

**A metabolic engineering approach for alpha-galactosylceramide production in
*S. cerevisiae***

Marc Caballé Margalef

<http://hdl.handle.net/10803/687826>

Data de defensa: 28-02-2023

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

DOCTORAL THESIS

Title A metabolic engineering approach for α -galactosylceramide production in *S. cerevisiae*

Presented by Marc Caballé Margalef

Centre IQS School of Engineering

Department Bioengineering

Directed by Dra. Magda Fajjes and Dr. Antoni Planas



This thesis was realised thanks to the grant for the hiring of predoctoral research personnel (FI) financed by the Department of Research and Universities of the Catalan Government and for European Social Fund [reference numbers 2018FI_B 00332, 2019 FI_B1 00167 & 2020FI_B2 00200], and grants BFU2016-77427-C2-1-R and PID2019-104350RB-I00 funded by MICIN/AEI.



A la meva família,
a la que toca i a la que es busca,



*“The most important step a man can take. It’s not the first one, is it?
It’s the next one. Always the next step.”*

Brandon Sanderson, Oathbringer



Agraïments

No es pot començar a llegir una tesi sense passar primer pels agraïments, ja que encara que la signi un sol, en una tesi hi ha participat moltes persones, tant a dins com a fora del laboratori. I tampoc es poden escriure els agraïments fins que la tesi ja, sí que sí, està acabada... així que aquestes són les últimes línies que afegiré, escrites amb una mescla d'alegria i pena, i amb això acaba l'etapa del doctorat.

Els primers agraïments són per als meus directors de tesi, la Dra. Magda Faijes i el Dr. Antoni Planas. Sense vosaltres el projecte no hauria estat possible, heu estat els meus pares científics. Primer que tot, gràcies per acollir-me al laboratori de Bioquímica i per la llibertat en escollir el projecte per començar una línia d'enginyeria metabòlica en llevat. Començar projectes així és dur, però deixa una bona satisfacció de feina feta per als que vinguin més endavant. Segon, per tot el recolzament que m'heu donat, tant a nivell científic per enfocar tots els experiments i com ser millor investigador, com a nivell personal, recolzant-me en moments difícils de la tesi, fer-me veure les parts positives dels resultats (especialment tu Magda, ja que amb això t'he donat bons mals de cap i has demostrat tenir una paciència envejable). Tot això m'ha fet créixer com a persona i com a professional, i vosaltres hi heu aportat molt.

Al Dr. Xevi Biarnés, moltes gràcies per la teva ajuda en aquest projecte. L'èxit de la identificació de l'enzim va ser també gràcies amb la teva aportació amb l'anàlisi bioinformàtic i les teves aportacions, però sobretot, per fer del laboratori de Bioquímica un lloc millor.

A la Dra. Cristina Fornaguera, per la teva ajuda amb l'anàlisi de les vesícules i, sobretot, amb el microscopi confocal. Et vas implicar i em vas dedicar un temps per ajudar-me amb tot això que considero molt valuós, i això ho agraeixo molt.

To Prof. Dr. Ir. Marjan de Mey, every chance you had to assist us with our project, which were a number of times, you did it with a smile. Thank you, your help was always priceless, and the project greatly improved with your contribution.

To the members of the Doctoral Thesis Tribunal, thank you for accepting this task and take the time to read and evaluate this thesis. Your comments were very much appreciated. I hope the reading was light and you could find some interesting insights.

A tot el laboratori de Bioquímica, moltes gràcies. Hi ha hagut millors i pitjors moments, però gràcies per fer un lloc amb tan bon ambient per treballar... Patri, moltes gràcies per donar-nos sempre un cop de mà i fer més fàcil portar el lab dia a dia, tu saps prou que a vegades n'acabem fins a dalt. Pau, t'he d'agrair molt que sempre que he vingut m'has ajudat i aportat la teva experiència amb el llevat, des de l'inici fins al final, i ha estat un plaer estar com auxiliar amb tu. Marc, moltes gràcies per les teves contribucions i consells que has anat aportant. Teresa, les teves aportacions en seminaris i presentacions sempre han estat molt instructives, i poder xerrar amb tu sempre ha estat un plaer, moltes gràcies per ser-hi.



