTATTGATGAAGCCAACCTGGAAAGCCATGGCACCTGGG
AAAAAGTTGGTGGTCATGAAGATCCGAGCTTTAATGTTCCGGGTGATGATCAGCATTGGCTGGGTGCCAGCCTGAGCCG
TGTGAAAAATATGATGGCACGTGATAAAAACCATGCCAGCATTCTGATTTGGAGCCTGGGTAATGAAAGCTATGCAGG
CACCGTTTTTGCACAGATGGCAGATTATGTTCGTAAAGCAGATCCGACCCGTGTTCAGCATTATGAAGGTGTTACCCAT
AACCGCAAATTCGATGATGCAACCCAGATTGAAAGCCGTATGTATGCACCGGCAAAAGTGATTGAAGAATACCTGACCA
ATAAACCTGCCAAACCGTTTATTAGCGTTGAATATGCACATGCCATGGGTAATAGCGTTGGCGATCTGGCAGCATATAC
CGCACTGGAAAAATATCCGCATTATCAGGGTGGTTTTATCTGGGATTGGATTGATCAGGGTCTGGAAAAAGATGGTCA
TCTGCTGTATGGTGGTGATTTCGATGATCGTCCGACCGATTATGAATTTTGTGGTAATGGTCTGGTGTTCGCCGATCGT
ACCGAAAGCCCGAAACTGGCAAATGTTAAAGCACTGTATGCAAACCTGAAACTGGAAGTTAAAGATGGTCAGCTGTTCC
TGAAAAACGATAACCTGTTTACCAATAGCAGCAGCTATTATTTCCTGACCAGCCTGCTGGTTGATGGTAAACTGACCTA
TCAGAGCCGTCCGCTGACCTTTGGTCTGGAACCGGGTGAAAGCGGCACCTTTGCACTGCCGTGGCCTGAAGTTGCAGAT
GAAAAAGGTGAAGTTGTTTATCGTGTTACCGCACACCTGAAAGAAGATCTGCCTTGGGCAGATGAAGGTTTTACCGTTG
CAGAAGCAGAAGAAGTTGCACAGAAACTGCCGGAATTTAAACCGGAAGGTCGTCCGGATCTGGTTGATAGCGATTACA
ATCTGGGTCTGAAAGGCAACAATTTTCAGATCCTGTTCAGCAAAGTTAAAGGTTGGCCTGTTAGCCTGAAATATGCAGG
TCGTGAATATCTGAAACGTCTGCCTGAATTTACCTTTTGGCGTGCACTGACCGATAATGATCGTGGTGCAGGTTATGGT
TATGATCTGGCACGTTGGGAAAATGCAGGTAAATATGCCCGTCTGAAAGATATTAGCTGCGAAGTAAAAGAAGATAGC
GTCCTGGTTAAAACCGCATTTACCCTGCCGGTTGCACTGAAAGGTGATCTGACCGTTACCTATGAAGTTGATGGTCGTG
GTAAAATTGCAGTGACCGCAGATTTTCCGGGTGCCGAAGAAGCAGGTCTGCTGCCTGCATTTGGTCTGAATCTGGCACT
GCCGAAAGAACTGACAGATTATCGTTATTATGGTCTGGGTCCGAATGAAAGCTATCCGGATCGTCTGGAAGGTAATTAT
CTGGGTATTTATCAGGGTGCCGTGAAAAAAAACTTTAGCCCGTATCTGCGTCCGCAAGAAACCGGTAATCGTAGCAAAG
TTCGTTGGTATCAGCTGTTTGATGAGAAAGGTGGCCTGGAATTTACAGCAAATGGTGCAGATCTGAACCTGAGTGCCCT
GCCGTATAGCGCAGCACAGATTGAAGCAGCAGATCATGCATTTGAACTGACCAATAACTATACCTGGGTTCGTGCCCTG
AGCGCACAGATGGGTGTTGGTGGTGATGATAGCTGGGGTCAGAAAGTTCATCCGGAATTTTGTCTGGATGCCCAGAAA
GCACGTCAGCTGCGTCTGGTTATTCAGCCGCTGCTGCTGAAAAAGCTTTAACTCGAG

P. thiaminolyticus

GGATCCACATATGACCACCCTGAGCTATGATGAAGGTCAGTTTAAAATGGGTGATCGTCCGATTCAGCTGATTAGCGGT
GCAATTCATTATTTTCGTATCGTTCCGGCATATTGGGAAGATCGTCTGCGTAAAATCAAAGCCATGGGTTGTAATTGCA
TCGAAACCTATGTTGCCTGGAATGTGCATGAACCGCGTGAAGGTGAATTTCATTTTGAACGTATGGCAGATGTTGCCGA
ATTTGTTCGTCTGGCAGGCGAACTGGGTCTGTATGTTATTGTTCGTCCGAGCCCGTATATTTGTGCAGAATGGGAATTT
GGTGGTCTGCCTGCATGGCTGCTGAAAGATGATATGCGTCTGCGTTGTAATGATCCGCGTTTTCTGGAAAAAGTTAGCG
CATATTATGATGCACTGCTGCCGCAGCTGACACCGCTGCTGGCAACCAAAGGTGGTCCGATTATTGCAGTTCAGATCGA
AAATGAATATGGCAGCTATGGTAATGATCAGGCATATCTGCAGGCACAGCGTGCAATGCTGATTGAACGTGGTGTTGA
TGTTCTGCTGTTTACCAGTGATGGTCCGCAGGATGACATGCTGCAGGGTGGTATGGCCGAAGGTGTTCTGGCAACCGTT
AATTTTGGTAGCCGTCCGAAAGAAGCATTCGATAAACTGAAAGAATATCAGCCGGATGGTCCTCTGATGTGTATGGAA
TATTGGAATGGTTGGTTTGATCACTGGTTTGAACCGCATCATACCCGTGATGCAAAAGATGCAGCACGTGTTCTGGACG
ATATGCTGGGTATGGGTGCAAGCGTTAACTTTTATATGGTTCATGGTGGCACCAATTTTGGTTTTGGTAGTGGTGCAAA
TCACAGCGATAAATATGAACCGACCGTTACCAGCTATGATTATGACGCAGCAATTAGCGAAGCCGGTGATCTGACCCCG
AAATATCATGCATTTCGTGAAGTGATTGGCAAATATGTTAGCCTGCCGGAAGGTGAACTGCCTGCAAATACCCCGAAAG
CAGATTATGGTAGCGTTCCGGTTACCCGTCGTGTTAAACTGTTTGATACCCTGGAACCGATGACCGAAGTTAAAGATAG
CATTTGTCCGGAACCTATGGAAAAATATGGTCAGAACAATGGCTTCATCGTTTATAGCACCCGTATTAGCGGTCCGCGT
CCGGAAAGCCGTCTGACCATTCAGGATGTTCGTGATCGTGCACTGGTTTTTCTGGATCGTAAACTGATTGGTGTTGTGG
AACGTTGGAATCCGCAGAGTCTGCCGGTTACCGTTCCGGAAGATGGTGCACAGCTGGATATTCTGGTTGAAAATATGGG

TCGTGTGAATTATGGTCCTCAGCTGTATGATCGTAAAGGTATTACCCATGGTGTGCGTCTGAATGGCCAGTTTCTGTTT
CATTGGGAAGTGCGTAGCCTGGAACTGGAAACCCTGGCAGGTCTGAGTTTTGATACCGCAGGCGCACGTGCATGGGAAG
AAGAACAGCCTGGTTTTTATGAAGCCAAACTGGTGATTGAAGATGAGCCGAAAGATACCTTTCTGCGTCTGGAAGGTT
GGAAAAAAGGTGTTGTTTTTGTGAACGGCTTTAACCTGGGTCGTTATTGGGAGGTGGGTCCGCAGCAGGCACTGTATGT
TCCAGCACCGGTTCTGCGTCAGGGTGAAAACGAAATTGTTGTGTTTGAACTGCACCGTGAAGGCGAACCGCTGCGTTTT
GAAGCAGCACCGAGCCTGAGCAAAAAGCTTTAACTCGAG

S. rectivirgula

GGATCCACATATGCGTCGTCGTATGCATCCGGCAGTTCTGCTGGCAACCCTGGCAGTTACCGCAGGTCTGCTGCTGCCTG
CAGCACATGCACAGCCGGTTGATGGTCCGAGCGATCCGGCAGAAATTGAAAGCTATCTGGAAGATCCGCAGCGTACCGG
TGAAGGTCAGCAGCCTCCGCATGCATTTCTGCGTCCGTATGCAGATGCAGAACAGGCACTGCGTAGCGCACGTGAGGAC
CTGTTTGCACCGGCAGATGAACCGACCCCGTGGACCCTGAGCCTGAATGGTCGTTGGCAGTTTCAGTATGCCGATCATT
ATCGTGATCTGCCTGCCGGTTGGCAGCGTGGTAGCGGTGAATGGGATGCAGTTGATGTTCCGGGTGTTTGGCAGCAACA
GGGTTATGATCGTCCGATTTATCGTAATGTTAGCAGCGAAATTGCACCGTATCGTCCGCCTGAAGTTCCGGATGATATC
AACCCGACCGGTGCATATCGTAAACAGTTTGAACTGCCTGAACATTGGGCAGGTCGTCGTCAGCTGCTGCGTTTTGAAG
GTGTTACCAGCGGTTATTTTGTTTGGGTGAATGGTCATTATGTGGGCTATGATCAGGGTGGTTATACCCCTGCAGAATT
TGATGTTACACCGCATCTGCGTCCTGGCACCAATACCATTGCAGTTCAGGTTCATCGTTGGAGCAGCGGTAGTTATCTG
GAAAATATGGATTTTTGGCACTTCAGCGGTATTTTTCGTGAAGTTCATCTGTATAGCGTTCCGCGTACCCACCTGGAAG
ATGTTACCGTTCGTACCGAACTGGATGATACCTATACCGATGCAACCCTGCGTCTGGGTGCCGAACTGCGTCATCTGCC
AGGCGGTACAACCGGTCGTCATGAACTGCGTGCCACCCTGTATGATCCGGATGGTAATCGTGTTGCAGAAGTTAGCGAA
CCGGTTGTTGTTGATCGTGAAACCGCAAGCGCAGAACTGGCACAGCATGTTGCAGCACCGCGTCTGTGGTCAGATGAAA
CTCCGCACCTGTATACCGCAGTTCTGGAACTGCTGCATAATGGTCAGGTTGTTCATACCACCCAGCAGCCGGTGGGTTT
TCGTAGCGTTGAAGTTCGTGATGAACAACTGCTGCTGAATGGCAAACCGGTTGAGCTGCGTGGCGTTAATCGCCATGAA
CATGATCCGCGTACAGGTCGTGCAGTTAGCCGTCAGCGCCAGCAGCAGGATGTTGAACTGCTGCGTCAGCATAACATTA
ATGCAGTTCGCACCAGCCATTATCCGAATGATCCGTATTGGTATCGTCTGGCAGATCATCATGGTATTCTGCTGGCAGA
TGAAGTTGATGTTGAAACCCATTATCGTGAGGATTGTAGCAATCCGGCAGATGATTGTCTGGCCGATCGTCCGGAATGG
CAGGCAGCATTTGCAGATCGTTTTCATGCACTGCTGGAACGCGATAAAAATCATCCGAGCGTGATTATTTGGGATACCG
GTAATGAAGCAGGTCTGGGTGCAGCACATTATGCAATGGCAGAACATGCACGTGCCACCGATCCGACCCGTCCGCTGTA
TCATCAGAGCAATAGTCCGGATGGTGATGCACCGTATGCAGATATTTGGGGTCCGCGTTATCCGAGTCCGCAGCGTCTG
GAAGAAATTGCAGATGAAACCACCAAACCGGTTGTTATGGGTGAATGGCTGCATGCAATGGGTAATAGCCTGGGTCAT
TATGAAGATATGTGGCGTACCATTCGTCGTGAACCGGCACTGCAGGGTGGTTTTGTTTGGGATTGGGTTGATCAGGGTC
TGTATCGTCCGCTGCGTACCACACCGGCAAATATTCCGGTTCATTATGGTGGTAATCCGCGTCCGGTTCCGGGTGTTAG
CGGTAATGGTCTGCTGCTGAGCGGTCTGGATGATTGGGTGGAAGCATATCGCGATCCGGCACTGGATATTACCGGTCGT
CAGGTTACCCTGGATATGTGGGTTAAACCGCTGAGCTGGCAGGGTAGCGGCACCTTTCTGAGCAAAGGTGATAAACAGT
GGGCTCTGCAGATGCCGGATCCTGAACATGTTGAATTTTTTGTGTATGGCGAAGATGATTGGCATACCGCACGTGTTCG
TATTCCGGATGACTGGTGGGGTAACTGGCATCGTCTGAGCGGTGTTTATGATGGTGAACAGGTTCGTCTGCATCTGGAT
GGCGAGCAGGTTGCAGCAACCCCGTATCGTGGTGAAATTGCGGCAAATACCATGTATACCGCAAGCATTGGTCGCAACC
AAGAAAAACATGGTGATAGCTTTGCAGGTCGTACCGCACATGCAGTTGTTGATAGCGTTCGTGTGTATGATACCGCACT
GACCCGTCAGCAGCTGGAAGATGATCCTGCAGAACGTGCGGTTCTGGCCCTGGATTTTGAAACCGAACAGCAGAATGGT
GAATTTCTGGATTATGGCAGCAGCAATTTTGTTGTGAATGGTCTGGTTAATGCCGATCGTACACCGCAGCCGGAACTGG
CACAGCTGGCATATAGCCATGCACCGGTTCGTTTTGAAGGCACCGGTGAACCGGGTGAATTCCGTGTTCATAATCGTAA
TCATACCCTGACCACCGAAGCATATGATCTGACCTGGCGTCTGGTTGAAGGTGGTCGTGATATTGCAAGCGGTGCACTG
GATGTTGTTGTTGGTCCGGGACAGCAGCAACTGGTTGAACTGGATCTGCCTCCGGCAGGCGGTACACAGCGTTATCTGG
TTCTGGAAGCACGTCAGCGTGCAGATGCACCGGGTGTTCCGCGTGGTCATCTGGTTGCAACAGAACAGCTGCCTGCGGG
TGGCACCCAGCCTCCGCCTCCACCTGAAAGCGGTGGTGAACCGCAGCCTCTGGCCGTGCATGAACATGGTAATGCCGTTG
AAGTTAGCGCACCGGATTTCACCTATACCTTTGATCGTGAAGCAGGCACCCTGACCGGTATGACCGCAAATGGTCAGCA
ACATCTGGTTCGTGGTCCGCAGCTGGATGTTTTTCGTGCACCGATTGGTAATGAATGGTCAGATTGGAGTGGTCTGGCA
CCGGAAGCACAGTTTCGTGCAGTTGGTCTGGATCGTCTGCGTACCGAAGTTACCGATTTTACCGTTAGCCAGCCTGCAC
CGGATCAGGTTCGTGTTAGTGTTCGTACCGATGTTTCAGCACCGGATGTGCCGGATGATGGTTTTCGTAGTACCTGGGT
TTATACCGTTGATGGTACAGGTGCAATTACCATTGATCATCAGGTTGATGCATATGGTGAACGTATGCGTGCACTGCCG
TGGCTGCCTCGTGTTGGTATGAGCATGGCAGTTCCGGATAGTTTTGGTCAGCTGGATTGGTATGGTCGCGGTCCGGGTG
AAAGCTATCCGGATCGTAAAGATGCACAGCATATTGGTCGTTGGAGCCAGACCGTGGATGAAAGCTGGTTTGATTTTCT
GCCACCGCAGGATAATGGTGTTAAAACCGATACCTGGTGGGCAGTTCTGAGTGGTGATCGTGGTGGTCTGCAGGTTTCA
GGTGATCGCCTGGCAATTGCAGCGGATCGTTTTAGCAATGCAGAACGCACCGATTTTGCACATCAGCTGCGTCGTGATG
ATTTTGTTACCCTGCATGTTAGCGCAGCAGTTACCGGTCTGGGCGATACCCCTGTGCCTGTTCAGCCGGAATATCGTGT
TGCTCCAGATCAGCAGCATAGCTATAGCGTGACCCTGCGTCCGGTGCCGAAGCTTTAACTCGAG

S. solfataricus

GGATCCACATATGTATAGCTTTCCGAACAGCTTTCGTTTTGGTTGGAGCCAGGCAGGTTTTCAGAGCGAAATGGGTACA
CCGGGTAGCGAAGATCCGAATACCGATTGGTATAAATGGGTTCATGATCCGGAAAACATGGCAGCAGGTCTGGTTAGCG
GTGATCTGCCGGAAAATGGTCCGGGTTATTGGGGTAACTATAAAACCTTTCATGACAACGCACAGAAAATGGGTCTGA
AAATTGCACGTCTGAATGTTGAATGGTCACGCATTTTTCCGAATCCGCTGCCTCGTCCGCAGAATTTTGATGAAAGCAA
ACAGGATGTTACCGAAGTGGAAATCAATGAGAATGAACTGAAACGCCTGGATGAGTATGCAAATAAAGATGCCCTGAA
TCACTATCGCGAGATCTTCAAAGATCTGAAAAGCCGTGGTCTGTATTTCATCCTGAATATGTATCATTGGCCTCTGCCG
CTGTGGCTGCATGATCCGATTCGTGTTCGTCGTGGTGATTTTACCGGTCCGAGCGGTTGGCTGAGCACCCGTACCGTTT
ATGAATTTGCACGTTTTAGCGCATATATCGCCTGGAAATTTGATGATCTGGTGGATGAATATAGCACCATGAATGAACC
GAATGTTGTTGGTGGTCTGGGTTATGTTGGTGTTAAAAGCGGTTTTCCGCCTGGTTATCTGTCATTTGAACTGAGCCGT
CGTGCAATGTATAACATTATTCAGGCACATGCACGTGCGTATGATGGCATTAAAAGCGTTAGCAAAAAACCGGTGGGT
ATCATTTATGCCAATAGCAGCTTTCAGCCGCTGACCGATAAAGATATGGAAGCAGTTGAAATGGCCGAAAATGATAAT
CGTTGGTGGTTTTTTGATGCCATTATTCGCGGTGAAATTACCCGTGGCAATGAAAAAATTGTTCGCGACGATCTGAAAG
GTCGTCTGGATTGGATTGGTGTGAACTATTATACCCGCACCGTTGTTAAACGTACCGAAAAAGGTTATGTTAGCCTGGG
TGGTTATGGTCATGGTTGTGAACGTAATAGCGTGAGCCTGGCAGGTCTGCCGACCAGCGATTTTGGCTGGGAATTTTTT
CCGGAAGGTCTGTATGATGTGCTGACCAAATATTGGAATCGCTATCACCTGTATATGTACGTGACCGAAAATGGTATTG
CAGATGATGCAGATTATCAGCGTCCGTATTATCTGGTGAGCCATGTTTATCAGGTTCATCGTGCAATTAATAGCGGTGC
AGATGTTCGCGGATATCTGCATTGGAGCCTGGCCGATAATTATGAATGGGCCAGCGGTTTTAGCATGCGTTTTGGTCTG
CTGAAAGTTGACTATAATACCAAACGTCTGTATTGGCGTCCGAGCGCACTGGTTTATCGTGAAATTGCAACCAATGGTG
CCATTACCGATGAAATTGAACATCTGAATAGCGTTCCGCCTGTTAAACCGCTGCGTCATAAGCTTTAACTCGAG