I als meus companys de doctorat... moltes gràcies un i mil cops. Sense vosaltres el doctorat hauria pogut ser un malson. Han estat tantes hores al laboratori... però també fora, i en tot moment l'ambient que s'ha creat ha estat excel·lent. Tot el suport que ha anat sorgint entre tots i l'amistat que ha anat sorgint no té preu, entrar a un lloc amb desconeguts i sortir-ne amb amics és sempre una gran sort. Aitor, moltes gràcies no només per la teva ajuda amb el model de l'enzim, sinó també pel bon rotllo que has portat i per ser el nostre subministrador de memes i vídeos virals, sempre ens has portat al dia. Álvaro, moltes gràcies per sestar sempre ajudant a tothom, conversar sobre actualitat "videojueguil" i reparar HPLC amb tu era un plaer la veritat (no sé si hi hauria cap HPLC sencer si no fos per això...). Bernat, vam posar el primer peu dintre el laboratori de Bioquímica alhora i ara posarem l'últim per sortir també "alhora"... Has estat un gran company, dins i fora del lab (i també torreta, Fira, ...), les excursions a la muntanya i les tardes de jocs al confinament sempre seran un gran record... gràcies de tot cor. Mireia, haver coincidit amb tu ha estat de les millors coses que em podia passar, has fet del laboratori un lloc ple de riure i bon rotllo. Gràcies per escoltar-me queixar-me 24/7 i donar-me suport i, encara així, continuar sent amiga meva i valor anar a fer passejos de "iaies". Ah, i sobretot per no treure les teves fotos que vaig anar penjant per tot el lab! Sergi, A.K.A. "Avui Fira?", moltes gràcies per ser-hi, hem de reconèixer que sense tu el lab hauria caigut molts cops... Em quedo amb el nostre projecte empenedor OliGO, i les tardes de Torreta que acabaven a la Fira... llàstima (o sort?) que la pandèmia ens ho va parar. Laia, vam coincidir quan ja estaves a l'últim any de tesi, i els somriures ja escassejaven, però encara així sempre ho feies i ens ajudaves en tot, gràcies. També molt memorable va ser el congrés a Toulouse que recordarem més com el que vam veure l'últim capítol de *Game of Thrones* a l'habitació d'un hotel amb dos alemanys. Núria, m'has ajudat en tot moment del projecte, sobretot en moments crítics, fins i tot quan ja havies marxat del laboratori. La teva ajuda ha estat incalculable (a pesar d'haver-te "aliat amb l'enemic"), moltes gràcies per tot. Cristina i Estela, amb vosaltres vaig coincidir un temps molt breu al laboratori, però suposo que la *old school* de Bioquímica sempre està present i en les trobades sempre ha estat un plaer poder compartir temps amb vosaltres. També una especial menció als estudiants que m'han ajudat amb el projecte. Mónica, vas ser la última i de les que més va aportar, espero poder tornar-t'ho donant-te un cop de mà quan ho necessitis. Jordi, a pesar de no ser estudiant meu, has estat una gran ajuda i espero poder també retornar-t'ho. També als altres que han passat per Bioquímica i m'ha ajudat o han deixat un bon record: Elena, Xavi M., Clàudia, Xavi C., Jordi, Alex... moltes gràcies a vosaltres també.

De IQS m'emporto també bons records de bona gent no solament de Bioquímica. Participar en la organització de les I Jornades de Doctorands IQS em va permetre conèixer gent d'altres laboratoris. I del contacte dia a dia amb el nostre laboratori veí m'emporto també molts records i bona gent. Pol, moltes gràcies per la teva ajuda i consells, i dono gràcies perquè el contacte i l'amistat també ha seguit fora del laboratori i del doctorat. A la Coral, que ja com a estudiant i després com companya m'ha recolzat i escoltat, espero haver-te tornat també aquest recolzament quan he pogut. Roberta, moltes gràcies també per ser una bona companya i intentar sempre donar un cop de mà. També m'emporto



grans records d'altres membres de Biomat, que sempre han aportat tan bon rotllo a les trobades a la Torreta... Alba, Laura, Cris, German, Robert, Joan... gràcies a vosaltres també.

Una part essencial de la meva vida durant aquests anys han estat els amics de sempre, els que ens coneixem de tota la vida, la colla de sempre... Tot i estar lluny en distància i no estar present en el dia a dia de la tesi, heu estat sempre allí per fer-me desconnectar i passar bons moments. I això m'ha donat molta vida en moments difícils de la tesi, encara que no sempre n'heu estat conscients... Sou molts, em doneu molt, i no acabaria un per un... només dir: GRÀCIES.

I també els amics més recents, que han anat apareixent a la meva vida en aquests últims anys i que s'hi ha quedat. Cristina, Maria, Inma, moltes gràcies per ser-hi sempre, cada dia quan arribava a casa, en dies d'estar enfonsat i dies d'estar eufòric. Ja soc casi menorquí ara. David, Ferran, a vosaltres que vau aparèixer en el màster i que també hem passat el camí de doctorat a la vegada, gràcies per donar suport sempre i intentar sempre estar en contacte a pesar de la distància. Vicent, Fani, a pesar de ser lluny, hem intentat estar sempre en contacte i m'heu donat suport, gràcies a vosaltres també. Marc, Vic, des de la Uni que vam començar junts i ara també hem passat pel doctorat alhora cadascú en un lloc diferent, gràcies per ser-hi sempre també.

I de les últimes persones en entrar a la meva vida has estat tu, Zoa, que vas arribar quan ja estava en la fase final de la tesi, que has passat per tot aquest últim tram amb frases de "no puc, he d'avançar tesi" o "buff, demà tinc dia llarg d'experimental", i no has fugit. Has aguantat i passat amb mi moments molt durs, i no puc dir quant t'ho arribo a agrair. Només espero que després de tancar aquesta etapa i d'obrir una nova, ens seguim estimant com fins ara. Gràcies des del fons del meu cor.

I finalment, agrair a la meva família. Als meus pares, Enriqueta i Joan Antoni, no només pel suport que m'han donat des del moment del primer fogós esforç i fins aquests últims mesos del doctorat, sinó també per educar-me i fer-me ser millor persona. A la meva germana Alba, que m'ha seguit i m'ha recolzat amb el que ha pogut i més, espero que sempre seguim cuidant l'un de l'altre com hem fet sempre. Als meus avis, Pepita, Pepe, Cinteta i Enric (qui ja no has pogut viure aquesta etapa del doctorat però sempre et porto al pensament), moltes gràcies per tota l'amor que m'heu donat, a mi i a tota la família, donant sempre tot el que hem necessitat, preguntant com va, com estem, i intentar entendre'm encara que us requerís esforç. També a la meva padrina Esther, que sempre m'ha recolzat, entès i ajudat a fer-me entendre als altres. I a la resta de família, padrí, tiets, cosins... moltes gràcies a tots per ser-hi.

I amb això acabo aquestes últimes línies i poso punt i final a la tesi. Ja només em queda desitjar que la tesi us sigui lleu de llegir, que us agradi i, sobretot, que us feu una idea de tot el que he estat fent durant aquests últims anys i al que tant temps i esforç i he dedicat.

"Encara queden moltes incògnites que resoldre, però...

... intentarem posar llum a la foscor...

... comencem."



INDEX

INDEX	i
Summary	v
Resum	vii
Resumen	ix
List of publications and communications	xi
Publications	xi
Oral communications	xi
Posters	xi
List of figures	xiii
List of tables	xvii
List of abbreviations	xix
<i>Introduction</i>	1
1.1 Glycolipids	3
1.1.1 Glycosphingolipids	5
1.1.2 Production of glycolipids	8
1.2 Metabolic engineering	11
1.2.1 Definition and evolution	11
1.2.2 Examples of metabolites obtained by metabolic engineering.....	13
1.3 Yeast as host for metabolic engineering	13
1.3.1 Organism description.....	14
1.3.2 Metabolism.....	16
1.3.3 Glycosphingolipids in yeast cell biology.....	22
1.3.4 Metabolic engineering of yeast <i>S. cerevisiae</i>	22
1.4 Framework of the project	23
<i>Objectives</i>	27
<i>Results</i>	29