T. neapolitana

GGATCCACATATGTATAGCTTTCCGAACAGCTTTCGTTTTGGTTGGAGCCAGGCAGGTTTTCAGAGCGAAATGGGTACA
CCGGGTAGCGAAGATCCGAATACCGATTGGTATAAATGGGTTCATGATCCGGAAAACATGGCAGCAGGTCTGGTTAGCG
GTGATCTGCCGGAAAATGGTCCGGGTTATTGGGGTAACTATAAAACCTTTCATGACAACGCACAGAAAATGGGTCTGA
AAATTGCACGTCTGAATGTTGAATGGTCACGCATTTTTCCGAATCCGCTGCCTCGTCCGCAGAATTTTGATGAAAGCAA
ACAGGATGTTACCGAAGTGGAAATCAATGAGAATGAACTGAAACGCCTGGATGAGTATGCAAATAAAGATGCCCTGAA
TCACTATCGCGAGATCTTCAAAGATCTGAAAAGCCGTGGTCTGTATTTCATCCTGAATATGTATCATTGGCCTCTGCCG
CTGTGGCTGCATGATCCGATTCGTGTTCGTCGTGGTGATTTTACCGGTCCGAGCGGTTGGCTGAGCACCCGTACCGTTT
ATGAATTTGCACGTTTTAGCGCATATATCGCCTGGAAATTTGATGATCTGGTGGATGAATATAGCACCATGAATGAACC
GAATGTTGTTGGTGGTCTGGGTTATGTTGGTGTTAAAAGCGGTTTTCCGCCTGGTTATCTGTCATTTGAACTGAGCCGT
CGTGCAATGTATAACATTATTCAGGCACATGCACGTGCGTATGATGGCATTAAAAGCGTTAGCAAAAAACCGGTGGGT
ATCATTTATGCCAATAGCAGCTTTCAGCCGCTGACCGATAAAGATATGGAAGCAGTTGAAATGGCCGAAAATGATAAT
CGTTGGTGGTTTTTTGATGCCATTATTCGCGGTGAAATTACCCGTGGCAATGAAAAAATTGTTCGCGACGATCTGAAAG
GTCGTCTGGATTGGATTGGTGTGAACTATTATACCCGCACCGTTGTTAAACGTACCGAAAAAGGTTATGTTAGCCTGGG
TGGTTATGGTCATGGTTGTGAACGTAATAGCGTGAGCCTGGCAGGTCTGCCGACCAGCGATTTTGGCTGGGAATTTTTT
CCGGAAGGTCTGTATGATGTGCTGACCAAATATTGGAATCGCTATCACCTGTATATGTACGTGACCGAAAATGGTATTG
CAGATGATGCAGATTATCAGCGTCCGTATTATCTGGTGAGCCATGTTTATCAGGTTCATCGTGCAATTAATAGCGGTGC
AGATGTTCGCGGATATCTGCATTGGAGCCTGGCCGATAATTATGAATGGGCCAGCGGTTTTAGCATGCGTTTTGGTCTG
CTGAAAGTTGACTATAATACCAAACGTCTGTATTGGCGTCCGAGCGCACTGGTTTATCGTGAAATTGCAACCAATGGTG
CCATTACCGATGAAATTGAACATCTGAATAGCGTTCCGCCTGTTAAACCGCTGCGTCATAAGCTTTAACTCGAG

A.2.2 Amino acid sequences

B. circulans

MKKAISCVFLISALILSSFQVPVQGQAMSKTTSAAGNSVSYDGERRVNFNENWRFQRETNGSIAGAQNPGFDDSSWRKLN
LPHDWSIELDFNKNSLATHEGGYLDGGIGWYRKTFTIPESMKGKRISLDFDGVYMNSTTYLNGEVLGTYPFGYNAFSYDI
SDKLYKDGRANVLVVKVNNTQPSSRWYSGSGIYRNVYLTVTDPIHVARYGTFVTTPNLEKSIKEDRADVNIKTKISNDAA
EAKQVKIKSTIYDGAGNTVQTVETEEKTAAAGTVTPFEQNTVIKQPKLWSIDKPYRYNLVTEVIVGGQTVDTYETKFGVR
YFKFDENEGFSLNGEYMKLHGVSMHHDLGALGAATNARGVERQMQIMKDMGVNAIRVTHNPASPELLEAANKLGLFII
EEAFDSWAQSKKPYDYGRFFNAWAEHDIKEMVDRGKNEPAIIMWSIGNEIYDTTNAAGVETARNLVGWVKEIDTTRPT
TIGEDKTRGDKVNVTPINSYIKEIFNIVDVVGLNYSENNYDGYHKQNPSWKLYGSETSSATRSRGVYTHPYQYNQSTKYA
DLQQSSYDNDYVGWGRTAEDAWKYDRDLKHIAGQFIWTGFDYIGEPTPYYNSYPAKSSYFGAVDTAGFPKDIFYYYQSQ
WKKEPMVHLLPHWNWKEGEKVRVLAYTNASKVELVLNGESLGEKNYDNKQTSWGAPYKETKDGKTYLEWAVPFKPG
KLEAVAKDENGKVIARDQVVTAGEPASVRLTADRKVVKADGTDLSFITADIVDSKGIVVPDADHLITFNVTGQGELAGVD
NGNASSVERYKDNKRKAFSGKALAIVQSSKLSGKITVHASVAGLSSDSTSVFTVTPADHDKKIVAGIDDVNLTVDVNEAPK
LPSEIKVYYSDESAAAKNVTWDEVDPKQYSTVGEFTVEGSVEGTSLKAKAFVIVKGIVAVKPYSTATKVGVQPVLPEKATL

LYSDGTTKGATVTWDEIPEDKLAKEGRFTVEGSVEGTDLKANVYVRVTNEVKSVNIMLQEQGSAYPKLEATFTNPADNL
QHLNDGIKSYTNNPVNRWTNWTRTPRDAGDSITVNFGKKHVINNLDLFVFTDSGTVVPEKAEVQYWDGTAWKDVENL
TQPSPYVVEKNELTFDAVATEKLKFHLTPSVKGKFLALTEAEVYADQIVMGETAKLQSITVNGKALEGFDHAKKNYELVL
PYGSELPKIEAAAADNATVTILPAFSYPGTAKLFVTSEDGKVTTEYSIGVSTEEPKLVSAELSADKTNVMEDDIIDLKVIGLF
ESKEKIDVTDSQPTYEFDQQIIKIEGNKLYALETGNVKVKVTVTYKGVSVTTPALEFTIAKNPAPKYITSLEPVTVVVKKGE
APELPATVVAHYNRGIPRDVKVKWERINPSKYQQLGEFTVSGMVEGTDIKAQAKVIVKGAVAVEDIRMAVLLKQMPQLP
GKVTVYYSDGAEEQRAVKWEEIPQEELENVGEFKVKGDVNGVKLKATATIRVTDEVGGEQNISRAKNGYEYPKAEASFT
NNGPGSSDRIEAINDDVISYEANPHNRWTNWQPVPRAGDWVSITFGDYEPTEYDVDSMEIHWFADHGTSYPERFQIEYK
SGDSWKEVTSLKSDPASPALGKANVYSFDRVKTSAIRVKMTAQAGKSLAITELKVFSKWPKAGTEPEVTDIKVGGKSILE
DFEQKGDHYEVTIDAGDANVMPKINVKAKDQTSITIVPAVTSPSTAKVIAKSEDGKKVKVYSIHYK

B. longum

MRARRDFAWPKLLTADGRGIAFGGDYNPDQWPEDVWDDDIRLMGQAGVNTVALAIFSWDRLQPTEDHWNFDWLDR
IIDKLGQAGIAVDLASATATAPLWLYENHPEVLPRDKYGHPVNAGSRQSWSPTSPVFKEYALTLCRKLAERYGTNPYVTA
WHMGNEYGWNNRDDYSDNALEAFRAWCRRKYGTIDALNQAWGTTFWGQEMTGFDEVLIPRFMGADSMVNPGQKLD
FERFGNDMLLDFYKAERDAIAEICPDKPFTTNFMISTDQCCMDYAAWAEEVNFVSNDHYFHEGESHLDELACSDALVDS
LALGKPWYVMEHSTSAVQWKPLNTRKRNGETVRDSLAHVAMGADAINFFQWRASAFGAEAFHSAMVPHAGENTKLF
RQVCELGATLQALADAGVQGSELAHADTAILFSAESEWATRSETLPSMKLNHWHDVRDWYRAFLNAGARADIVPLAYD
WSSYKTIVLPTVLMLSDADTRRLAGFVQDGGRVVVGYATGLLDERFHTWLGGYPGAGDGLLRSMLGVRGEEFNILGTET
EDEPSEIRLASTGDSPTMDGAVTRLWQNDVTVAGPHVQVLAAYAGEEANEWELDGTAAITRNTYGEGEAYFLGCDLGV
SDLTRFVGGWLAARPQDGRQPEANLRSPASGVTSDVLHTVRQSDDAIFDFYLTRGKSDVELRDIAGEPIVLFRAERGSDG
GAYTVHRNGVLVMKRPNPSV

G. kaustophilus

DKVIDILYDHGVYINLGTATATTPAWFVKKYPDSLPIDESGVILSFGSRQHYCPNHPQLITHIKRLVRAIAERYKNHPALK
MWHVNNEYACHVSKCFCENCAVAFRKWLKERYKTIDELNERWGTNFWGQRYNHWDEINPPRKAPTFINPSQELDYYR
FMNDSILKLFLTEKEILREVTPDIPVSTNFMGSFKPLNYFQWAQHVDIVTWDSYPDPREGLPIQHAMMNDLMRSLRKGQ
PFILMEQVTSHVNWRDINVPKPPGVMRLWSYATIARGADGIMFFQWRQSRAGAEKFHGAMVPHFLNENNRIYREVTQL
GQELKKLDCLVGSRIKAEVAIIFDWENWWAVELSSKPHNKLRYIPIVEAYYRELYKRNIAVDFVRPSDDLTKYKVVIAPM
LYMVKEGEDENLRQFVANGGTLIVSFFSGIVDENDRVHLGGYPGPLRDILGIFVEEFVPYPETKVNKIYSNDGEYDCTTWA
DIIRLEGAEPLATFKGDWYAGLPAVTRNCYGKGEGIYVGTYPDSNYLGRLLEQVFAKHHINPILEVAENVEVQQRETDEW
KYLIIINHNDYEVTLSLPEDKIYQNMIDGKCFRGGELRIQGVDVAVLREHDEAGKV

L. delbrueckii

MSNKLVKEKRVDQADLAWLTDPEVYEVNTIPPHSDHESFQSQEELEEGKSSLVQSLDGDWLIDYAENGQGPVNFYAEDF
DDSNFKSVKVPGNLELQGFGQPQYVNVQYPWDGSEEIFPPQIPSKNPLASYVRYFDLDEAFWDKEVSLKFDGAATAIYV
WLNGHFVGYGEDSFTPSEFMVTKFLKKENNRLAVALYKYSSASWLEDQDFWRMSGLFRSVTLQAKPRLHLEDLKLTAS
LTDNYQKGKLEVEANIAYRLPNASFKLEVRDSEGDLVAEKLGPIRSEQLEFTLADLPVAAWSAEKPNLYQVRLYLYQAGS
LLEVSRQEVGFRNFELKDGIMYLNGQRIVFKGANRHEFDSKLGRAITEEDMIWDIKTMKRSNINAVRCSHYPNQSLFYRL

CDKYGLYVIDEANLESHGTWEKVGGHEDPSFNVPGDDQHWLGASLSRVKNMMARDKNHASILIWSLGNESYAGTVFAQ
MADYVRKADPTRVQHYEGVTHNRKFDDATQIESRMYAPAKVIEEYLTNKPAKPFISVEYAHAMGNSVGDLAAYTALEK
YPHYQGGFIWDWIDQGLEKDGHLLYGGDFDDRPTDYEFCGNGLVFADRTESPKLANVKALYANLKLEVKDGQLFLKND
NLFTNSSSYYFLTSLLVDGKLTYQSRPLTFGLEPGESGTFALPWPEVADEKGEVVYRVTAHLKEDLPWADEGFTVAEAEE
VAQKLPEFKPEGRPDLVDSDYNLGLKGNNFQILFSKVKGWPVSLKYAGREYLKRLPEFTFWRALTDNDRGAGYGYDLAR
WENAGKYARLKDISCEVKEDSVLVKTAFTLPVALKGDLTVTYEVDGRGKIAVTADFPGAEEAGLLPAFGLNLALPKELTD
YRYYGLGPNESYPDRLEGNYLGIYQGAVKKNFSPYLRPQETGNRSKVRWYQLFDEKGGLEFTANGADLNLSALPYSAAQI
EAADHAFELTNNYTWVRALSAQMGVGGDDSWGQKVHPEFCLDAQKARQLRLVIQPLLLK

P. thiaminolycticus

MTTLSYDEGQFKMGDRPIQLISGAIHYFRIVPAYWEDRLRKIKAMGCNCIETYVAWNVHEPREGEFHFERMADVAEFVR
LAGELGLYVIVRPSPYICAEWEFGGLPAWLLKDDMRLRCNDPRFLEKVSAYYDALLPQLTPLLATKGGPIIAVQIENEYGS
YGNDQAYLQAQRAMLIERGVDVLLFTSDGPQDDMLQGGMAEGVLATVNFGSRPKEAFDKLKEYQPDGPLMCMEYWNG
WFDHWFEPHHTRDAKDAARVLDDMLGMGASVNFYMVHGGTNFGFGSGANHSDKYEPTVTSYDYDAAISEAGDLTPKY
HAFREVIGKYVSLPEGELPANTPKADYGSVPVTRRVKLFDTLEPMTEVKDSICPEPMEKYGQNNGFIVYSTRISGPRPESR
LTIQDVRDRALVFLDRKLIGVVERWNPQSLPVTVPEDGAQLDILVENMGRVNYGPQLYDRKGITHGVRLNGQFLFHWEV
RSLELETLAGLSFDTAGARAWEEEQPGFYEAKLVIEDEPKDTFLRLEGWKKGVVFVNGFNLGRYWEVGPQQALYVPAPV
LRQGENEIVVFELHREGEPLRFEAAPSLSK

S. solftaricus

MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPGYWGNYKTFHDNAQKMGL
KIARLNVEWSRIFPNPLPRPQNFDESKQDVTEVEINENELKRLDEYANKDALNHYREIFKDLKSRGLYFILNMYHWPLPL
WLHDPIRVRRGDFTGPSGWLSTRTVYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSRR
AMYNIIQAHARAYDGIKSVSKKPVGIIYANSSFQPLTDKDMEAVEMAENDNRWWFFDAIIRGEITRGNEKIVRDDLKGRL
DWIGVNYYTRTVVKRTEKGYVSLGGYGHGCERNSVSLAGLPTSDFGWEFFPEGLYDVLTKYWNRYHLYMYVTENGIAD
DADYQRPYYLVSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFGLLKVDYNTKRLYWRPSALVYREIATNGA
ITDEIEHLNSVPPVKPLRH

S. rectivirgula

MRRRMHPAVLLATLAVTAGLLLPAAHAQPVDGPSDPAEIESYLEDPQRTGEGQQPPHAFLRPYADAEQALRSAREDLFA
PADEPTPWTLSLNGRWQFQYADHYRDLPAGWQRGSGEWDAVDVPGVWQQQGYDRPIYRNVSSEIAPYRPPEVPDDIN
PTGAYRKQFELPEHWAGRRQLLRFEGVTSGYFVWVNGHYVGYDQGGYTPAEFDVTPHLRPGTNTIAVQVHRWSSGSYL
ENMDFWHFSGIFREVHLYSVPRTHLEDVTVRTELDDTYTDATLRLGAELRHLPGGTTGRHELRATLYDPDGNRVAEVSE
PVVVDRETASAELAQHVAAPRLWSDETPHLYTAVLELLHNGQVVHTTQQPVGFRSVEVRDEQLLLNGKPVELRGVNRH
EHDPRTGRAVSRQRQQQDVELLRQHNINAVRTSHYPNDPYWYRLADHHGILLADEVDVETHYREDCSNPADDCLADRP
EWQAAFADRFHALLERDKNHPSVIIWDTGNEAGLGAAHYAMAEHARATDPTRPLYHQSNSPDGDAPYADIWGPRYPS
PQRLEEIADETTKPVVMGEWLHAMGNSLGHYEDMWRTIRREPALQGGFVWDWVDQGLYRPLRTTPANIPVHYGGNP
RPVPGVSGNGLLLSGLDDWVEAYRDPALDITGRQVTLDMWVKPLSWQGSGTFLSKGDKQWALQMPDPEHVEFFVYGE
DDWHTARVRIPDDWWGNWHRLSGVYDGEQVRLHLDGEQVAATPYRGEIAANTMYTASIGRNQEKHGDSFAGRTAHA
VVDSVRVYDTALTRQQLEDDPAERAVLALDFETEQQNGEFLDYGSSNFVVNGLVNADRTPQPELAQLAYSHAPVRFEGT
GEPGEFRVHNRNHTLTTEAYDLTWRLVEGGRDIASGALDVVVGPGQQQLVELDLPPAGGTQRYLVLEARQRADAPGVP
RGHLVATEQLPAGGTQPPPPPESGGEPQPLAVHEHGNAVEVSAPDFTYTFDREAGTLTGMTANGQQHLVRGPQLDVFR
APIGNEWSDWSGLAPEAQFRAVGLDRLRTEVTDFTVSQPAPDQVRVSVRTDVSAPDVPDDGFRSTWVYTVDGTGAITID
HQVDAYGERMRALPWLPRVGMSMAVPDSFGQLDWYGRGPGESYPDRKDAQHIGRWSQTVDESWFDFLPPQDNGVKT
DTWWAVLSGDRGGLQVSGDRLAIAADRFSNAERTDFAHQLRRDDFVTLHVSAAVTGLGDTPVPVQPEYRVAPDQQHSY
SVTLRPVP

T. neapolitana

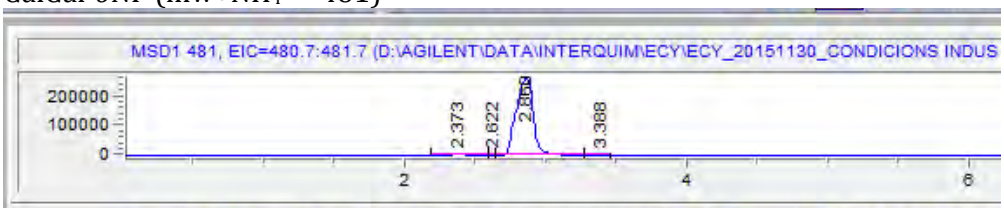
MLGVCYYPEHWGVEKVEEDFRRMKELGIEYVRIGEFAWSRIEPERGKFNWEWLDRTLEAAERMGLKIILGTPTATPPK
WLVDEHSEILPVDREGRVKNFGSRRHYCFSSPVYREEAKRIVSIVAGRYGKHPAVVGWQTDNEYGCHDTVRCYCPRCKK
AFQKWLERRYEGDIDKLNRAWGTVFWSQEYRSFEEIELPNLTPAGPNPSHLLDYYRFASDQVVEFNRLQVEIIRELSPGR
FVTHNFMAGFTDFDHYKISKDLDFASWDNYPLGHTLVFLRAKGESKNPFNRVGHPDIISFSHDLYRGVGRGRFWVMEQ
QAGPVNWAPYNLWPAEGAVRLWTWQAFAHGAEVVSYFRWRQAPFAQEQMHSGLLAPDSSPSQGYQEVKQVFEELKG
VDLSEPVKSEVALVFDYETAWVFSIQPHGEGVNYLDLVLRFYSALRGLGLNVDIVSADSPLDGYRMVVVPSLAIVREEALE
NFKKYSGVLVIGPRSGSKTETFQIPEEMPPGLLKEIIPVEVKRVESLGTGNVEVLLWDGKKYPVTIWREDVDPTIAEVVAR