Chapter 1 Design of a metabolic engineering strategy to produce α-galactosylceramide	31
1.1 Optimizing host metabolism	33
1.1.1 Ceramide and glycosphingolipid metabolic pathways: a bibliography study	33
1.1.2 Strain design	39
1.1.3 Engineering approach	41
1.2 Summary	42
Chapter 2 Engineering a <i>S. cerevisiae</i> strain metabolism to optimize phytoceramide production	45
2.1 Redirection of metabolic flux on sphingolipid pathway	47
2.1.1 Design and construction of <i>SUR2</i> and <i>ISC1</i> yeast expression vectors	47
2.1.2 Analysis of pTDH3-SUR2 and pTEF2-ISC1 strains	57
2.1.3 <i>SCS7</i> gene knock-out	60
2.2 Summary	69
2.3 Annexes Chapter 2	70
Chapter 3 Identification and characterization of an α-galactosylceramide synthase	77
3.1 Identification of an α-galactosylceramide synthase from <i>Bacteroides fragilis</i>...	79
3.1.1 Literature study	79
3.1.2 <i>In silico</i> methods for α -GalCer synthase identification	80
3.1.3 Analysis of candidate sequences expressed with pET28a vector in <i>E. coli</i>	86
3.1.4 Expression analysis of Bf3149 in pET22b-Strep-SUMO vector	93
3.1.5 Comparison of activity between vectors with BF3149 and purified fraction	96
3.2 Characterization of Bf3149	98
3.2.1 Specific activity	98
3.2.2 Enzyme stability	101
3.2.3 Metal binding.....	101
3.2.4 Optimal pH and temperature	103
3.2.5 Enzyme kinetics and substrate affinity	104
3.2.6 Ceramide in mixed micelles and activator lipids	105
3.2.7 Structure analysis	107
3.3 Summary	110

3.4	Annexes Chapter 3	112
	Chapter 4 Engineering a <i>S. cerevisiae</i> strain to produce α-galactosylceramide	123
4.1	Expression of BF9343_3149 in <i>S. cerevisiae</i>	125
4.1.1	Expression vectors for BF3149 in <i>S. cerevisiae</i> for different protein localization	125
4.1.2	Expression vectors p2a33-BF3149	143
4.1.3	Analysis of engineered yeast strains with p2a33 vectors	150
4.2	Summary.....	159
4.3	Annexes Chapter 4	160
	Final discussion	165
	Conclusions	171
	Material and methods	177
	Basic biochemistry and molecular biology protocols and materials.....	179
	In silico analysis.....	180
	Cloning, protein expression and purification in <i>E. coli</i>.....	181
	Cloning of <i>B. fragilis</i> GT sequences into <i>E. coli</i> expression vector.....	181
	Expression of GT candidates and α -GalCer_GT in <i>E. coli</i>	181
	Protein purification.....	182
	Galactosylceramide transferase activity assay	182
	Modeled 3D structure and ligand docking.	184
	Yeast adapted RNA extraction protocol for qPCR analysis	184
	Yeast Bf3149 expression analysis.....	184
	Bibliography	187
	Publications	201

Index



Summary

Glycolipids are molecules of high-added value due to their amphipathic properties, which involve them in several important cellular processes, as well as in many applications such as biosurfactants or vaccine adjuvants. Glycosphingolipids (GSL) is one of the glycolipids families which are composed of a ceramide lipid bound to a sugar moiety. Among GSL, α -galactosylceramide (α -GalCer) has shown immunostimulatory properties which is interesting for biomedical applications. Production of α -GalCer is difficult for the complex chemical synthesis to obtain the specific regio- and stereospecificity.

In this thesis we report an alternative approach using metabolic engineering to produce α -galactosylceramide in the yeast *Saccharomyces cerevisiae*. Metabolic engineering is based on the use of a biological platform that provides the precursors -UDP-galactose and phytoceramide in this case- for the synthesis of the molecule. In addition, a glycosyltransferase capable of transferring galactose to the ceramide acceptor should be present in the host. However, no such enzyme was identified previously to the start of this project. For the implementation of our approach, four steps were necessary: (1) redesign of the host metabolism to forward lipid production to phytoceramide pool, (2) perform modifications on the host lipid metabolism, (3) identification of an α -GalCer glycosyltransferase, and (4) expression of the identified enzyme in the engineered host considering the different yeast compartments.

In the first step, three genes were identified to be key elements to increase phytoceramide pool in yeast, and three gene modifications were planned: *SUR2* and *ISC1* overexpression, and *SCS7* deletion. Secondly, *S. cerevisiae* RH6082 strain was engineered with these modifications. *SUR2* and *ISC1* genes were cloned into two different yeast expression vectors. Expression was analyzed by qPCR, however, only *SUR2* proved to be overexpressed in the transformed strain. In addition, CRISPR/Cas9 technology was used for *SCS7* gene knockout. The plasmids containing *SCS7* gRNA for Cas9 targeting and *SCS7* donor DNA plasmid were successfully constructed, however, no *SCS7* knockout resulted after yeast transformation.

Thirdly, four potential α -GalCer glycosyltransferases were identified from *Bacteroides fragilis* among a pool of GT4 glycosyltransferases using *in silico* methods. Galactosyltransferase activity assay determined that only BF9343_3149 candidate formed galactosylceramide. To improve solubility, the enzyme was expressed as a SUMO fusion protein, with Strep tag for purification. The enzyme proved to be a non-processive glycosyltransferase that prefers UDP-Gal over UDP-Glc as donor substrate and showed no metal dependence. The reaction occurs when ceramide is solubilized with BSA but it is inhibited when ceramide is presented in DOPG/DOPC micelles.

Finally, glycosyltransferase BF9343_3149 was cloned into a yeast expression vector using different relocation strategies. However, engineered yeast strains did not show any GalCer glycosyltransferase activity, and confocal microscopy showed no co-location of enzyme and substrate.

Summary





Resum

Els glicolípidis són molècules amb un alt valor afegit degut a les seves propietats amfipàtiques, fet que les involucra en diferents processos cel·lulars importants, així com moltes aplicacions com bio-surfactant o adjuvants en vacunes. Els glicoesfingolípidis (GSL) són un grup de glicolípidis que es componen per una ceramida (lípid) unida a un glúcid. Entre els GSL, la α -galactosilceramida ha mostrat propietats immunoestimulants, fet que la fa interessant per aplicacions biomèdiques. La producció de α -galactosilceramida és difícil per la complexa síntesi química per obtenir regio- i estèreo-especificitat. En aquesta tesi presentem una alternativa per a la producció de α -galactosilceramida utilitzant enginyeria metabòlica en el llevat *Saccharomyces cerevisiae*. L'enginyeria metabòlica es basa en l'ús d'una plataforma biològica que proveeix els precursors (UDP-galactosa i fitoceramida en aquest cas) per a la síntesi de la molècula. A més, una glicosiltransferasa capaç de transferir la galactosa a l'acceptor ceramida hauria de ser present en aquest hoste. Tot i això, aquest enzim no s'havia identificat abans de l'inici d'aquest projecte. Per a la implementació de la nostra proposta són necessàries quatre etapes: (1) re-disseny del metabolisme de l'hoste per a la re-direcció de la producció lipídica cap a fitoceramida, (2) implementar les modificacions en el metabolisme lipídic de l'hoste, (3) identificar una α -galactosilceramida glicosiltransferasa, i (4) expressar aquest enzim identificat en l'hoste modificat considerant els diferents compartiments cel·lulars del llevat.

En la primera etapa, tres gens van ser identificats com a element clau per incrementar el contingut de fitoceramida en el llevat, i per tant, es van proposar 3 modificacions: sobre-expressió de *SUR2* i *ISC1*, i deleció de *SCS7*.

Segon, la soca de *S. cerevisiae* RH6082 va ser modificada amb les 3 modificacions. Els gens *SUR2* i *ISC1* van ser clonats en dos vectors d'expressió en llevat diferents. La seva expressió va ser analitzada per qPCR, tot i així, només *SUR2* va demostrar ser sobre-expressat en la soca transformada. A més, la deleció de *SCS7* es va fer amb la tecnologia CRISPR/Cas9. Els plasmidis que contenien el gARN de *SCS7* per a la re-direcció de Cas9 i el plasmidi amb el ADN donador de *SCS7* van ser construïts amb èxit, però la deleció del gen no es va donar després d'analitzar les colònies de llevat transformades.