FEDGPGAIFRKENVYYLAFWPDREFLLDFFEKLMKEIGYTVRRLPEGVRIQKRGKYTFAFNFSPHETEVEIPDDARIVLGD
RKVEPYSLVVWEEV

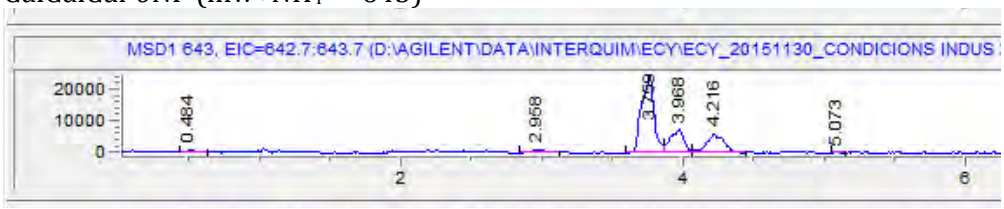
A. 3 Scan chromatograms

E. coli chromatograms in mode scan extracting corresponding ions

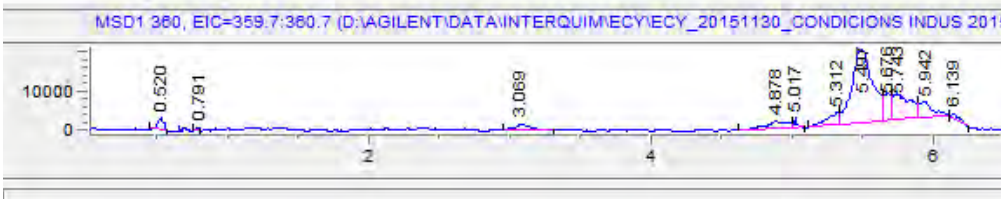
GalGal-*o*NP (mw+NH_4^+ = 481)



GalGalGal-*o*NP (mw+NH_4^+ = 643)

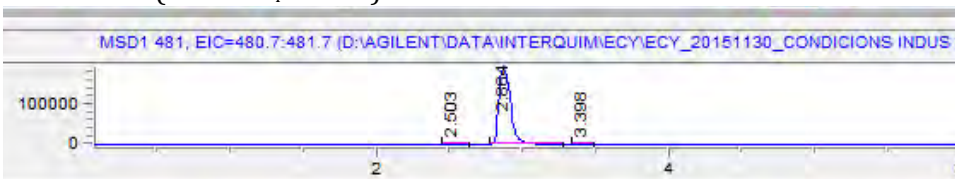


GalGal (mw+NH_4^+ = 360)

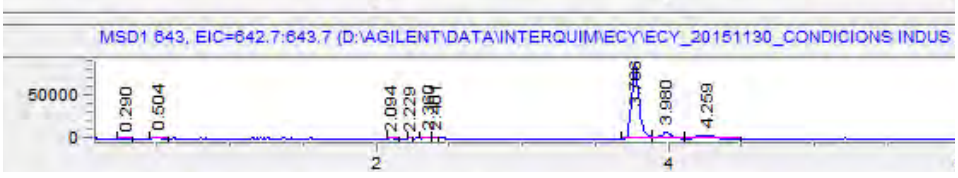


[redacted] chromatograms in mode scan extracting corresponding ions

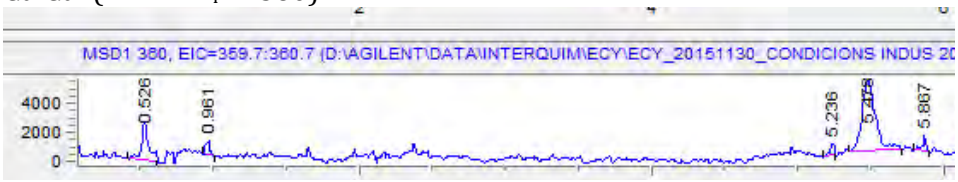
GalGal-*o*NP (mw+NH_4^+ = 481)



GalGalGal-*o*NP (mw+NH_4^+ = 643)

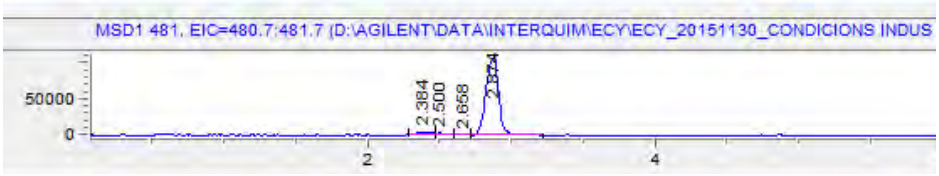


GalGal (mw+NH_4^+ = 360)

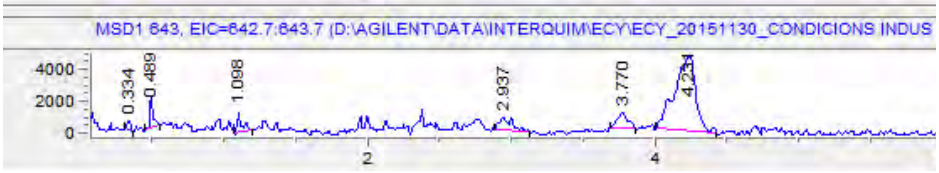


L. delbrueckii chromatograms in mode scan extracting corresponding ions

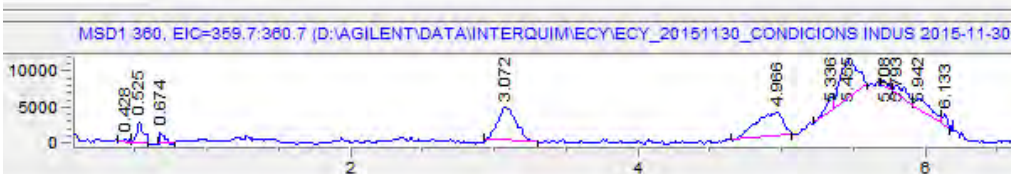
GalGal-*o*NP (mw+NH_4^+ = 481)



GalGalGal-*o*NP (mw+NH_4^+ = 643)

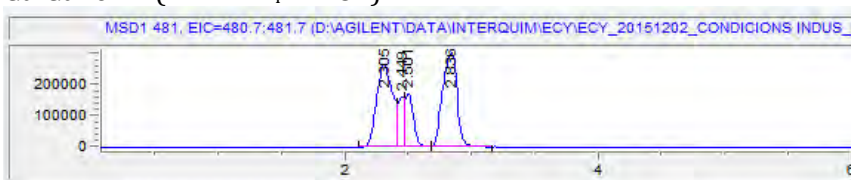


GalGal (mw+NH_4^+ = 360)

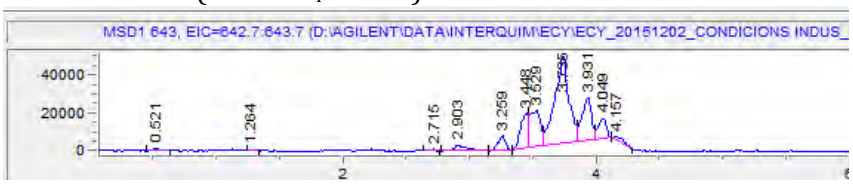


S. rectivirgula chromatograms in mode scan extracting corresponding ions

GalGal-*o*NP (mw+NH_4^+ = 481)

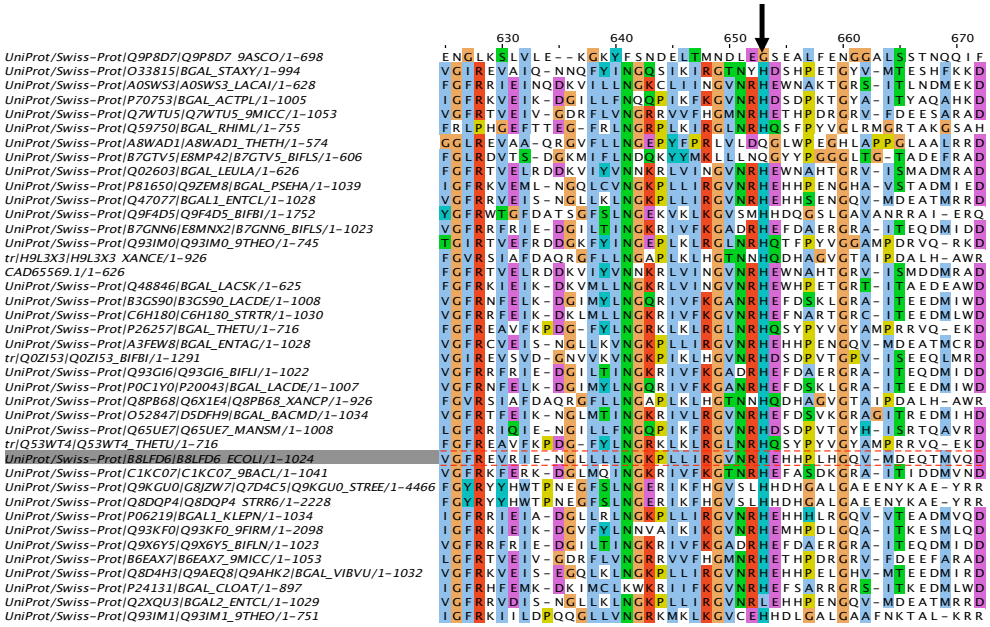


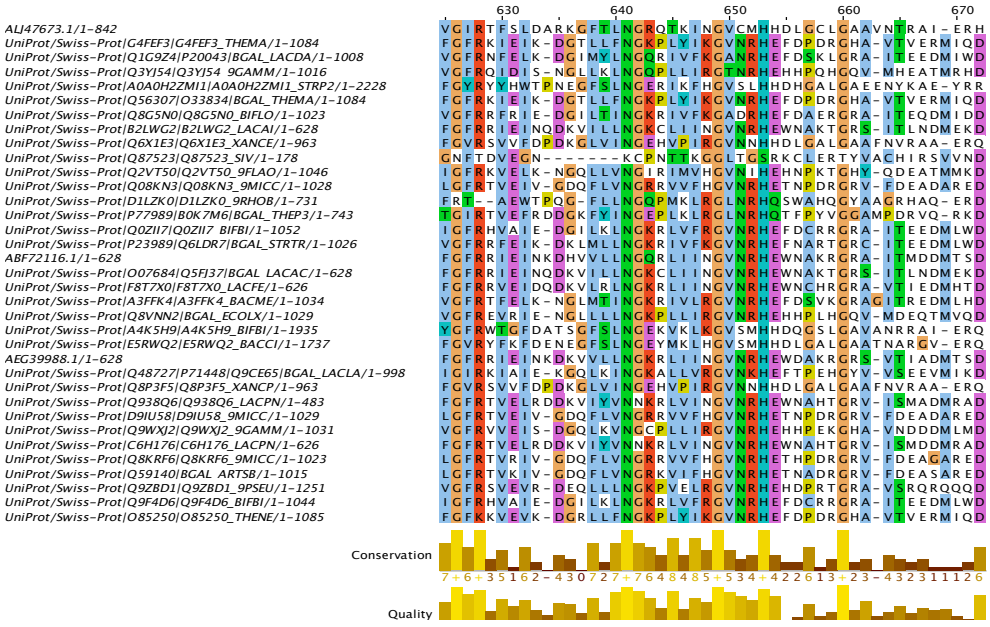
GalGalGal-*o*NP (mw+NH_4^+ = 643)