Tercer, quatre potencials α -GalCer glicosiltransferases van ser identificades de *Bacteroides fragilis* entre un grup de glicosiltransferases GT4 mitjançant mètodes *in silico*. L'assaig d'activitat glicosiltransferasa va determinar que només el candidat BF9343_3149 produïa galactosilceramida. Per incrementar la solubilitat de l'enzim, es va expressar fusionat amb la proteïna SUMO, juntament amb un Strep-tag per purificar-la. L'enzim va demostrar ser una glicosiltransferasa no processiva, amb preferència per UDP-Gal abans que UDP-Glc com a substrat donador, i que no mostra dependència de cations metàl·lics. La reacció es dona amb la ceramida solubilitzada amb BSA, però és inhibida quan aquest substrat es presenta en micel·les de DOPG/DOPC.

Finalment, la glicosiltransferasa BF9343_3149 va ser clonada en un vector d'expressió en llevat utilitzant diferents estratègies de re-localització de proteïnes. Tot i això, cap de les soques modificades

Resum



va presentar activitat glicosiltransferasa, i la microscopia confocal va mostrar que no hi ha colocalització d'enzim i substrat.

Resumen

Los glicolípidos son moléculas con un alto valor añadido debido a sus propiedades anfipáticas, hecho que los implica en diferentes procesos celulares importantes, así como muchas aplicaciones como biosurfactantes o adyuvantes en vacunas. Los glicosfingolípidos (GSL) son un grupo de glicolípidos que se componen por una ceramida (lípidos) unida a un glúcido. Entre los GSL, la α -galactosilceramida ha mostrado propiedades inmunoestimulantes, lo que la hace interesante para aplicaciones biomédicas. La producción de α -galactosilceramida es difícil por la compleja síntesis química para obtener regio- y estereo-especificidad.

En esta tesis presentamos una alternativa para la producción de α -galactosilceramida usando ingeniería metabólica en la levadura *Saccharomyces cerevisiae*. La ingeniería metabólica se basa en el uso de una plataforma biológica que provee los precursores (UDP-galactosa y fitoceramida en este caso) para la síntesis de la molécula. Además, una glicosiltransferasa capaz de transferir la galactosa al aceptor ceramida debería estar presente en este huésped. Asimismo, esta enzima no había sido identificada antes del inicio de este proyecto. Para la implementación de nuestra propuesta son necesarias cuatro etapas: (1) rediseño del metabolismo del huésped para la redirección de la producción lipídica hacia la fitoceramida, (2) implementar las modificaciones en el metabolismo lipídico del huésped, (3) identificar una α -galactosilceramida glicosiltransferasa, i (4) expresar esta enzima identificada en el huésped modificado considerando los diferentes compartimentos celulares de la levadura.

En la primera etapa, tres genes fueron identificados como elementos clave para incrementar el contenido de fitoceramida en la levadura, i por tanto, se propusieron 3 modificaciones: sobreexpresión de *SUR2* y *ISC1*, i delección de *SCS7*.

Segundo, la cepa de *S. cerevisiae* RH6082 fue modificada con las 3 modificaciones. Los genes *SUR2* y *ISC1* se clonaron en dos vectores de expresión en levadura diferentes. Su expresión fue analizada por qPCR, aunque solamente *SUR2* demostró ser sobreexpresado en la cepa transformada. Además, la delección de *SCS7* fue implementada mediante la tecnología CRISPR/Cas9. Los plásmidos que contenían el gen de *SCS7* para la redirección de Cas9 y el plásmido con el ADN donador de *SCS7* fueron contruidos con éxito, sin embargo, la delección del gen no se observó después de analizar las colonias de levadura transformadas.

Tercero, cuatro potenciales α -GalCer glicosiltransferasas fueron identificadas de *Bacteroides fragilis* entre un grupo de glicosiltransferasas GT4 mediante métodos *in silico*. El ensayo de actividad glicosiltransferasa determinó que solamente el candidato BF9343_3149 producía galactosilceramida. Para incrementar la solubilidad de la enzima, se expresó fusionándola con la proteína SUMO, junto con un Strep-tag para purificarla. La enzima demostró ser una glicosiltransferasa no procesiva, con preferencia para UDP-Gal antes que UDP-Glc como sustrato donador, y que no muestra dependencia

de cationes metálicos. La reacción se da con la ceramida solubilizada con BSA, pero es inhibida cuando este sustrato se presenta en micelas de DOPG/DOPC.

Finalmente, la glicosiltransferasa BF9343_3149 se clonó en un vector de expresión en levadura usando diferentes estrategias de relocalización de proteínas. Aun así, ninguna de las cepas modificadas presentó actividad glicosiltransferasa, y la microscopia confocal mostró que no hay co-localización de enzima y sustrato.