A.4 Protein alignment

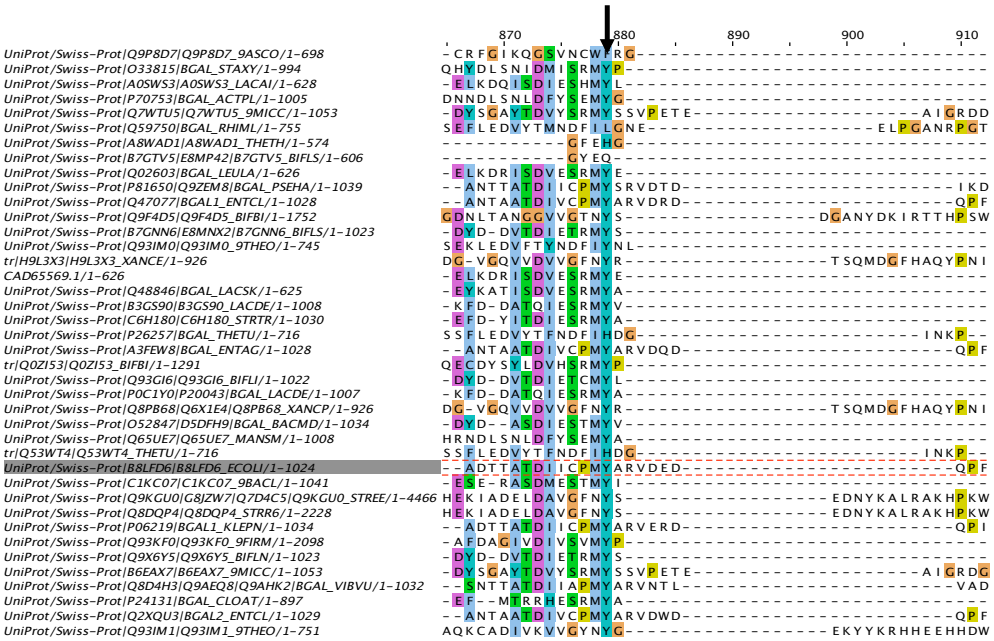
A.4.1 L359

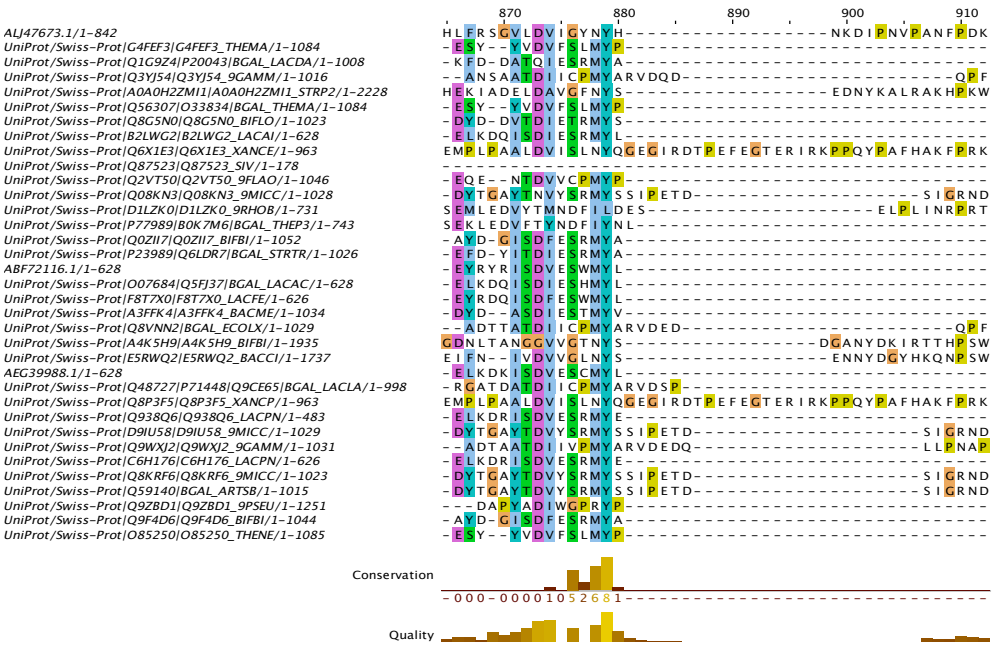




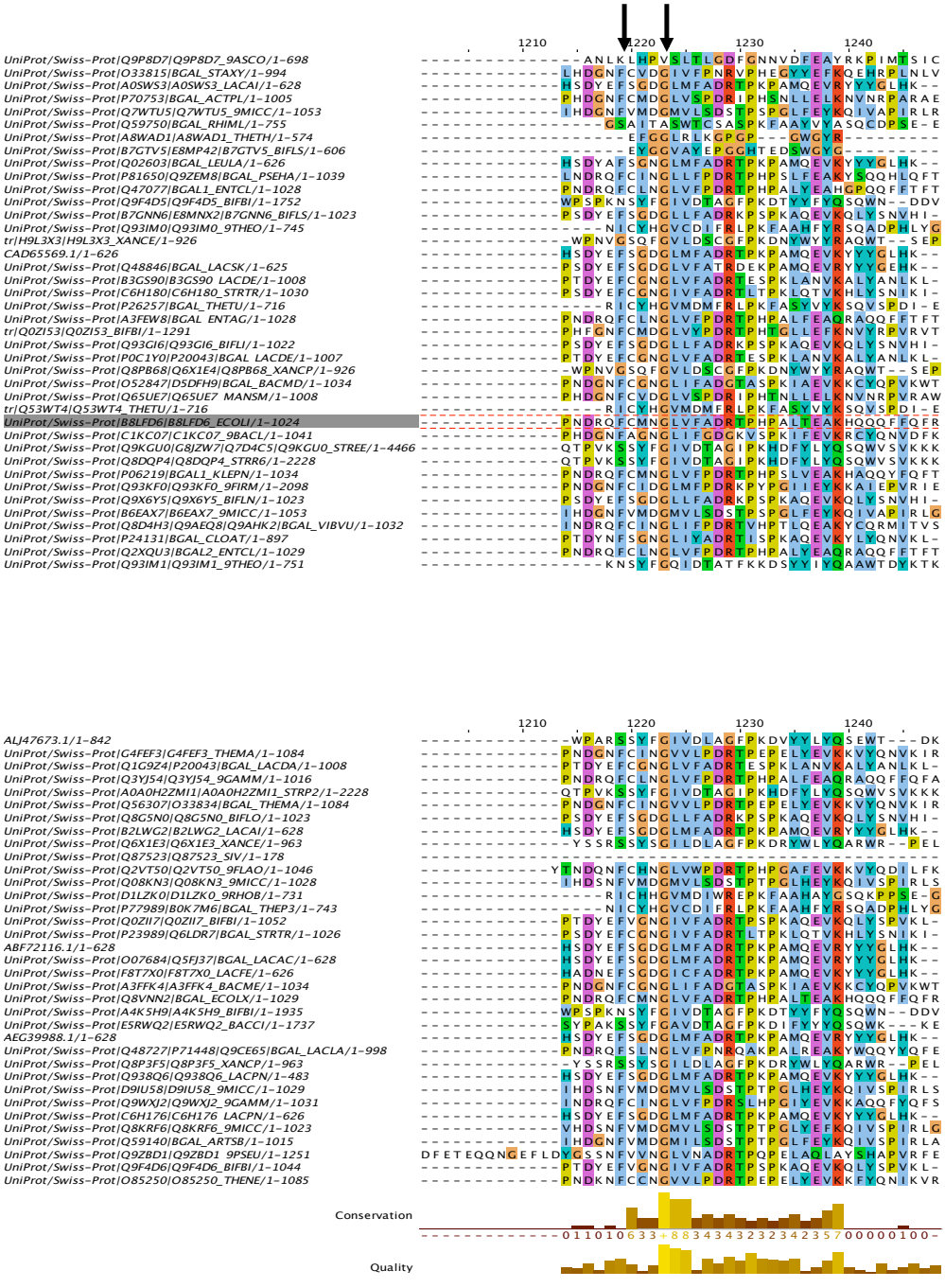
A.4.2 R390 and H420

A.4.3 Y505

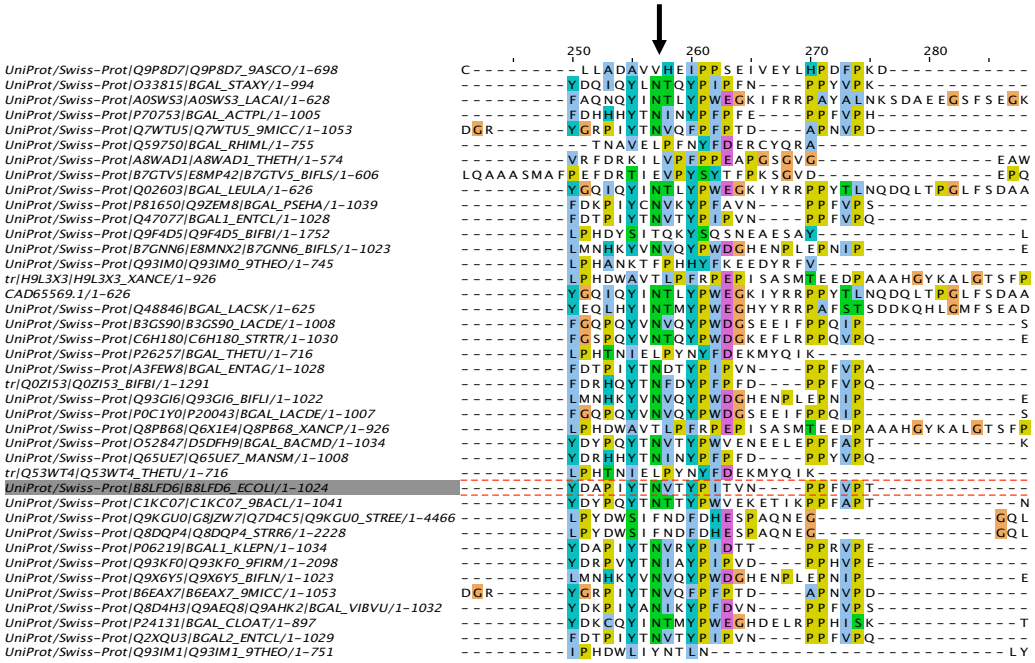

870
880
890
900
910
UniProt/Swiss-Prot|Q9P8D7|Q9P8D7_9ASCO/1-698
UniProt/Swiss-Prot|O33815|BGAL_STAXY/1-994
UniProt/Swiss-Prot|A0SWS3|A0SWS3_LACAI/1-628
UniProt/Swiss-Prot|P70753|BGAL_ACTPL/1-1005
UniProt/Swiss-Prot|Q7WTU5|Q7WTU5_9MICC/1-1053
UniProt/Swiss-Prot|Q59750|BGAL_RHIML/1-755
UniProt/Swiss-Prot|A8WAD1|A8WAD1_THETH/1-574
UniProt/Swiss-Prot|B7GTV5|E8MP42|B7GTV5_BIFLS/1-606
UniProt/Swiss-Prot|Q02603|BGAL_LEULA/1-626
UniProt/Swiss-Prot|P81650|Q9ZEM8|BGAL_PSEHA/1-1039
UniProt/Swiss-Prot|Q47077|BGAL1_ENTCL/1-1028
UniProt/Swiss-Prot|Q9F4D5|Q9F4D5_BIFBI/1-1752
UniProt/Swiss-Prot|B7GNN6|E8MNX2|B7GNN6_BIFLS/1-1023
UniProt/Swiss-Prot|Q93IM0|Q93IM0_9THEO/1-745
tr|H9L3X3|H9L3X3_XANCE/1-926
CAD65569.1/1-626
UniProt/Swiss-Prot|Q48846|BGAL_LACSK/1-625
UniProt/Swiss-Prot|B3GS90|B3GS90_LACDE/1-1008
UniProt/Swiss-Prot|C6H180|C6H180_STRTR/1-1030
UniProt/Swiss-Prot|P26257|BGAL_THETU/1-716
UniProt/Swiss-Prot|A3FEW8|BGAL_ENTAG/1-1028
tr|Q0ZI53|Q0ZI53_BIFBI/1-1291
UniProt/Swiss-Prot|Q93GI6|Q93GI6_BIFLI/1-1022
UniProt/Swiss-Prot|P0C1Y0|P20043|BGAL_LACDE/1-1007
UniProt/Swiss-Prot|Q8PB68|Q6X1E4|Q8PB68_XANCP/1-926
UniProt/Swiss-Prot|O52847|D5DFH9|BGAL_BACMD/1-1034
UniProt/Swiss-Prot|Q65UE7|Q65UE7_MANSM/1-1008
tr|Q53WT4|Q53WT4_THETU/1-716
UniProt/Swiss-Prot|B8LFD6|B8LFD6_ECOLI/1-1024
UniProt/Swiss-Prot|C1KC07|C1KC07_9BACL/1-1041
UniProt/Swiss-Prot|Q9KGU0|G8JZW7|Q7D4C5|Q9KGU0_STREE/1-4466
UniProt/Swiss-Prot|Q8DQP4|Q8DQP4_STRR6/1-2228
UniProt/Swiss-Prot|P06219|BGAL1_KLEPN/1-1034
UniProt/Swiss-Prot|Q93KF0|Q93KF0_9FIRM/1-2098
UniProt/Swiss-Prot|Q9X6Y5|Q9X6Y5_BIFLN/1-1023
UniProt/Swiss-Prot|B6EAX7|B6EAX7_9MICC/1-1053
UniProt/Swiss-Prot|Q8D4H3|Q9AEQ8|Q9AHK2|BGAL_VIBVU/1-1032
UniProt/Swiss-Prot|P24131|BGAL_CLOAT/1-897
UniProt/Swiss-Prot|Q2XQU3|BGAL2_ENTCL/1-1029
UniProt/Swiss-Prot|Q93IM1|Q93IM1_9THEO/1-751

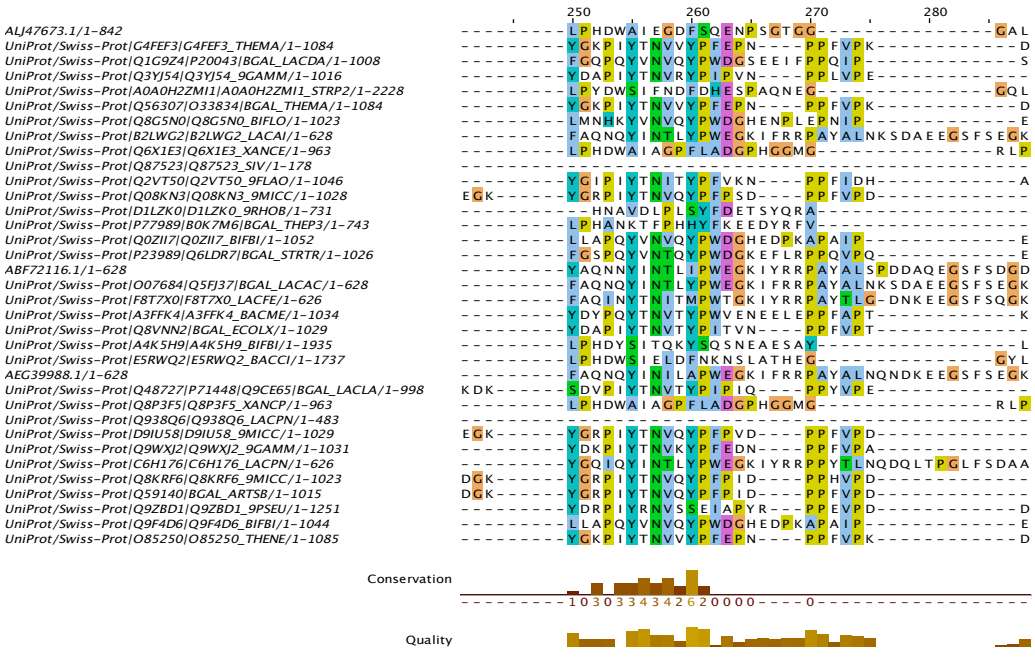

870
880
890
900
910
ALJ47673.1/1-842
UniProt/Swiss-Prot|G4FEF3|G4FEF3_THEMA/1-1084
UniProt/Swiss-Prot|Q1G9Z4|P20043|BGAL_LACDA/1-1008
UniProt/Swiss-Prot|Q3YJ54|Q3YJ54_9GAMM/1-1016
UniProt/Swiss-Prot|A0A0H2ZMI1|A0A0H2ZMI1_STRP2/1-2228
UniProt/Swiss-Prot|Q56307|O33834|BGAL_THEMA/1-1084
UniProt/Swiss-Prot|Q8G5N0|Q8G5N0_BIFLO/1-1023
UniProt/Swiss-Prot|B2LWG2|B2LWG2_LACAI/1-628
UniProt/Swiss-Prot|Q6X1E3|Q6X1E3_XANCE/1-963
UniProt/Swiss-Prot|Q87523|Q87523_SIV/1-178
UniProt/Swiss-Prot|Q2VT50|Q2VT50_9FLAO/1-1046
UniProt/Swiss-Prot|Q08KN3|Q08KN3_9MICC/1-1028
UniProt/Swiss-Prot|D1LZK0|D1LZK0_9RHOB/1-731
UniProt/Swiss-Prot|P77989|B0K7M6|BGAL_THEP3/1-743
UniProt/Swiss-Prot|Q0ZII7|Q0ZII7_BIFBI/1-1052
UniProt/Swiss-Prot|P23989|Q6LDR7|BGAL_STRTR/1-1026
ABF72116.1/1-628
UniProt/Swiss-Prot|O07684|Q5FJ37|BGAL_LACAC/1-628
UniProt/Swiss-Prot|F8T7X0|F8T7X0_LACFE/1-626
UniProt/Swiss-Prot|A3FFK4|A3FFK4_BACME/1-1034
UniProt/Swiss-Prot|Q8VNN2|BGAL_ECOLX/1-1029
UniProt/Swiss-Prot|A4K5H9|A4K5H9_BIFBI/1-1935
UniProt/Swiss-Prot|E5RWQ2|E5RWQ2_BACCI/1-1737
AEG39988.1/1-628
UniProt/Swiss-Prot|Q48727|P71448|Q9CE65|BGAL_LACLA/1-998
UniProt/Swiss-Prot|Q8P3F5|Q8P3F5_XANCP/1-963
UniProt/Swiss-Prot|Q938Q6|Q938Q6_LACPN/1-483
UniProt/Swiss-Prot|D9IU58|D9IU58_9MICC/1-1029
UniProt/Swiss-Prot|Q9WXJ2|Q9WXJ2_9GAMM/1-1031
UniProt/Swiss-Prot|C6H176|C6H176_LACPN/1-626
UniProt/Swiss-Prot|Q8KRF6|Q8KRF6_9MICC/1-1023
UniProt/Swiss-Prot|Q59140|BGAL_ARTSB/1-1015
UniProt/Swiss-Prot|Q9ZBD1|Q9ZBD1_9PSEU/1-1251
UniProt/Swiss-Prot|Q9F4D6|Q9F4D6_BIFBI/1-1044
UniProt/Swiss-Prot|O85250|O85250_THENE/1-1085
Conservation
Quality

A.4.4 N606 and F603



A.4.5 N104 and V105





BIBLIOGRAPHY

1. Silanikove, N., Leitner, G. & Merin, U. The interrelationships between lactose intolerance and the modern dairy industry: Global perspectives in evolutional and historical backgrounds. *NutrientsNutrients* **7,** 7312–7331 (2015).

2. Pilson, M. E. Q. & Kelly, A. L. Composition of the Milk from Zalophus californianus, the California Sea Lion. *Science (80-.).* **135,** 104–105 (1962).

3. Vesa, T. H., Marteau, P. & Korpela, R. Lactose Intolerance. *J. Am. Coll. Nutr.* **19,** 165S–175S (2000).

4. Fox, P. F. & Mcsweeney, P. L. H. Dairy Chemistry and Biochemistry. *Dairy Chem. Biochem.* **1542,** 478 (1998).

5. Wong, S. Y. & Hartel, R. W. Crystallization in Lactose Refining-A Review. *J. Food Sci.* **79,** 257–272 (2014).

6. McSweeney, P. L. H. . F. P. F. *Advanced Dairy Chemistry*. (Springer, 2009).

7. Cochet, B. *et al.* Effects of lactose on intestinal calcium absorption in normal and lactase-deficient subjects. *Gastroenterology* **84,** 935–940 (1983).

8. Thureen, P. J. & Hay, W. W. *Neonatal Nutrition and Metabolism*. 344–345 (Cambridge University Press, 2012). at <https://books.google.de/books?id=Hn7A_OJ2JP8C>

9. Zeng, X. M., Martin, G. P., Marriott, C. & Pritchard, J. The influence of crystallization conditions on the morphology of lactose intended for use as a carrier for dry powder aerosols. *J. Pharm. Pharmacol.* **52,** 633–643 (2000).

10. Kuhn, N. J. & White, A. The Topography of Lactose Synthesis. *J. Mammary Gland Biol. Neoplasia* **14,** 213–220 (2009).

11. Petry, K. G. & Reichardt, J. K. V. The fundamental importance of human galactose metabolism: Lessons from genetics and biochemistry. *Trends Genet.* **14,** 98–102 (1998).

12. Prado, E. L. & Dewey, K. G. Nutrition and brain development in early life. *Nutr. Rev.* **72,** 267–284 (2014).

13. Mills, S., Ross, R. P., Hill, C., Fitzgerald, G. F. & Stanton, C. Milk intelligence: Mining milk for bioactive substances associated with human health. *Int. Dairy J.* **21,** 377–401 (2011).

14. Coelho, A. I., Berry, G. T. & Rubio-Gozalbo, M. E. Galactose metabolism and health. *Curr. Opin. Clin. Nutr. Metab. Care* **18,** 422–7 (2015).

15. Rosenthal, M. D. & Glew, R. H. *Medical Biochemistry*. (John Wiley & Sons, Inc., 2009).

16. Roser, M. *et al.* Metabolism of galactose in the brain and liver of rats and its conversion into glutamate and other amino acids. *J. Neural Transm.* **116,** 131–139 (2009).

17. Salkovic-Petrisic, M. *et al.* Long-term oral galactose treatment prevents cognitive deficits in male Wistar rats treated intracerebroventricularly with streptozotocin. *Neuropharmacology* **77,** 68–80 (2014).

18. Naim, H. Y. Molecular and cellular aspects and regulation of intestinal lactase-phlorizin hydrolase. *Histol. Histopathol.* **16,** 553–561 (2001).

19. Simoons, F. J. Age of onset of lactose malabsorption. *Pediatrics* **66,** 646–648 (1980).

20. Wang, Y. *et al.* The genetically programmed down-regulation of lactase in children. *Gastroenterology* **114,** 1230–1236 (1998).

21. Deng, Y., Misselwitz, B., Dai, N. & Fox, M. Lactose intolerance in adults: Biological mechanism and dietary management. *Nutrients* **7,** 8020–8035 (2015).

22. Matthews, S., Waud, J., Roberts, A. & Campbell, A. Systemic lactose intolerance: a new perspective on an old problem. *Postgrad. Med. J.* **81,** 167–173 (2005).

23. Torniainen, S. *et al.* Four novel mutations in the lactase gene (LCT) underlying congenital lactase deficiency (CLD). *BMC Gastroenterol.* **9,** 8 (2009).

24. Kuokkanen, M. *et al.* Mutations in the translated region of the lactase gene (LCT) underlie congenital lactase deficiency. *Am. J. Hum. Genet.* **78,** 339–344 (2006).

25. Robayo-Torres, C. C. & Nichols, B. L. Molecular Differentiation of Congenital Lactase Deficiency from Adult-Type Hypolactasia. *Nutr. Rev.* **65,** 95–98 (2007).

26. Campbell, A. K., Waud, J. P. & Matthews, S. B. *The molecular basis of lactose intolerance*. *Sci. Prog.* **92,** 241–287 (2009).

27. Uchida, N. *et al.* Two novel mutations in the lactase gene in a Japanese infant with congenital lactase deficiency. *Tohoku J. Exp. Med.* **227,** 69–72 (2012).

28. Canani, R. B. *et al.* Diagnosing and treating intolerance to carbohydrates in children. *Nutrients* **8,** (2016).

29. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30,** 233–237 (2002).

30. Swallow, D. M. Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.* **37,** 197–219 (2003).

31. Ingram, C. J. E., Mulcare, C. a., Itan, Y., Thomas, M. G. & Swallow, D. M. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124,** 579–591 (2009).

32. Itan, Y., Jones, B. L., Ingram, C. J. E., Swallow, D. M. & Thomas, M. G. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* **10,** 36 (2010).

33. Gerbault, P. *et al.* Evolution of lactase persistence: an example of human niche construction. *Philos. Trans. R. Soc. B Biol. Sci.* **366,** 863–877 (2011).

34. Aoki, K. A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 2929–33 (1986).

35. Holden, C. & Mace, R. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum. Biol.* **69,** 605–628 (1997).

36. Ferguson, A. & Maxwell, J. D. Genetic aetiology of lactose intolerance. *Lancet* 188–191 (1967).

37. Sahi, T. & Launiala, K. More evidence for the recessive inheritance of selective adult type lactose malabsorption. *Gastroenterology* **73,** 231—232 (1977).

38. Ho, M. W., Povey, S. & Swallow, D. Lactase polymorphism in adult British natives: estimating allele frequencies by enzyme assays in autopsy samples. *Am. J. Hum. Genet.* **34,** 650–7 (1982).

39. Flatz, G. Gene-dosage effect on intestinal lactase activity demonstrated in vivo. *Am. J. Hum. Genet.* **36,** 306–10 (1984).

40. Wang, Y. *et al.* The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. *Hum. Mol. Genet.* **4,** 657–662 (1995).

41. Mulcare, C. a *et al.* The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* **74,** 1102–1110 (2004).

42. Ingram, C. J. E. *et al.* A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? *Hum. Genet.* **120,** 779–788 (2007).

43. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Rev. Cancer* **39,** 31–40 (2016).

44. Järvelä, I. E. Molecular genetics of adult-type hypolactasia. *Scand. J. Clin. Lab. Invest.* **65,** 535–540 (2005).

45. Perino, A., Cabras, S., Obinu, D. & Cavalli Sforza, L. Lactose intolerance: a non-allergic disorder often managed by allergologists. *Eur. Ann. Allergy Clin. Immunol.* **41,** 3–16 (2009).

46. Arola, H. Diagnosis of hypolactasia and lactose malabsorption. *Scand. J. Gastroenterol. Suppl.* **202,** 26–35 (1994).