List of publications and communications

Publications

1. Caballé, M., Faijes, M., & Planas, A. (2022). Characterization of a Glycolipid Synthase Producing α -Galactosylceramide in *Bacteroides fragilis*. *International journal of molecular sciences*, 23(22), 13975. <https://doi.org/10.3390/ijms232213975>

Oral communications

1. "A novel approach for the production of glycosphingolipids in *Saccharomyces cerevisiae*". Introductory Workshop on Biomedical Glycoscience, 3-5 June 2019, San Sebastián (Spain).
2. "Enginyeria metabòlica per a la producció de glicoesfingolípid en *Saccharomyces cerevisiae*". I Jornada Doctorands IQS, 16-17 May 2019, Barcelona (Spain).

Posters

1. "Tuning phosphatidic and UDP-glucose metabolic pathways for glycolipid production". Carbohydrate Bioengineering Meeting 13, 19-22 May 2019, Toulouse, (France).
2. "A novel approach for the production of glycosphingolipids in *Saccharomyces cerevisiae*". Introductory Workshop on Biomedical Glycoscience, 3-5 June 2019, San Sebastián (Spain).
3. "Enginyeria metabòlica per a la producció de glicoesfingolípid en *Saccharomyces cerevisiae*". I Jornada Doctorands IQS, 16-17 May 2019, Barcelona (Spain).


List of publications and communications





List of figures

- Figure 0.1** Example of glycolipid: chemical structure of β -galactosyldiacylglycerol. Page 3
- Figure 0.2** Chemical structure of different glycolipids. Page 4
- Figure 0.3** Chemical structure of KRN7000, α -galactosylceramide. Page 6
- Figure 0.4** CD1d/ α -galactosylceramide complex and its immune system activation scheme. Page 7
- Figure 0.5** Enzymatic reaction of β -glucocerebroside synthesis by UGCG glycosyltransferase. Page 8
- Figure 0.6** Glycosyltransferases catalytic mechanism of action. Page 9
- Figure 0.7** The five different levels of metabolic engineering as identified by Erb *et al.* Page 12
- Figure 0.8** α -galactosylceramide synthesis reaction catalyzed by BF9343_3149 glycosyltransferase from *B. fragilis*. Page 14
- Figure 0.9** *S. cerevisiae* cellular organization with organelles. Page 15
- Figure 0.10** *S. cerevisiae* metabolism map. Page 17
- Figure 0.11** Nucleotide-sugar synthesis pathway in *S. cerevisiae*. Page 18
- Figure 0.12** Ceramide synthesis pathway in *S. cerevisiae*. Page 19
- Figure 0.13** Yeast ceramide structures. Page 20
- Figure 0.14** Complex sphingolipids (GSL) synthesis pathway in *S. cerevisiae*. Page 21
- Figure 1.1** Sphingosine and ceramide species in yeast *S. cerevisiae* and *Homo sapiens*. Page 34
- Figure 1.2** Metabolic pathway map of modifications from bibliography. Page 35
- Figure 1.3** Scheme of metabolic flux redirection approach on new *S. cerevisiae* strain. Page 40
- Figure 1.4** Scheme of metabolic engineering strategy on creating a new *S. cerevisiae* strain for production of α -galactosylceramide. Page 43
- Figure 2.1** Vector map of pTDH3-ScoMG517 (A), pTDH3-SUR2 (B) and pTEF2-ISC1 (C). Page 49
- Figure 2.2** CPEC strategy followed for *SUR2* (Top) and *ISC1* (Bottom) cloning. Page 51
- Figure 2.3** Agarose gel electrophoresis results of PCR for pTDH3-SUR2 and pTEF2-ISC1 vector synthesis. Page 53
- Figure 2.4** Agarose gel electrophoresis with different amplifications. Page 53
- Figure 2.5** Digestion analysis approach of pTDH3-SUR2 and pTEF1-ISC1 vectors. Page 54
- Figure 2.6** Agarose gel electrophoresis results of Colony PCR for pTDH3-SUR2. Page 55
- Figure 2.7** Agarose gel electrophoresis results of restriction analysis of positive colonies of transformation with pTDH3-SUR2. Page 56
- Figure 2.8** Agarose gel electrophoresis results of restriction analysis of transformed colonies with pTEF2-ISC1. Page 56
- Figure 2.9** qPCR scheme for *SUR2* and *ISC1* expression analysis. Page 58
- Figure 2.10** Graph bar of the normalized qPCR results of *SUR2* expression. Page 59
- Figure 2.11** Vector map of p414-TEF-Cas9 plasmid. Vector elements are marked in colored arrows in the center, the name of the vector and its size. Page 60
- Figure 2.12** CPEC strategy followed for *SCS7* gRNA plasmid construction. Page 61
- Figure 2.13** CPEC strategy followed for donor DNA plasmid construction. Page 62
- Figure 2.14** Vector map of *SCS7* donor DNA plasmid (pET22b-donorDNA-SCS7) (A) and *SCS7* gRNA plasmid (B). Page 64