47. Dahlqvist, A. Method for assay of intestinal disaccharidases. *Anal. Biochem.* **7,** 18–25 (1964).

48. Dahlqvist, A. Assay of intestinal disaccharidases. *Anal. Biochem.* **22,** 99–107 (1968).

49. Rasinperä, H. *et al.* A genetic test which can be used to diagnose adult-type hypolactasia in children. *Gut* **53,** 1571–1576 (2004).

50. Poulter, M. *et al.* The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann. Hum. Genet.* **67,** 298–311 (2003).

51. Ridefelt, P. & Håkansson, L. D. Lactose intolerance: lactose tolerance test versus genotyping. *Scand. J. Gastroenterol.* **40,** 822–826 (2005).

52. Di Rienzo, T. *et al.* Lactose intolerance: from diagnosis to correct management. *Eur. Rev. Med. Pharmacol. Sci.* **17,** 18–25 (2013).

53. Rosado, J. L. & Solomons, N. W. Sensitivity and specificity of the hydrogen breath-analysis test for detecting malabsorption of physiological doses of lactose. *Clin. Chem.* **29,** 545–548 (1983).

54. Beyerlein, L. *et al.* Correlation between symptoms developed after the oral ingestion of 50 g lactose and results of hydrogen breath testing for lactose intolerance. *Aliment. Pharmacol. Ther.* **27,** 659–665 (2008).

55. Calloway, D. J., Murphy, E. L. & Bauer, D. METHODS AND TECHNICS Determination of l , actose intolerance by Breath Analysis. *Am. J. Digetsive Dis.* **14,** 811–815 (1969).

56. Metz, G., Peters, T. J., Jenkins, D. J. A., Newman, A. & Blendis, L. M. Breath Hydrogen as a Diagnostic Method for Hypolactasia. *Lancet* 1155–1157 (1975).

57. Newcomer, A. D. . & McGill, D. B. Lactose tolerance test in adults with normal lactase activity. *Gastroenterology* **50,** 340–346 (1966).

58. Szilagyi, A. Adult lactose digestion status and effects on disease. *Can. J. Gastroenterol. Hepatol.* **29,** 149–56 (2015).

59. Isokoski, M., Jussila, J. & Sarna, S. A simple screening method for lactose malabsorption. *Gastroenterology* **62,** 28–32 (1972).

60. Arola, H. *et al.* Comparison of Indirect Diagnostic Methods for Hypolactasia. *Scand. J. Gastroenterol.* **23,** 351–357 (1988).

61. Jussila, J. Diagnosis of Lactose Malabsorption by the Lactose Tolerance Test with Peroral Ethanol Administration. *Scand. J. Gastroenterol.* **4,** 361–368 (1969).

62. Martínez-Pardo, M., Montes, P. G., Martin-Lomas, M. & Sols, A. Intestinal lactase evaluation in vivo with 3-methyllactose. *FEBS Lett.* **98,** 99–102 (1979).

63. Fernandez-Mayoralas, A. & Martin-Lomas, M. 4-0-β-D-Galactopyranosyl-3-O-Methyl-D-Glucose: Thesis and Application To the Evaluation Lactase*. *Carbohydr. Res.* **140,** 81–91 (1985).

64. Rivera-Sagredo, a, Cañada, F. J., Nieto, O., Jimenez-Barbero, J. & Martín-Lomas, M. Substrate specificity of small-intestinal lactase. Assessment of the role of the substrate hydroxyl groups. *Eur. J. Biochem.* **209,** 415–22 (1992).

65. Csáky, T. Z. & Glenn, J. E. Urinary recovery of 3-Methylglucose administered to Rats. *Am. J. Physiol.* **188,** 159–163 (1957).

66. Aragón, J. J. *et al.* Evaluation of rat intestinal lactase in vivo with 4-galactosylxylose. *Clin. Chim. acta* **210,** 221–226 (1992).

67. Rivera-sagredo, A., Martin-lomast, M., Jimenez-barbero, J., Villanueva, D. & Aragon, J. J. 4-O-β-D-Galactopyranosyl-D-xylose: a new synthesis and application to the evaluation of intestinal lactase. *Carbohydr. Res.* **228,** 129–135 (1992).

68. Craig, R. M. & Atkinson, A. J. D-Xylose Testing : A Review. *Gastroenterologytr* **95,** 223–231 (1988).

69. Hermida, C. *et al.* Improvement and validation of d-xylose determination in urine and serum as a new tool for the noninvasive evaluation of lactase activity in humans. *J. Clin. Lab. Anal.* **28,** 478–486 (2014).

70. Eberts, T. J., Sample, R. H. B., Glick, M. R. & Ellis, G. H. A simplified, colorimetric micromethod for xylose in serum or urine, with phloroglucinol. *Clin. Chem.* **25,** 1440–1443 (1979).

71. Roe, J. H. & Rice, W. W. A photometric method for the determination of free pentoses in animal tissues. *J. Biol. Chem* **173,** 507–512 (1948).

72. Trinder, P. Micro-determination of xylose in plasma. *Analyst* **100,** 12–15 (1975).

73. Buttery, J. E., Kua, S. L. & De Witt, G. F. The ortho-toluidine method for blood and urine xylose. *Clin. Chim. Acta.* **64,** 325–328 (1975).

74. Braidwood, J. L. & Smith, M. B. A sensitive automated method for the determination of serum and urine xylose. *Clin. Biochem.* **9,** 19–21 (1976).

75. Johnson, S. L. & Mayersohn, M. Quantitation of xylose from plasma and urine by capillary column gas chromatography. *Clin. Chim. Acta* **137,** 13–20 (1984).

76. Mori, L. A method for determination of xylose utilizing glucose dehydrogenase. *Clin. Biochem.* **27,** 21–23 (1994).

77. Hermida, C. *et al.* Phase I and phase IB clinical trials for the noninvasive evaluation of intestinal lactase with 4-galactosylxylose (gaxilose). *J. Clin. Gastroenterol.* **47,** 501–508 (2013).

78. Aragon, J. J. *et al.* Noninvasive diagnosis of hypolactasia with 4-Galactosylxylose (Gaxilose): a multicentre, open-label, phase IIB-III nonrandomized trial. *J. Clin. Gastroenterol.* **48,** 29–36 (2014).

79. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).

80. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** 490–495 (2014).

81. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16,** 29–37 (2006).

82. Breton, C., Fournel-Gigleux, S. & Palcic, M. M. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **22,** 540–549 (2012).

83. Larriba, G., Elorza, M. V., Villanueva, J. R. & Sentandreu, R. Participation of dolichol phospho-mannose in the glycosylation of yeast wall manno-proteins at the polysomal level. *FEBS Lett.* **71,** 316–320 (1976).

84. Centrone, C. A. & Lowary, T. L. Synthesis and Antituberculosis Activity of C-Phosphonate Analogues of Decaprenolphosphoarabinose, a Key Intermediate in the Biosynthesis of Mycobacterial Arabinogalactan and Lipoarabinomannan. *J. Org. Chem.* **67,** 8862–8870 (2002).

85. Schuman, B., Evans, S. V. & Fyles, T. M. Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *PLoS One* **8,** (2013).

86. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28,** 131–141 (2014).

87. De Luca, C. *et al.* Enzymic Synthesis of Hyaluronic Acid with Regeneration of Sugar Nucleotides. *J. Am. Chem. Soc.* **117,** 5869–5870 (1995).

88. Davies, G. J. & Henrissat, B. Structures and Mechanisms of Glycosyl Hydrolases. *Structure* **3,** 853 – 859 (1995).

89. Faijes, M. & Planas, A. In vitro synthesis of artificial polysaccharides by glycosidases and glycosynthases. *Carbohydr. Res.* **342,** 1581–94 (2007).

90. Kittl, R. & Withers, S. G. New approaches to enzymatic glycoside synthesis through directed evolution. *Carbohydr. Res.* **345,** 1272–9 (2010).

91. Planas, A., Faijes, M. & Codera, V. in *Carbohydr. Chem. State Art Challenges Drug Dev. - An Overv. Struct. Biol. Roles, Synth. Methods Appl. As Ther. (Cipolla, Laura)* (World Scientific Publishing Company, 2015). at <https://books.google.de/books?id=rNy3CgAAQBAJ>

92. Cantarel, B. I. *et al.* The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* **37,** 233–238 (2009).

93. Desmet, T. & Soetaert, W. Enzymatic glycosyl transfer: mechanisms and applications. *Biocatal. Biotransformation* **29,** 1–18 (2011).

94. McCarter, J. D. & Withers, S. G. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4,** 885–892 (1994).

95. Hahn, M., Pons, J., Planas, A., Querol, E. & Heinemann, U. Crystal structure of Bacillus licheniformis 1,3-1,4-β-D-glucan 4-glucanohydrolase at 1.8 A resolution. *FEBS Lett.* **374,** 221–224 (1995).

96. Aragunde, H., Castilla, E., Biarnés, X., Faijes, M. & Planas, A. A transitional hydrolase to glycosynthase mutant by Glu to Asp substitution at the catalytic nucleophile in a retaining glycosidase. *Carbohydr. Res.* **389,** 85–92 (2014).

97. Withers, S. G. Mechanisms of glycosyl transferases and hydrolases. *Carbohydr. Polym.* **44,** 325–337 (2001).

98. Davies, G. J. *et al.* Snapshots along an Enzymatic Reaction Coordinate: Analysis of a Retaining β-Glycoside Hydrolase,. *Biochemistry* **37,** 11707–11713 (1998).

99. Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J. & Withers, S. G. Changing Enzymic Reaction Mechanisms by Mutagenesis: Conversion of a Retaining Glucosidase to an Inverting Enzyme. *J. Am. Chem. Soc.* **116,** 11594–11595 (1994).

100. Terwisscha van Scheltinga, a C. *et al.* Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry* **34,** 15619–15623 (1995).

101. Iseli, B., Armand, S., Boller, T., Neuhaus, J. M. & Henrissat, B. Plant chitinases use two different hydrolytic mechanisms. *FEBS Lett.* **382,** 186–188 (1996).

102. Tews, I., Terwisscha Van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. Substrate-assisted catalysis unifies two families of chitinolytic enzymes. *J. Am. Chem. Soc.* **119,** 7954–7959 (1997).

103. Vic, G., Biton, J., Le Beller, D., Michel, J. M. & Thomas, D. Enzymatic glucosylation of hydrophobic alcohols in organic medium by the reverse hydrolysis reaction using almond-beta-D-glucosidase. *Biotechnol. Bioeng.* **46,** 109–116 (1995).

104. Panintrarux, C., Adachi, S., Araki, Y., Kimura, Y. & Matsuno, R. Equilibrium yield of n-alkyl-β-d-glucoside through condensation of glucose and n-alcohol by β-glucosidase in a biphasic system. *Enzyme Microb. Technol.* **17,** 32–40 (1995).

105. Fan, J.-Q. *et al.* Enhanced Transglycosylation Activity of Arthrobacter protophormiae Endo-thermodynamic controlled synthesis-N-acetylglucosaminidase in Media containg organic solvents. *J. Biol. Chem.* **270,** 17723–17729 (1995).

106. Zeuner, B., Riisager, A., Mikkelsen, J. D. & Meyer, A. S. Improvement of trans-sialylation versus hydrolysis activity of an engineered sialidase from Trypanosoma rangeli by use of co-solvents. *Biotechnol. Lett.* **36,** 1315–1320 (2014).

107. Ajisaka, K., Nishida, H. & Fujimoto, H. The synthesis of oligosaccharides by the reversed hydrolysis reaction of β-glucosidase at high substrate concentration and at high temperature. *Biotechnol. Lett.* **9,** 243–248 (1987).

108. Mladenoska, I., Winkelhausen, E. & Kuzmanova, S. Transgalactosylation / Hydrolysis Ratios of Various β-Galactosidases Catalyzing Alkyl-β-Galactoside Synthesis in Single-Phased Alcohol Media. *Food Technol. Biotechnol* **46,** 311–316 (2008).

109. Osman, A., Tzortzis, G., Rastall, R. A. & Charalampopoulos, D. BbgIV is an important Bifidobacterium β-galactosidase for the synthesis of prebiotic galactooligosaccharide at high temperatures. *J. Agric. Food Chem.* **60,** 740–748 (2012).

110. Planas, A. & Faijes, M. Glycosidases and Glycosynthases in enzymatic synthesis of oligosaccharides. An overview. *Afinidad* **59,** 295–313 (2002).

111. Usui, T., Matsui, H. & Isobe, K. Enzymic synthesis of useful chito-oligosaccharides utilizing transglycosylation by chitinolytic enzymes in a buffer containing ammonium sulfate. *Carbohydr. Res.* **203,** 65–77 (1990).

112. Fujimoto, H., Nishida, H. & Ajisaka, K. Enzymatic Syntheses of Glucobioses by a Condensation Reaction with alpha-glucosidase, beta-glucosdase and glucoamylase. *Agric. Biol. Chem.* **52,** 1345–1351 (1988).

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

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

115. Kim, Y.-W., Chen, H. & Withers, S. G. Enzymatic transglycosylation of xylose using a glycosynthase. *Carbohydr. Res.* **340,** 2735–2741 (2005).

116. Hommalai, G., Withers, S. G., Chuenchor, W., Cairns, J. R. K. & Svasti, J. Enzymatic synthesis of cello-oligosaccharides by rice BGlu1 β-glucosidase glycosynthase mutants. *Glycobiology* **17,** 744–753 (2007).

117. Pengthaisong, S., Withers, S. G., Kuaprasert, B., Svasti, J. & Ketudat Cairns, J. R. The role of the oligosaccharide binding cleft of rice BGlu1 in hydrolysis of cellooligosaccharides and in their synthesis by rice BGlu1 glycosynthase. *Protein Sci.* **21,** 362–372 (2012).

118. Henze, M. *et al.* Rational design of a glycosynthase by the crystal structure of β-galactosidase from Bacillus circulans (BgaC) and its use for the synthesis of N-acetyllactosamine type 1 glycan structures. *J. Biotechnol.* **191,** 78–85 (2014).

119. Faijes, M., Fairweather, J. K., Driguez, H. & Planas, A. Oligosaccharide Synthesis by Coupled endo-Glycosynthases of Different Specificity: A Straightforward Preparation of Two Mixed-Linkage Hexasaccharide Substrates of 1,3/1,4-β-Glucanases. *Chem. – A Eur. J.* **7,** 4651–4655 (2001).

120. Pérez, X., Faijes, M. & Planas, A. Artificial Mixed-Linked β-Glucans Produced by Glycosynthase-Catalyzed Polymerization: Tuning Morphology and Degree of Polymerization. *Biomacromolecules* **12,** 494–501 (2011).

121. Codera, V., Gilbert, H. J., Faijes, M. & Planas, A. Carbohydrate Binding Module assisting glycosynthase-catalyzed polymerizations. *Biochem. J.* (2015). doi:10.1042/BJ20150420

122. Ben-David, A., Shoham, G. & Shoham, Y. A Universal Screening Assay for Glycosynthases: Directed Evolution of Glycosynthase XynB2(E335G) Suggests a General Path to Enhance Activity. *Chem. Biol.* **15,** 546–551 (2008).

123. Andres, E., Aragunde, H. & Planas, A. Screening glycosynthase libraries with a fluoride chemosensor assay independently of enzyme specificity: identification of a transitional hydrolase to synthase mutant. *Biochem. J.* **458,** 355–363 (2014).

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

125. Wilson, D. S. & Keefe, A. D. in *Curr. Protoc. Mol. Biol.* (John Wiley & Sons, Inc., 2001). at <http://dx.doi.org/10.1002/0471142727.mb0803s51>

126. Chusacultanachai, S. & Yuthavong, Y. in *Parasite Genomics Protoc.* (Melville, S. E.) 319–333 (Humana Press, 2004). at <https://doi.org/10.1385/1-59259-793-9:319>

127. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2,** 891–903 (2007).

128. Reetz, M. T., Wang, L. W. & Bocola, M. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angew. Chemie - Int. Ed.* **45,** 1236–1241 (2006).

129. Bissaro, B. *et al.* Molecular Design of Non-Leloir Furanose-Transferring Enzymes from an α-L-Arabinofuranosidase: A Rationale for the Engineering of Evolved Trasnglycosylases. *ACS Catal.* **5,** 4598–4611 (2015).

130. Feng, H.-Y. *et al.* Converting a β-glycosidase into a β-transglycosidase by directed evolution. *J. Biol. Chem.* **280,** 37088–97 (2005).

131. Hou, X. E. Z., Ang, Y. W., Eehan, E. J. M. & Hen, L. C. Mutation of a Conserved Tryptophan in the Chitin-Binding Cleft of Serratia marcescens Chitinase A Enhances Transglycosylation. **70,** 243–251 (2006).

132. Merceron, R. *et al.* The molecular mechanism of thermostable α-galactosidases AgaA and AgaB explained by x-ray crystallography and mutational studies. *J. Biol. Chem.* **287,** 39642–39652 (2012).

133. Teze, D. *et al.* Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. *Protein Eng. Des. Sel.* **27,** 13–9 (2014).

134. Kuriki, T. *et al.* Controlling substrate preference and transglycosylation activity of Neupullulanase by manipulating steric constraint and hydrophobicity in active center. *J. Biol. Chem.* **271,** 17321–17329 (1996).

135. Yin, J. *et al.* Structural basis and catalytic mechanism for the dual functional endo-beta-N-acetylglucosaminidase A. *PLoS One* **4,** e4658 (2009).

136. Teze, D. *et al.* Dynamic of conserved water molecules in family 1 glycoside hydrolases : an X-ray , DXMS and Molecular Dynamics study. **52,** 5900–5910 (2013).

137. Husain, Q. β-galactosidases and their potential applications: a review. *Crit. Rev. Biotechnol.* **30,** 41–62 (2010).

138. Asraf, S. S. & Gunasekaran, P. Current trends of ß-galactosidase research and application. 880–890 (2010).

139. Shukla, T. P. & Wierzbicki, L. E. β-galactosidase technology: A solution to the lactose problem. *C R C Crit. Rev. Food Technol.* **5,** 325–356 (1975).

140. Ladero, M., Perez, M. T., Santos, a. & Garcia-Ochoa, F. Hydrolysis of lactose by free and immobilized β-galactosidase from thermus sp. strain T2. *Biotechnol. Bioeng.* **81,** 241–252 (2003).

141. Panesar, P. S., Panesar, R., Singh, R. S., Kennedy, J. F. & Kumar, H. Microbial production, immobilization and applications of β-D-galactosidase. *J. Chem. Technol. Biotechnol.* **81,** 530–543 (2006).

142. Mahoney, R. R. Galactosyl-oligosaccharide formation during lactose hydrolysis: A review. *Food Chem.* **63,** 147–154 (1998).

143. Panesar, P. S., Kumari, S. & Panesar, R. Potential Applications of Immobilized β-Galactosidase in Food Processing Industries. *Enzyme Res.* **2010,** 473137 (2010).

144. Haider, T. & Husain, Q. Hydrolysis of milk/whey lactose by β-galactosidase: A comparative study of stirred batch process and packed bed reactor prepared with calcium alginate entrapped enzyme. *Chem. Eng. Process. Process Intensif.* **48,** 576–580 (2009).

145. Das, B., Roy, A. P., Bhattacharjee, S., Chakraborty, S. & Bhattacharjee, C. Lactose hydrolysis by β-galactosidase enzyme: Optimization using response surface methodology. *Ecotoxicol. Environ. Saf.* **121,** 244–252 (2015).

146. Klein, M. P. *et al.* High stability of immobilized β-d-galactosidase for lactose hydrolysis and galactooligosaccharides synthesis. *Carbohydr. Polym.* **95,** 465–470 (2013).

147. Dwevedi, A. & Kayastha, A. M. Optimal immobilization of β-galactosidase from Pea (PsBGAL) onto Sephadex and chitosan beads using response surface methodology and its applications. *Bioresour. Technol.* **100,** 2667–2675 (2009).

148. Garcia-Galan, C., Berenguer-Murcia, Á., Fernandez-Lafuente, R. & Rodrigues, R. C. Potential of Different Enzyme Immobilization Strategies to Improve Enzyme Performance. *Adv. Synth. Catal.* **353,** 2885–2904 (2011).

149. Gosling, A., Stevens, G. W., Barber, A. R., Kentish, S. E. & Gras, S. L. Recent advances refining galactooligosaccharide production from lactose. *Food Chem.* **121,** 307–318 (2010).

150. Tzortzis, G. & Vulevic, J. in *Prebiotics Probiotics Sci. Technol.* (Charalampopoulos, D. & Rastall, R. A.) 207–244 (Springer New York, 2009). at <https://doi.org/10.1007/978-0-387-79058-9_7>

151. Park, A.-R. & Oh, D.-K. Galacto-oligosaccharide production using microbial β-galactosidase: current state and perspectives. *Appl. Microbiol. Biotechnol.* **85,** 1279–86 (2010).

152. Torres, D. P. M., Gonçalves, M. D. P. F., Teixeira, J. a. & Rodrigues, L. R. Galacto-Oligosaccharides: Production, Properties, Applications, and Significance as Prebiotics. *Compr. Rev. Food Sci. Food Saf.* **9,** 438–454 (2010).

153. Petzelbauer, I., Reiter, A., Splechtna, B., Kosma, P. & Nidetzky, B. Transgalactosylation by thermostable β-glycosides from Pyrococcus furiosus and Sulfolobus solfataricus. *Eur. J. Biochem.* **267,** 5055–5066 (2000).

154. Park, H. Y., Kim, H. J., Lee, J. K., Kim, D. & Oh, D. K. Galactooligosaccharide production by a thermostable β-galactosidase from Sulfolobus solfataricus. *World J. Microbiol. Biotechnol.* **24,** 1553–1558 (2008).

155. Boon, M. a, Oost, I. J. V. a N. D. E. R., Vos, W. M. D. E., Janssen, a E. M. & Riet, K. V. a N. T. Synthesis of Oligosaccharides Catalyzed by Thermostable β-Glucosidase from Pyrococcus furiosus. *Appl. Biochem. Biotechnol.* 269–278 (1999).

156. Ji, E. S., Park, N. H. & Oh, D. K. Galacto-oligosaccharide production by a thermostable recombinant β-galactosidase from Thermotoga maritima. *World J. Microbiol. Biotechnol.* **21,** 759–764 (2005).

157. Talens-Perales, D., Polaina, J. & Marín-Navarro, J. Structural Dissection of the Active Site of Thermotoga maritima β-Galactosidase Identifies Key Residues for Transglycosylating Activity. *J. Agric. Food Chem.* **64,** 2917–2924 (2016).

158. Ball, J. C., Puckett, L. G. & Bachas, L. G. Covalent Immobilization of β-Galactosidase onto a Gold-Coated Magnetoelastic Transducer via a Self-Assembled Monolayer: Toward a Magnetoelastic Biosensor. *Anal. Chem.* **75,** 6932–6937 (2003).

159. Lukacheva, L. V *et al.* Determination of glucose and lactose in food products with the use of biosensors based on Berlin blue. *J. Anal. Chem.* **62,** 388–393 (2007).

160. Amárita, F., Fernández, C. R. & Alkorta, F. Hybrid biosensors to estimate lactose in milk. *Anal. Chim. Acta* **349,** 153–158 (1997).

161. Jacob, F. & Monod, J. Genetic regulatory mechanims in the synthesis of preoteins. *J. Mol. Biol* **3,** 318–356 (1961).

162. Muller-Hill, B. & Kania, J. Lac repressor can fused to β-galactosidase. *Nature* **249,** 561–563 (1974).

163. Brake, A. J., Fowlert, A. V, Zabin, I., Kania, J. & Muller-Hills, B. β-Galactosidase chimeras: Primary structure of a lac repressor-β-galactosidase protein. *Proc. Natl. Acad. Sci. U.S.A.* **75,** 4824–4827 (1978).

164. Shuman, H. a & Silhavy, T. J. The art and design of genetic screens: Escherichia coli. *Nat. Rev. Genet.* **4,** 419–431 (2003).

165. Juers, D. H., Matthews, B. W. & Huber, R. E. LacZ β-galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Sci.* **21,** 1792–807 (2012).

166. Talens-Perales, D., Górska, A., Huson, D. H., Polaina, J. & Marín-Navarro, J. Analysis of Domain Architecture and Phylogenetics of Family 2 Glycoside Hydrolases (GH2). 1–17 (2016).

167. Juers, D. H. *et al.* High resolution refinement of β-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. *Protein Sci.* **9,** 1685–99 (2000).

168. Juers, D. H. *et al.* A structural view of the action of Escherichia coli (lacZ) β-galactosidase. *Biochemistry* **40,** 14781–94 (2001).

169. Gallagher, C. N. & Huber, R. E. Monomer–Dimer Equilibrium of Uncomplemented M15 β-Galactosidase from Escherichia coli. *Biochemistry* **36,** 1281–1286 (1997).

170. Jacobson, R. H., Zhang, X.-J., DuBose, R. F. & Matthews, B. W. Three-dimensional structure of β-galactosidase from E. coli. *Nature* **369,** 761–766 (1994).

171. Matthews, B. W. The structure of E. coli β-galactosidase. *C. R. Biol.* **328,** 549–556 (2005).

172. Juers, D. H., Huber, R. E. & Matthews, B. W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between beta-galactosidase and other glycohydrolases. *Protein Sci.* **8,** 122–36 (1999).

173. Huber, R. E., Hakda, S., Cheng, C., Cupples, C. G. & Edwards, R. A. Trp-999 of β-Galactosidase (Escherichia coli) Is a Key Residue for Binding, Catalysis, and Synthesis of Allolactose, the Natural Lac Operon Inducer. *Biochemistry* **42,** 1796–1803 (2003).

174. Juers, D. H., Hakda, S., Matthews, B. W. & Huber, R. E. Structural Basis for the Altered Activity of Gly794 Variants of Escherichia coli β-Galactosidase. *Biochemistry* **42,** 13505–13511 (2003).

175. Jancewicz, L. J. *et al.* Ser-796 of β-galactosidase (Escherichia coli) plays a key role in maintaining a balance between the opened and closed conformations of the catalytically important active site loop. *Arch. Biochem. Biophys.* **517,** 111–122 (2012).

176. Wheatley, R. W., Lo, S., Jancewicz, L. J., Dugdale, M. L. & Huber, R. E. Structural explanation for allolactose (lac operon inducer) synthesis by lacZ β-galactosidase and the evolutionary relationship between allolactose synthesis and the lac repressor. *J. Biol. Chem.* **288,** 12993–13005 (2013).

177. Dugdale, M. L. *et al.* Importance of Arg-599 of β-galactosidase (Escherichia coli) as an anchor for the open conformations of Phe-601 and the active-site loop. *Biochem. Cell Biol.* **88,** 969–979 (2010).

178. Dugdale, M. L., Dymianiw, D. L., Minhas, B. K., D'Angelo, I. & Huber, R. E. Role of Met-542 as a guide for the conformational changes of Phe-601 that occur during the reaction of β-galactosidase (Escherichia coli). *Biochem. Cell Biol.* **88,** 861–869 (2010).

179. López, R. & Fernández-Mayoralas, A. Controlling yield and regioselectivity in the enzymatic synthesis of β-D-galactopyranosyl-β-D-xylopyranosides. *Tetrahedron Lett.* **33,** 5449–5452 (1992).

180. Lopez, R. & Fernhdez-mayoralas, A. Enzymatic β-Galactosidation of Modified Monosaccharides: Study of the Enzyme Selectivity for the Acceptor and Its Application to the Synthesis of Disaccharides. *J Org Chem* **59,** 737–745 (1994).

181. Aragón, J. J. *et al.* A direct enzymatic synthesis of β-D-galactopyranosyl-D-xylopyranosides and their use to evaluate rat intestinal lactase activity in vivo. *Carbohydr. Res.* **290,** 209–16 (1996).

182. Montero, E., Alonso, J., Cañada, F. J., Fernández-Mayoralas, A. & Martín-Lomas, M. Regioselectivity of the enzymatic transgalactosidation of D- and L- xylose catalysed by β-galactosidases. *Carbohydr. Res.* **305,** 383–391 (1997).

183. Hermida, C., Corrales, G., Cañada, F. J., Aragón, J. J. & Fernández-Mayoralas, A. Optimizing the enzymatic synthesis of β-D-galactopyranosyl-D-xyloses for their use in the evaluation of lactase activity in vivo. *Bioorg. Med. Chem.* **15,** 4836–40 (2007).

184. Hermida, C., Corrales, G., Cañada, F. J., Aragón, J. J. & Fernández-Mayoralas, A. Optimizing the enzymatic synthesis of beta-D-galactopyranosyl-D-xyloses for their use in the evaluation of lactase activity in vivo. *Bioorg. Med. Chem.* **15,** 4836–40 (2007).

185. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

186. Rose, P. W. *et al.* The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **45,** D271–D281 (2017).

187. Martinez-Bilbao, M., Holdsworthz, R. E., Edwardsa, L. A. & Huberl, R. E. A Highly Reactive & Galactosidase (Escherichia coli) Resulting from a substitution of an Aspartic Acid for Gly-794. *J. Biol. Chem.* **266 (8),** 4979–4986 (1991).

188. Henrissat, B., Vegetales, M. & Grenoble, F. A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochem. J.* **280,** 309–316 (1991).

189. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., Vos, W. M. De & de Vos, W. M. Characterization of the celB gene coding for β-glucosidase from the hyperthermophilic archaeon Pyrococcus furiosus and its expression and site-directed mutation in Escherichia coli. *J. Bacteriol.* **177,** 7105–11 (1995).

190. Hansson, T., Kaper, T., van Der Oost, J., de Vos, W. M. & Adlercreutz, P. Improved oligosaccharide synthesis by protein engineering of β-glucosidase CelB from hyperthermophilic Pyrococcus furiosus. *Biotechnol. Bioeng.* **73,** 203–210 (2001).

191. Lin, X., Zhang, W., Chen, Y., Yao, B. & Zhang, Z. F. Overexpression of celB gene coding for beta-glucosidase from Pyrococcus furiosus using a baculovirus expression vector system in silkworm, Bombyx mori. *Z. Naturforsch. C.* **61,** 595–600 (2006).

192. Böhmer, N., Lutz-Wahl, S. & Fischer, L. Recombinant production of hyperthermostable CelB from Pyrococcus furiosus in Lactobacillus sp. *Appl. Microbiol. Biotechnol.* **96,** 903–912 (2012).

193. Dong, Q., Yan, X., Zheng, M. & Yang, Z. Characterization of an extremely thermostable but cold-adaptive β-galactosidase from the hyperthermophilic archaeon Pyrococcus furiosus for use as a recombinant aggregation for batch lactose degradation at high temperature. *J. Biosci. Bioeng.* **117,** 706–710 (2014).

194. Lang, M., Kamrat, T. & Nidetzky, B. Influence of Ionic Liquid Cosolvent on Transgalactosylation Reactions Catalyzed by Thermostable β-Glycosylhydrolase CelB from Pyrococcus Furiosus. *Biotechnol. Bioeng.* **95 (6),** 1093–1100 (2006).

195. Dabrowski, S., Sobiewska, G., Maciunska, J., Synowiecki, J. & Kur, J. Cloning, Expression, and Purification of the His-Tagged Thermostable β-Galactosidase form Pyrococcus woesei in Escherichia coli and Somo Properties of the Isolated Enzyme. *Protein Expr. Purif.* **19,** 107–112 (2000).

196. Synowiecki, J. & Maciunska, J. Isolation and some properties of the thermostable β-galactosidase of Pyrococcus woesei expressed in Escherichia coli. *J. Food Biochem.* **26,** 49–62 (2002).

197. Wanarska, M., Kur, J., Pladzyk, R. & Turkiewicz, M. Thermostable Pyrococcus woesei β-D-galactosidase - high level expression , purification and biochemical properties. *Crit. Rev.* **52,** 781–787 (2005).

198. Xiong, A. S. *et al.* Directed evolution of a β-galactosidase from Pyrococcus woesei resulting in increased thermostable beta-glucuronidase activity. *Appl. Microbiol. Biotechnol.* **77,** 569–578 (2007).

199. Henze, M., Merker, D. & Elling, L. Microwave-assisted synthesis of glycoconjugates by transgalactosylation with recombinant thermostable β-glycosidase from Pyrococcus. *Int. J. Mol. Sci.* **17,** (2016).

200. Park, A. R., Kim, H. J., Lee, J. K. & Oh, D. K. Hydrolysis and transglycosylation activity of a thermostable recombinant β-glycosidase from sulfolobus acidocaldarius. *Appl. Biochem. Biotechnol.* **160,** 2236–2247 (2010).

201. Grogan, D. W. Evidence that β-Galactosidase of Sulfolobus solfataricus Is Only One of Several Activities of a Thermostable β-d-Glycosidase. *Appl. Environ. Microbiol.* **57,** 1644–1649 (1991).

202. Moracci, M., Lavolpe, A., Pulitzer, J. F., Rossi, M. & Ciaramella, M. Expression of the Thermostable β-Galactosidase Gene from the Archaebacterium Sulfolobus-Solfataricus in Saccharomyces-Cerevisiae and Characterization of a New Inducible Promoter for Heterologous Expression. *J. Bacteriol.* **174,** 873–882 (1992).

203. Kim, Y. S., Park, C. S. & Oh, D. K. Lactulose production from lactose and fructose by a thermostable β-galactosidase from Sulfolobus solfataricus. *Enzyme Microb. Technol.* **39,** 903–908 (2006).

204. Sun, X., Duan, X., Wu, D., Chen, J. & Wu, J. Characterization of Sulfolobus solfataricus β-galactosidase mutant F441Y expressed in Pichia pastoris. *J. Sci. Food Agric.* **94,** 1359–1365 (2014).

205. Yu, L. & O'Sullivan, D. J. Production of galactooligosaccharides using a hyperthermophilic β-galactosidase in permeabilized whole cells of Lactococcus lactis. *J. Dairy Sci.* **97,** 694–703 (2014).

206. Cho, Y.-J., Shin, H.-J. & Bucke, C. Purification and biochemical properties of a galactooligosaccharide producing β -galactosidase from Bullera singularis. *Biotechnol. Lett.* **25,** 2107–11 (2003).

207. Gorin, P. a. J., Spencer, J. F. T. & Phaff, H. J. The Structures of Galactosyl-Lactose and Galactobiosyl- Lactose Produced From Lactose By Sporobolomyces Singularis. *Can. J. Chem.* **42,** 1341–1344 (1964).

208. Gorin, P. A. J. & Pei, H. J. The synthesis of β-galacto- and β-gluco-pyranosyl disaccharides by Sporobolomyces Singularis. *Can. J. Chem.* **42,** 2307–2317 (1964).

209. Ishikawa, E., Sakai, T., Ikemura, H., Matsumoto, K. & Abe, H. Identification, cloning, and characterization of a Sporobolomyces singularis β-galactosidase-like enzyme involved in galacto-oligosaccharide production. *J. Biosci. Bioeng.* **99,** 331–339 (2005).

210. Shin, H. J., Park, J. M. & Yang, J. W. Continuous production of galacto-oligosaccharides from lactose by Bullera singularis β-galactosidase immobilized in chitosan beads. *Process Biochem.* **33,** 787–792 (1998).

211. Nakagawa, T., Fujimoto, Y., Ikehata, R., Miyaji, T. & Tomizuka, N. Purification and molecular characterization of cold-active β-galactosidase from Arthrobacter psychrolactophilus strain F2. *Appl. Microbiol. Biotechnol.* **72,** 720–725 (2006).

212. Nakagawa, T., Ikehata, R., Myoda, T., Miyaji, T. & Tomizuka, N. Overexpression and functional analysis of cold-active β-galactosidase from Arthrobacter psychrolactophilus strain F2. *Protein Expr. Purif.* **54,** 295–299 (2007).

213. Karasová-Lipovová, P. *et al.* The cloning, purification and characterisation of a cold-active β-galactosidase from the psychrotolerant Antarctic bacterium Arthrobacter sp. C2-2. *Enzyme Microb. Technol.* **33,** 836–844 (2003).

214. Pawlak-Szukalska, A., Wanarska, M., Popinigis, A. T. & Kur, J. A novel cold-active β-d-galactosidase with transglycosylation activity from the Antarctic Arthrobacter sp. 32cB - Gene cloning, purification and characterization. *Process Biochem.* **49,** 2122–2133 (2014).

215. Song, J. *et al.* Cloning and Expression of a β-Galactosidase Gene of *Bacillus circulans*. *Biosci. Biotechnol. Biochem.* **75,** 1194–1197 (2011).

216. Song, J. *et al.* The discoidin domain of Bacillus circulans β-galactosidase plays an essential role in repressing galactooligosaccharide production. *Biosci. Biotechnol. Biochem.* **77,** 73–9 (2013).

217. Bultema, J. B., Kuipers, B. J. H. & Dijkhuizen, L. Biochemical characterization of mutants in the active site residues of the β-galactosidase enzyme of Bacillus circulans ATCC 31382. *FEBS Open Bio* **4,** 1015–1020 (2014).

218. Ishikawa, K. *et al.* Crystal structure of β-galactosidase from Bacillus circulans ATCC 31382 (BgaD) and the construction of the thermophilic mutants. *FEBS J.* **282,** 2540–2552 (2015).

219. Rodriguez-Colinas, B., Poveda, A., Jimenez-Barbero, J., Ballesteros, A. O. & Plou, F. J. Galacto-oligosaccharide synthesis from lactose solution or skim milk using the β-galactosidase from Bacillus circulans. *J. Agric. Food Chem.* **60,** 6391–8 (2012).

220. Yin, H., Pijning, T., Meng, X., Dijkhuizen, L. & van Leeuwen, S. S. Engineering of the Bacillus circulans β-Galactosidase Product Specificity. *Biochemistry* acs.biochem.7b00032 (2017). at <http://pubs.acs.org/doi/abs/10.1021/acs.biochem.7b00032>

221. Li, Y. *et al.* Purification and characterization of a novel β-galactosidase with transglycosylation activity from bacillus megaterium 2-37-4-1. *Appl. Biochem. Biotechnol.* **158,** 192–199 (2009).

222. Møller, P. L. *et al.* Intra- and Extracellular β-galactosidases from Bifidobacterium bifidum and B. infantis: Molecular Cloning, Heterologous Expression, and Comparative Characterization. *Appl. Environ. Microbiol.* **67,** 2276–2283 (2001).

223. Dumortier, V., Brassart, C. & Bouquelet, S. Purification and properties of β-D-galactosidase from Bifidobacterium bifidum exhibiting a transgalactosylation reaction. *Biotechnol. Appl. Biochem.* **19,** 341–354 (1994).

224. Goulas, T. K., Goulas, A. K., Tzortzis, G. & Gibson, G. R. Molecular cloning and comparative analysis of four β-galactosidase genes from Bifidobacterium bifidum NCIMB41171. *Appl. Microbiol. Biotechnol.* **76,** 1365–1372 (2007).

225. Goulas, A., Tzortzis, G. & Gibson, G. R. Development of a process for the production and purification of α- and β-galactooligosaccharides from Bifidobacterium bifidum {NCIMB} 41171. *Int. Dairy J.* **17,** 648–656 (2007).

226. Goulas, T., Goulas, A., Tzortzis, G. & Gibson, G. R. Comparative analysis of four β-galactosidases from Bifidobacterium bifidum NCIMB41171: purification and biochemical characterisation. *Appl. Microbiol. Biotechnol.* **82,** 1079–88 (2009).

227. Goulas, T., Goulas, A., Tzortzis, G. & Gibson, G. R. Expression of four β-galactosidases from Bifidobacterium bifidum NCIMB41171 and their contribution on the hydrolysis and synthesis of galactooligosaccharides. *Appl. Microbiol. Biotechnol.* **84,** 899–907 (2009).

228. Lu, L., Xiao, M., Xu, X., Li, Z. & Li, Y. A novel β-galactosidase capable of glycosyl transfer from Enterobacter agglomerans B1. *Biochem. Biophys. Res. Commun.* **356,** 78–84 (2007).

229. Lu, L., Xiao, M., Li, Z., Li, Y. & Wang, F. A novel transglycosylating β-galactosidase from Enterobacter cloacae B5. *Process Biochem.* **44,** 232–236 (2009).

230. Lu, L. *et al.* Synthesis of novel galactose containing chemicals by beta-galactosidase from Enterobacter cloacae B5. *Bioresour. Technol.* **101,** 6868–72 (2010).

231. Qi, T., Gu, G., Xu, L., Xiao, M. & Lu, L. Efficient synthesis of tyrosol galactosides by the β-galactosidase from Enterobacter cloacae B5. *Appl. Microbiol. Biotechnol.* (2017). at <http://link.springer.com/10.1007/s00253-017-8249-x>

232. Hall, B. G. & Reeve, E. C. R. A third β-galactosidase in a strain of Klebsiella that possesses two lac genes. *J. Bacteriol.* **132,** 219–223 (1977).

233. Hall, B. G. Properties of β-galactosidase III: Implications for entry of galactosides into Klebsiella. *J. Bacteriol.* **142,** 433–438 (1980).

234. Wang, H. *et al.* Expression and Characterization of Two β -Galactosidases from Klebsiella pneumoniae 285 in Escherichia coli and their Application in the Enzymatic Synthesis of Lactulose and 1-Lactulose. *Zeitschrift für Naturforsch. C* 1–9 (2014). doi:10.5560/ZNC.2014-0061

235. Padilla, B. *et al.* Evaluation of oligosaccharide synthesis from lactose and lactulose using β-galactosidases from Kluyveromyces isolated from artisanal cheeses. *J. Agric. Food Chem.* **60,** 5134–41 (2012).

236. Shen, Q. *et al.* Enzymatic synthesis and identification of oligosaccharides obtained by transgalactosylation of lactose in the presence of fructose using β-galactosidase from Kluyveromyces lactis. *Food Chem.* **135,** 1547–54 (2012).

237. Verma, M. L., Barrow, C. J., Kennedy, J. F. & Puri, M. Immobilization of β-d-galactosidase from Kluyveromyces lactis on functionalized silicon dioxide nanoparticles: characterization and lactose hydrolysis. *Int. J. Biol. Macromol.* **50,** 432–7 (2012).

238. Pereira-Rodríguez, A. *et al.* Structural basis of specificity in tetrameric Kluyveromyces lactis β-galactosidase. *J. Struct. Biol.* **177,** 392–401 (2012).

239. Nguyen, T. H. *et al.* Characterization and molecular cloning of a heterodimeric β-galactosidase from the probiotic strain Lactobacillus acidophilus R22. *FEMS Microbiol. Lett.* **269,** 136–144 (2007).

240. Nguyen, T.-T. *et al.* Homodimeric β-galactosidase from Lactobacillus delbrueckii subsp. bulgaricus DSM 20081: expression in Lactobacillus plantarum and biochemical characterization. *J. Agric. Food Chem.* **60,** 1713–21 (2012).

241. Iqbal, S., Nguyen, T.-H., Nguyen, T. T., Maischberger, T. & Haltrich, D. β-Galactosidase from Lactobacillus plantarum WCFS1: biochemical characterization and formation of prebiotic galacto-oligosaccharides. *Carbohydr. Res.* **345,** 1408–1416 (2010).

242. Iqbal, S. *et al.* Characterization of a heterodimeric GH2 β-galactosidase from lactobacillus sakei Lb790 and formation of prebiotic galacto-oligosaccharides. *J. Agric. Food Chem.* **59,** 3803–3811 (2011).

243. Turkiewicz, M. *et al.* Antarctic marine bacterium Pseudoalteromonas sp. 22b as a source of cold-adapted β-galactosidase. *Biomol. Eng.* **20,** 317–324 (2003).

244. Makowski, K. *et al.* Immobilized preparation of cold-adapted and halotolerant Antarctic β-galactosidase as a highly stable catalyst in lactose hydrolysis. *FEMS Microbiol. Ecol.* **59,** 535–542 (2007).

245. Makowski, K., Białkowska, a., Olczak, J., Kur, J. & Turkiewicz, M. Antarctic, cold-adapted β-galactosidase of Pseudoalteromonas sp. 22b as an effective tool for alkyl galactopyranosides synthesis. *Enzyme Microb. Technol.* **44,** 59–64 (2009).

246. Nakao, M. *et al.* Purification and Characterization of a Thermostable β-Galactosidase with High Transgalactosylation Activity from Saccharopolyspora-Rectivirgula. *Appl. Microbiol. Biotechnol.* **40,** 657–663 (1994).

247. Inohara-Ochiai, M. *et al.* Unique primary structure of a thermostable multimetal β-galactosidase from Saccharopolyspora rectivirgula. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1388,** 77–83 (1998).

248. Inohara-Ochiai, M. *et al.* Deletion and insertion of a 192-residue peptide in the active-site domain of glycosyl hydrolase family-2 β-galactosidases. *J. Biosci. Bioeng.* **93,** 575–583 (2002).

249. Harada, M. *et al.* Divalent metal ion requirements of a thermostable multimetal β-galactosidase from Saccharopolyspora rectivirgula. *J. Biol. Chem.* **269,** 22021–22026 (1994).

250. Nakayama, T. *et al.* Mutational analysis of the role of His452 of Saccharopolyspora rectivirgula β-galactosidase. *J. Biosci. Bioeng.* **90,** 535–539 (2000).

251. Smart, J. B. Transferase reactions of the β-galactosidase from Streptococcus thermophilus. *Appl. Microbiol. Biotechnol.* **34,** 495–501 (1991).

252. Greenberg, N. A. & Mahoney, R. R. Production and Characterization of β-Galactosidase from Streptococcus thermophilus. *J. Food Sci.* **47,** 1824–1835 (1982).

253. Rao, V. R. Production of Beta-Galactosidase from Streptococcus thermophilus Grown in Whey. *Appl. Environ. Microbiol.* **34,** 185–188 (1977).

254. Kim, C. S., Ji, E. S. & Oh, D. K. Characterization of a thermostable recombinant β-galactosidase from Thermotoga maritima. *J. Appl. Microbiol.* **97,** 1006–1014 (2004).

255. Marín-Navarro, J., Talens-Perales, D., Oude-Vrielink, A., Cañada, F. J. & Polaina, J. Immobilization of thermostable β-galactosidase on epoxy support and its use for lactose hydrolysis and galactooligosaccharides biosynthesis. *World J. Microbiol. Biotechnol.* **30,** 989–998 (2014).

256. Moore, J. B., Markiewicz, P. & Miller, J. H. Identification and sequencing of the Thermotoga maritima lacZ gene, part of a divergently transcribed operon. *Gene* **147,** 101–106 (1994).

257. Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. **399,** 323–329 (1999).

258. Toba, T. & Adachi, S. Hydrolysis of Lactose by Microbial β-Galactosidases. Formation of Oligosaccharides with Special Reference to 2-0-β-D-galactopyranosyl-D-glucose. *J. Dairy Sci.* **61,** 33–38 (1978).

259. Zheng, P. *et al.* Production of galacto-oligosaccharides by immobilized recombinant β-galactosidase from Aspergillus candidus. *Biotechnol. J.* **1,** 1464–1470 (2006).

260. Albayrak, N. & Yang, S. Production of Galacto-Oligosaccharides From Lactose by Aspergillus oryzae β -Galactosidase Immobilized on Cotton Cloth. (2001).

261. Gaur, R., Pant, H., Jain, R. & Khare, S. Galacto-oligosaccharide synthesis by immobilized Aspergillus oryzae β-galactosidase. *Food Chem.* **97,** 426–430 (2006).

262. Irazoqui, G. *et al.* Characterization of galactosyl derivatives obtained by transgalactosylation of lactose and different polyols using immobilized β-galactosidase from Aspergillus oryzae. *J. Agric. Food Chem.* **57,** 11302–11307 (2009).

263. Vera, C., Guerrero, C. & Illanes, A. Determination of the transgalactosylation activity of Aspergillus oryzae β-galactosidase: effect of pH, temperature, and galactose and glucose concentrations. *Carbohydr. Res.* **346,** 745–52 (2011).

264. Ito, Y. & Sasaki, T. Cloning and Characterization of the Gene Encoding a Novel β-Galactosidase from Bacillus circulans. *Biosci. Biotech. Biochem* **61,** 1270–1276 (1997).

265. Coombs, J. M. & Brenchley, J. E. Biochemical and Phylogenetic Analyses of a Cold-Active β-Galactosidase from the Lactic Acid Bacterium Carnobacterium pisciola BA. *Appl. Environ. Microbiol* **65,** 5443–5450 (1999).

266. Coombs, J. & Brenchley, J. E. Characterization of two New Glycosyl Hydrolases from the Lactic Acid Bacterium Carnobacterium piscicola strain BA. *Appl. Environ. Microbiol.* **67,** 5094–5099 (2001).

267. Gamauf, C. *et al.* Characterization of the bga1-encoded glycoside hydrolase family 35 β-galactosidase of Hypocrea jecorina with galacto-β-D-galactanase activity. *FEBS J.* **274,** 1691–1700 (2007).

268. Benešová, E., Lipovová, P., Dvořáková, H. & Králová, B. β-D-Galactosidase from Paenibacillus thiaminolyticus catalyzing transfucosylation reactions. *Glycobiology* **20,** 442–451 (2009).

269. Li, Y., Lu, L., Wang, H., Xu, X. & Xiao, M. Cell surface engineering of a β-galactosidase for galactooligosaccharide synthesis. *Appl. Environ. Microbiol.* **75,** 5938–5942 (2009).

270. Zinin, A. I. *et al.* 1-O-Acetyl-β-d-galactopyranose: a novel substrate for the transglycosylation reaction catalyzed by the β-galactosidase from Penicillium sp. *Carbohydr. Res.* **337,** 635–642 (2002).

271. Lauro, B. Di *et al.* Isolation and characterization of a new family 42 β-galactosidase from the thermoacidophilic bacterium Alicyclobacillus acidocaldarius: Identification of the active site residues. *Biochim. Biophys. Acta - Proteins Proteomics* **1784,** 292–301 (2008).

272. Strazzulli, A. *et al.* Introducing transgalactosylation activity into a family 42 β-galactosidase. *Glycobiology* **27,** 425–437 (2017).

273. Tang, L. *et al.* Lactulose biosynthesis by β-galactosidase from a newly isolated Arthrobacter sp. *J. Ind. Microbiol. Biotechnol.* **38,** 471–476 (2011).

274. Mozaffar, Z., Nakanishi, K., Matsuno, R. & Kamikubo, T. Purification and Properties of β-Galactosidases from Bacillus circulans. *Agric. Biol. Chem* **48,** 3053–3061 (1984).

275. Vetere, a & Paoletti, S. Separation and characterization of three beta-galactosidases from Bacillus circulans. *Biochim. Biophys. Acta* **1380,** 223–31 (1998).

276. Hinz, S. W. a, Van Den Broek, L. a M., Beldman, G., Vincken, J. P. & Voragen, A. G. J. β-Galactosidase from Bifidobacterium adolescentis DSM20083 prefers β(1,4)-galactosides over lactose. *Appl. Microbiol. Biotechnol.* **66,** 276–284 (2004).

277. Van Laere, K. M. J., Abee, T., Schols, H. a., Beldman, G. & Voragen, A. G. J. Characterization of a novel β-galactosidase from Bifidobacterium adolescentis DSM 20083 active towards transgalactooligosaccharides. *Appl. Environ. Microbiol.* **66,** 1379–1384 (2000).

278. Rabiu, B. a, Jay, A. J., Gibson, G. R. & Rastall, R. a. Synthesis and Fermentation Properties of Novel Galacto- Oligosaccharides by β-Galactosidases from Bifidobacterium Species. *Appl. Environ. Microbiol.* **67,** 2526–2530 (2001).

279. Tzortzis, G., Goulas, A. K. & Gibson, G. R. Synthesis of prebiotic galactooligosaccharides using whole cells of a novel strain, Bifidobacterium bifidum NCIMB 41171. *Appl. Microbiol. Biotechnol.* **68,** 412–416 (2005).

280. Osman, A., Tzortzis, G., Rastall, R. A. & Charalampopoulos, D. A comprehensive investigation of the synthesis of prebiotic galactooligosaccharides by whole cells of Bifidobacterium bifidum NCIMB 41171. *J. Biotechnol.* **150,** 140–148 (2010).

281. Viborg, A. H. *et al.* Distinct substrate specificities of three glycoside hydrolase family 42 β-galactosidases from Bifidobacterium longum subsp. infantis ATCC 15697. *Glycobiology* **24,** 208–16 (2014).

282. Viborg, A. H. *et al.* A β 1-6 / β 1-3 galactosidase from Bifidobacterium animalis subsp . lactis Bl-04 gives insight into sub-specificities of β -galactoside catabolism within Bifidobacterium. 1–17 (2014).

283. Hsu, C. a, Lee, S. L. & Chou, C. C. Enzymatic production of galactooligosaccharides by beta-galactosidase from Bifidobacterium longum BCRC 15708. *J. Agric. Food Chem.* **55,** 2225–30 (2007).

284. Saishin, N., Ueta, M., Wada, A. & Yamamoto, I. Properties of β-galactosidase purified from Bifidobacterium longum subsp. longum JCM 7052 grown on gum arabic. *J. Biol. Macromol* **10,** 23–31 (2010).

285. Park, A. R. & Oh, D. K. Effects of galactose and glucose on the hydrolysis reaction of a thermostable β-galactosidase from Caldicellulosiruptor saccharolyticus. *Appl. Microbiol. Biotechnol.* **85,** 1427–1435 (2010).

286. Kim, Y.-S., Yeom, S.-J. & Oh, D.-K. Reduction of galactose inhibition via the mutation of β-galactosidase from Caldicellulosiruptor saccharolyticus for lactose hydrolysis. *Biotechnol. Lett.* **33,** 353–358 (2011).

287. Kim, Y.-S. & Oh, D.-K. Lactulose production from lactose as a single substrate by a thermostable cellobiose 2-epimerase from Caldicellulosiruptor saccharolyticus. *Bioresour. Technol.* **104,** 668–672 (2012).

288. Hirata, H., Negoro, S. & Okada, H. Molecular basis of isozyme formation of β-galactosidases in Bacillus stearothermophilus: isolation of two beta-galactosidase genes, bgaA and bgaB. *J. Bacteriol.* **160,** 9–14 (1984).

289. Placier, G., Watzlawick, H., Rabiller, C. & Mattes, R. Evolved β-galactosidases from Geobacillus stearothermophilus with improved transgalactosylation yield for galacto-oligosaccharide production. *Appl. Environ. Microbiol.* **75,** 6312–6321 (2009).

290. Wu, Y. *et al.* Enhancing the production of galacto-oligosaccharides by mutagenesis of Sulfolobus solfataricus β-galactosidase. *Food Chem.* **138,** 1588–1595 (2013).

291. Lu, L., Xiao, M., Li, Z., Li, Y. & Wang, F. A novel transglycosylating β-galactosidase from Enterobacter cloacae B5. *Process Biochem.* **44,** 232–236 (2009).

292. Smart, J. G. T. C. J. The effect of cations on the hydrolysis of lactose and the transferase reactions catalysed by β-galactosidase from six strains of lactic acid bacteria. *Appl Microbiol Biotechnol* **46,** 22–27 (1996).

293. Vasiljevic, T. & Jelen, P. Oligosaccharide production and proteolysis during lactose hydrolysis using crude cellular extracts from lactic acid bacteria. *Lait* **83,** 453–467 (2003).

294. Song, J. *et al.* Causes of the Production of Multiple Forms of β-Galactosidase by Bacillus circulans. *Biosci. Biotechnol. Biochem.* **75,** 268–278 (2011).

295. Warmerdam, A., Paudel, E., Jia, W., Boom, R. M. & Janssen, A. E. M. Characterization of β-galactosidase isoforms from Bacillus circulans and their contribution to GOS production. *Appl. Biochem. Biotechnol.* **170,** 340–58 (2013).

296. Cheng, Y.-M., Hsieh, F.-C. & Meng, M. Functional analysis of conserved aromatic amino acids in the discoidin domain of Paenibacillus β-1,3-glucanase. *Microb. Cell Fact.* **8,** 62 (2009).

297. Hansen, O., J**?**rgensen, F. & Stougaard, P. High-efficiency synthesis of oligosaccharides with a truncated β-galactosidase from Bifidobacterium bifidum. *Appl. Microbiol. Biotechnol.* **57,** 647–652 (2001).

298. Hung, M. & Lee, B. H. Cloning and expression of β-galactosidase gene from Bifidobacterium infantis into Escherichia coli. **20,** 659–660 (1998).

299. Hung, M., Xia, Z. & Hu, N. Molecular and Biochemical Analysis of Two β-Galactosidases from Bifidobacterium infantis HL96. **67,** 4256–4263 (2001).

300. Hung, M.-N. & Lee, B. H. Purification and characterization of a recombinant β-galactosidase with transgalactosylation activity from Bifidobacterium infantis HL96. *Appl. Microbiol. Biotechnol.* **58,** 439–45 (2002).

301. Carter, P. Site-directed mutagenesis. *Biochem. J.* **237,** 1–7 (1986).

302. Cobb, R. E., Chao, R. & Zhao, H. Directed Evolution: Past, Present and Future. **59,** 1432–1440 (2013).

303. Arnold, F. H. & Georgiou, G. Directed Evolution Library Creation. **231,** 1–13 (2003).

304. Lutz, S. Beyond directed evolution - semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21,** 734–743 (2011).

305. Shin, H. & Cho, B. K. Rational protein engineering guided by deep mutational scanning. *Int. J. Mol. Sci.* **16,** 23094–23110 (2015).

306. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29,** 245–251 (2016).

307. Ruff, A. J., Dennig, A. & Schwaneberg, U. To get what we aim for - Progress in diversity generation methods. *FEBS J.* **280,** 2961–2978 (2013).

308. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).

309. Xiao, H., Bao, Z. & Zhao, H. High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res.* **54,** 4011–4020 (2015).

310. Koné, F. M. T., Le Béchec, M., Sine, J.-P., Dion, M. & Tellier, C. Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.* **22,** 37–44 (2009).

311. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

312. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30,** S162–S173 (2009).

313. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37,** D387–D392 (2009).

314. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–W258 (2014).

315. Eramian, D. *et al.* A composite score for predicting errors in protein structure models. *Protein Sci.* **15,** 1653–1666 (2006).

316. Wong, T. S., Zhurina, D. & Schwaneberg, U. The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* **9,** 271–288 (2006).

317. Rasila, T. S., Pajunen, M. I. & Savilahti, H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. *Anal. Biochem.* **388,** 71–80 (2009).

318. McCullum, E. O., Williams, B. A. R., Zhang, J. & Chaput, J. C. in *Vitr. Mutagen. Protoc. Third Ed.* (Braman, J.) 103–109 (Humana Press, 2010). doi:10.1007/978-1-60761-652-8_7

319. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 10747–10751 (1994).

320. Coco, W. M. *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19,** 354–359 (2001).

321. You, C. & Percival Zhang, Y. H. Easy preparation of a large-size random gene mutagenesis library in Escherichia coli. *Anal. Biochem.* **428,** 7–12 (2012).

322. Voigt, C. *Synthetic Biology, Part B: Computer Aided Design and DNA Assembly*. 400–405 (Elsevier Science, 2011). at <https://books.google.de/books?id=9uPvZWiabr4C>

323. Pai, J. C., Entzminger, K. C. & Maynard, J. a. Restriction enzyme-free construction of random gene mutagenesis libraries in E. coli. *Anal. Biochem.* **421,** 640–648 (2013).

324. Luhe, A. L., Ting, E. N. Y., Tan, L., Wu, J. & Zhao, H. Engineering of small sized DNAs by error-prone multiply-primed rolling circle amplification for introduction of random point mutations. *J. Mol. Catal. B Enzym.* **67,** 92–97 (2010).

325. Kipnis, Y., Dellus-Gur, E. & Tawfik, D. S. TRINS: A method for gene modification by randomized tandem repeat insertions. *Protein Eng. Des. Sel.* **25,** 437–444 (2012).

326. Liu, J. & Cropp, A. T. Rational protein sequence diversification by multi-codon scanning mutagenesis. *Methods Mol Biol* **978,** 217–228 (2013).

327. Wong, T. S. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* **32,** 26e–26 (2004).

328. Ostermeier, M., Nixon, A. E., Shim, J. H. & Benkovic, S. J. Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 3562–3567 (1999).

329. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D. & Benkovic, S. J. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 11248–11253 (2001).

330. Sieber, V., Martinez, C. a & Arnold, F. H. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19,** 456–460 (2001).

331. Kolkman, J. A. & Stemmer, W. P. C. Directed evolution of proteins by exon shuffling. *Nat Biotech* **19,** 423–428 (2001).

332. Hiraga, K. & Arnold, F. H. General method for sequence-independent site-directed chimeragenesis. *J. Mol. Biol.* **330,** 287–296 (2003).

333. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* **4,** e6441 (2009).

334. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6,** 242–251 (2011).

335. Quan, J. & Tian, J. in *DNA Cloning Assem. Methods* (Valla, S. & Lale, R.) 103–117 (Humana Press, 2014). doi:10.1007/978-1-62703-764-8_8

336. Dalonso, N. *et al.* Sequence-independent cloning methods for long DNA fragments applied to synthetic biology. *Anal. Biochem.* **530,** 5–8 (2017).

337. Mangas-Sánchez, J. & Adlercreutz, P. Enzymatic preparation of oligosaccharides by transglycosylation: A comparative study of glucosidases. *J. Mol. Catal. B Enzym.* **122,** 51–55 (2015).

338. Boudabbous, M. *et al.* Trans-glycosylation capacity of a highly glycosylated multi-specific beta-glucosidase from Fusarium solani. *Bioprocess Biosyst. Eng.* **40,** 559–571 (2017).

339. Qin, Z. *et al.* Catalytic Mechanism of a Novel Glycoside Hydrolase Family 16 "Elongating" beta-Transglycosylase. *J. Biol. Chem.* **292,** 1666–1678 (2017).

340. Mallakuntla, M. K., Vaikuntapu, P. R., Bhuvanachandra, B., Das, S. N. & Podile, A. R. Transglycosylation by a chitinase from Enterobacter cloacae subsp. cloacae generates longer chitin oligosaccharides /631/45/603 /631/45/607 /82/29 /82/58 /82/83 /38/77 /82/16 article. *Sci. Rep.* **7,** 1–12 (2017).

341. Huber, R. E., Kurz, G. & Wallenfels, K. A Quantitation of the Factors Which Affect the Hydrolase and Transgalactosylase Activities of β-Galactosidase (E.coli) on Lactose. *Biochemistry* **15 (9),** 1994–2001 (1976).

342. Huber, R. E., Kurz, G. & Wallenfels, K. A quantitation of the factors which affect the hydrolase and transgalactosylase activities of β-galactosidase (E. coli) on lactose. *Biochemistry* **15,** 1994–2001 (1976).

343. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal Escherichia coli. *Nat. Rev. Microbiol.* **8,** 207 (2010).

344. Hoffmann, H. & Roggenkamp, A. Population Genetics of the Nomenspecies Enterobacter cloacae Population Genetics of the Nomenspecies Enterobacter cloacae. *Appl. Environ. Microbiol.* **69,** 5306–5318 (2003).

345. Liu, W.-Y., Wong, C.-F., Chung, K. M.-K., Jiang, J.-W. & Leung, F. C.-C. Comparative Genome Analysis of Enterobacter cloacae. *PLoS One* **8,** e74487 (2013).

346. Faijes, M., Pérez, X., Pérez, O. & Planas, A. Glycosynthase Activity of Bacillus licheniformis 1,3-1,4-β-Glucanase Mutants: Specificity, Kinetics, and Mechanism. *Biochemistry* **42,** 13304–13318 (2003).

347. Li, T., Tong, X., Yang, Q., Giddens, J. P. & Wang, L.-X. Glycosynthase Mutants of Endoglycosidase S2 Show Potent Transglycosylation Activity and Remarkably Relaxed Substrate Specificity for Antibody Glycosylation Remodeling. *J. Biol. Chem.* **291,** 16508–16518 (2016).

348. Ohnuma, T., Dozen, S., Honda, Y., Kitaoka, M. & Fukamizo, T. A glycosynthase derived from an inverting chitinase with an extended binding cleft. *J. Biochem.* **160,** 93–100 (2016).

349. Giddens, J. P., Lomino, J. V, Amin, M. N. & Wang, L.-X. Endo-F3 Glycosynthase Mutants Enable Chemoenzymatic Synthesis of Core-fucosylated Triantennary Complex Type Glycopeptides and Glycoproteins. *J. Biol. Chem.* **291,** 9356–9370 (2016).

350. Fernandes, P. A. & Ramos, M. J. QM / MM Studies on the -Galactosidase Catalytic Mechanism : Hydrolysis and Transglycosylation Reactions. *J. Chem. Theory Comput* **6,** 421–433 (2010).

351. Raich, L. *et al.* A Trapped Covalent Intermediate of a Glycoside Hydrolase on the Pathway to Transglycosylation. Insights from Experiments and Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **138,** 3325–3332 (2016).

352. Badieyan, S., Bevan, D. R. & Zhang, C. Probing the Active Site Chemistry of β-Glucosidases along the Hydrolysis Reaction Pathway. *Biochemistry* **51,** 8907–8918 (2012).

353. Feng, H. Y. *et al.* Converting a β-glycosidase into a β-transglycosidase by directed evolution. *J. Biol. Chem.* **280,** 37088–37097 (2005).

354. Teze, D. *et al.* Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase 5 65. 1–7 (2013).

355. Matsui, I. *et al.* Roles of the aromatic residues conserved in the active center of Saccharomycopsis α-amylase for transglycosylation and hydrolysis activity. *Biochemistry* **33,** 451–458 (1994).

356. Teze, D., Daligault, F., Ferrières, V., Sanejouand, Y. H. & Tellier, C. Semi-rational approach for converting a GH36 α-glycosidase into an α-transglycosidase. *Glycobiology* **25,** 420–427 (2015).

357. Zakariassen, H., Hansen, M. C., Joranli, M., Eijsink, V. G. H. & Sorlie, M. Mutational effects on transglycosylating activity of family 18 chitinases and construction of a hypertransglycosylating mutant. *Biochemistry* **50,** 5693–5703 (2011).

358. Bobrov, K. S., Borisova, A. S., Eneyskaya, E. V & Ivanen, D. R. Improvement of the Efficiency of Transglycosylation Catalyzed by α Galactosidase from Thermotoga maritima by Protein Engineering. *Biochemistry* **78,** 1112–1123 (2013).

359. Hansson, T. & Adlercreutz, P. Enhanced transglucosylation/hydrolysis ratio of mutants of Pyrococcus furiosus β-glucosidase: Effects of donor concentration, water content, and temperature on activity and selectivity in hexanol. *Biotechnol. Bioeng.* **75,** 656–665 (2001).

360. Kuroki, R., Weaver, L. H. & Matthews, B. W. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8949–8954 (1999).

361. Hinz, S. W. a *et al.* Increasing the transglycosylation activity of α-galactosidase from Bifidobacterium adolescentis DSM 20083 by site-directed mutagenesis. *Biotechnol. Bioeng.* **93,** 122–31 (2006).

362. Roth, N. J., Rob, B. & Huber, R. E. His-357 of β-Galactosidase (E.coli) Interacts with the C3 Hydroxyl in the Transition State and Helps to Mediate Catalysis. *Biochemistry* **1,** 10099–10107 (1998).

363. Huber, R. E., Hlede, I. Y., Roth, N. J., McKenzie, K. C. & Ghumman, K. K. His-391 of β-galactosidase (*Escherichia coli*) promotes catalyes by strong interactions with the transition state. *Biochem. Cell Biol.* **79,** 183–193 (2001).

364. Saumonneau, A. *et al.* Design of an α-l-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology* **26,** 261–269 (2015).

365. Osanjo, G. *et al.* Directed evolution of the α-L-fucosidase from Thermotoga maritima into an α-L-transfucosidase. *Biochemistry* **46,** 1022–1033 (2007).

366. Wheatley, R. W. *et al.* Substitution for Asn460 Cripples β-galactosidase (Escherichia coli) by increasing substrate affinity and decreasing transition state stability. *Arch. Biochem. Biophys.* **521,** 51–61 (2012).

367. Sierks, M. R. & Svensson, B. Kinetic identification of a hydrogen bonding pair in the glucoamylase-maltose transition state complex. *Protein Eng.* **5,** 185–188 (1992).

368. Notenboom, V. *et al.* Insights into transition state stabilization of the β-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. *Nat. Struct. Biol.* **5,** 812 (1998).

369. Brás, N. F., Ramos, M. J. & Fernandes, P. A. DFT studies on the β-glycosidase catalytic mechanism: The deglycosylation step. *J. Mol. Struct. THEOCHEM* **946,** 125–133 (2010).

370. Nerinckx, W., Desmet, T., Piens, K. & Claeyssens, M. An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy. *FEBS Lett.* **579,** 302–312 (2005).

371. Cupples, C. G. & Miller, J. H. Effects of amino acid substitutions at the active site in Escherichia coli β-galactosidase. *Genetics* **120,** 637–44 (1988).

372. Ring, M., Bader, D. E. & Huber, R. E. Site-directed mutagenesis of beta-galactosidase (E. coli) reveals that tyr-503 is essential for activity. *Biochem. Biophys. Res. Commun.* **152,** 1050–1055 (1988).

373. Ring, M. & Huber, R. E. Multiple replacements establish the importance of tyrosine-503 in beta-galactosidase (Escherichia coli). *Arch. Biochem. Biophys.* **283,** 342–350 (1990).

374. Penner, R. M., Roth, N. J., Rob, B., Lay, H. & Huber, R. E. Tyr-503 of beta-galactosidase (Escherichia coli) plays an important role in degalactosylation. *Biochem. Cell Biol.* **77,** 229–236 (1999).

375. Xu, J., McRae, M. A. A., Harron, S., Rob, B. & Huber, R. E. A study of the relationships of interactions between Asp-201, Na+ or K+, and galactosyl C6 hydroxyl and their effects on binding and reactivity of beta-galactosidase. *Biochem. Cell Biol.* **82,** 275–284 (2004).

376. Kelly, R. M. *et al.* Elimination of competing hydrolysis and coupling side reactions of a cyclodextrin glucanotransferase by directed evolution. *Biochem. J.* **413,** 517–525 (2008).

377. Hu, X., Robin, S., O'Connell, S., Walsh, G. & Wall, J. G. Engineering of a fungal β-galactosidase to remove product inhibition by galactose. *Appl. Microbiol. Biotechnol.* **87,** 1773–1782 (2010).

378. Pessela, B. C. C. *et al.* The immobilization of a thermophilic β-galactosidase on Sepabeads supports decreases product inhibition: Complete hydrolysis of lactose in dairy products. *Enzyme Microb. Technol.* **33,** 199–205 (2003).

379. Guven, R. G., Kaplan, A., Guven, K., Matpan, F. & Dogru, M. Effects of various inhibitors on β-galactosidase purified from the thermoacidophilic Alicyclobacillus acidocaldarius subsp. Rittmannii isolated from Antarctica. *Biotechnol. Bioprocess Eng.* **16,** 114–119 (2011).

380. Wells, J. a, Powers, D. B., Bott, R. R., Graycar, T. P. & Estell, D. a. Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1219–23 (1987).

381. Wilson, C. & Agard, D. A. Engineering substarte specificty. *Curr. Opin. Struct. Biol.* **1,** 617–623 (1991).

382. Ielasi, F. S., Verhaeghe, T., Desmet, T. & Willaert, R. G. Engineering the carbohydrate-binding site of Epa1p from Candida glabrata: Generation of adhesin mutants with different carbohydrate specificity. *Glycobiology* **24,** 1312–1322 (2014).

383. Juers, D. H. *et al.* Direct and indirect roles of His-418 in metal binding and in the activity of beta-galactosidase (E. coli). *Protein Sci.* **18,** 1281–92 (2009).

384. Lundemo, P., Adlercreutz, P. & Karlsson, E. N. Improved transferase/hydrolase ratio through rational design of a family 1 β-glucosidase from Thermotoga neapolitana. *Appl. Environ. Microbiol.* **79,** 3400–3405 (2013).

385. Nerinckx, W., Desmet, T. & Claeyssens, M. A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. *FEBS Lett.* **538,** 1–7 (2003).

386. Splechtna, B. *et al.* Production of prebiotic galacto-oligosaccharides from lactose using β-galactosidases from Lactobacillus reuteri. *J. Agric. Food Chem.* **54,** 4999–5006 (2006).

387. Nordvang, R. T. *et al.* It all starts with a sandwich: Identification of sialidases with trans-glycosylation activity. *PLoS One* **11,** 1–16 (2016).

388. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **9,** 1–5 (1996).

389. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30,** 2785–2791 (2009).

390. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31,** 455–461 (2010).

391. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

392. Adams, P. D. *et al.* The Phenix Software for Automated Determination of Macromolecular Structures. *Methods* **55,** 94–106 (2011).

393. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

394. Wu, M. *et al.* Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science (80-.).* **350,** (2015).

395. Kramhoeller, B., Pischetsrieder, M. & Severin, T. Maillard reactions of lactose and maltose. *J. Agric. Food Chem.* **41,** 347–351 (1993).

396. Van Boekel, M. a. J. S. Effect of Heating on Maillard Reaction in Milk. *Food Chem.* **62,** 403–414 (1998).

397. Hellwig, M. & Henle, T. Baking, Ageing, Diabetes: A Short History of the Maillard Reaction. *Angew. Chemie Int. Ed.* **53,** 10316–10329 (2014).

398. Harvey, C. B. *et al.* Regional localization of the lactase-phlorizin hydrolase gene, LCT, to chromosome 2q21. *Ann. Hum. Genet.* **57,** 179–185 (1993).

399. Mantei, N. *et al.* Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *EMBO J.* **7,** 2705–13 (1988).

400. Hermida, C., Corrales, G., Martı, O. H., Ferna, A. & Arago, J. J. Noninvasive Evaluation of Intestinal Lactase with 4-Galactosylxylose : Comparison with 3- and 2-Galactosylxylose and Optimization of the Method in Rats. *Clin. Chem.* **52,** 270–277 (2006).

401. Ohkohchi, N., Himukai, M., Igarashi, Y. & Kasai, M. Mechanism of D-xylose transport in human small intestine. *J. Pediatr. Gastroenterol. Nutr.* **5,** 372–378 (1986).

402. Crane, R. K. Intestinal absorption. *Encycl. Fish Physiol.* 1311–1321 (1960). at <http://www.sciencedirect.com/science/article/pii/B9780123745538000721>

403. Sols, A. The hexokinase activity of the intestinal mucosa. *Biochim. Biophys. Acta* **19,** 144–152 (1956).

404. Pitkänen, E. The conversion of d-xylose into d-threitol in patients without liver disease and in patients with portal liver cirrhosis. *Clin. Chim. Acta* **80,** 49–54 (1977).

405. Chi, C., Chang, H. M., Li, Z., Jameel, H. & Zhang, Z. A method for rapid determination of sugars in lignocellulose prehydrolyzate. *BioResources* **8,** 172–181 (2013).

406. Langgley, K. E., Villarejo, M. R., Fowler, V. A., Zamenhof, P. J. & Zabin, I. Molecular Basis of β-galactosidase α-complentation. *Proc. Natl. Acad. Sci. U.S.A.* **72,** 1254–1257 (1976).

407. Cronan Jr., J. E., Narasimhan, M. L. & Rawlings, M. Insertional restoration of β-galactosidase α-compentation (white-to-blue colony screening) facilities assembly of synthetic genes. *Gene* **70,** 161–169 (1988).

408. Ullmann, A. Complementation in β-Galactosidase: From Protein Structure to Genetic Engineering. *Bioessays* **14,** 201–205 (1992).

409. Fowler, V. A. & Zabin, I. Acid Sequence of β-Galactosidase. *J. Biol. Chem.* **253,** 5521–5525 (1978).

410. Kalnins, a, Otto, K., Rüther, U. & Müller-Hill, B. Sequence of the lacZ gene of Escherichia coli. *EMBO J.* **2,** 593–7 (1983).

411. Lee, N., Francklyn, C. & Hamilton, E. P. Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proc. Natl. Acad. Sci.* **84,** 8814–8818 (1987).

412. Kasner, E., Hunter, C. a, Ph, D., Kariko, K. & Ph, D. Advances in Homology Protein Structure Modeling. *Curr Protein Pept Sci* **7,** 217–227 (2006).

413. Fiser, A. Template-Base Protein Structure Modeling. *Methods Mol Biol* **673,** 73–94 (2010).

414. Fiser, A. & Sali, A. in *Protein Struct.* 167–2006 (2003).

415. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9,** 1797–1804 (2008).

416. Nov, Y. When second best is good enough: Another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.* **78,** 258–262 (2012).

417. Hughes, M. D., Nagel, D. A., Santos, A. F., Sutherland, A. J. & Hine, A. V. Removing the Redundancy From Randomised Gene Libraries. *J. Mol. Biol.* **331,** 973–979 (2003).