- 
- Figure 2.15** Agarose gel electrophoresis results of PCR amplification of SCS7 gRNA and donor DNA plasmids. Page 66
- Figure 2.16** Agarose gel electrophoresis results of Colony PCR for SCS7 donor DNA plasmid. Page 67
- Figure 2.17** Colony PCR analysis of SCS7 knockout colonies. Page 68
- Figure 3.1** Neighbor-Joining phylogenetic tree of annotated GT4 enzymes. Page 81
- Figure 3.2** BLAST search of candidate sequences. Page 85
- Figure 3.3** Scheme for ordered GT sequences with additional bp that include restriction sites. Page 86
- Figure 3.4** Strategy followed for GT cloning into pET28a vector. Page 87
- Figure 3.5** Vector maps for the final assembled vectors for *B. fragilis* candidate GT expression. Page 88
- Figure 3.6** Agarose gel electrophoresis results of PCR for pET28a-GT vector synthesis. Page 89
- Figure 3.7** Colony PCR analysis of pET28a-GT vectors. Page 90
- Figure 3.8** SDS-PAGE gel electrophoresis results of expression of *B. fragilis* GT candidates in *E. coli*. Page 91
- Figure 3.9** Time-course reaction monitoring of galactosyltransferase activity of a cell-free extract expressing all GT candidates. Page 92
- Figure 3.10** Time-course reaction monitoring of galactosyltransferase activity of a cell-free extract expressing BF9343_3149. Page 92
- Figure 3.11** Vector map of pET22b-Strep-SUMO-BF3149. Page 94
- Figure 3.12** CPEC strategy followed for pET22b-Strep-SUMO-BF3149 construction. Page 94
- Figure 3.13** Agarose gel electrophoresis results of PCR for pET22b-Strep-SUMO-BF3149 construction. Page 95
- Figure 3.14** Ceramide-NBD activity assay of α -GalCer_GT (BF9343_3149) cell-free extracts from different expression systems. Page 96
- Figure 3.15** SDS-PAGE gel electrophoresis results of purification of construct Strep-SUMO-BF3149. Page 97
- Figure 3.16** Time-course reaction monitoring of galactosyltransferase activity of purified α -GalCer synthase. Page 99
- Figure 3.17** A) Time-course of BF3149 α -GalCer GT reaction. B) Linear dependence of initial rates with enzyme concentration. Page 100
- Figure 3.18** Stability of α -GalCer synthase from *B. fragilis*. Page 101
- Figure 3.19** Effect of metal cations on α -GalCer synthase activity. Page 102
- Figure 3.20** Optimal pH and temperature profiles of α -GalCer synthase activity. Page 103
- Figure 3.21** Kinetics of α -GalCer synthase. Page 104
- Figure 3.22** α -GalCer synthase activity on UDP-Gal and UDP-Glc donor substrates. Page 105
- Figure 3.23** Reaction monitoring of α -GalCer GT activity using ceramide in mixed micelles as acceptor. Page 106
- Figure 3.24** Structural alignment of 10 selected GT4 enzymes. Page 108
- Figure 3.25** UDP-galactose donor location in modeled α -GalCer GT with identified interacting residues. Page 109
- Figure 3.26** Structural superposition of the modeled α -GalCer GT with bound UDP-Gal and PimA from *Mycobacterium smegmatis* with bound GDP-Man. Page 109



- Figure S3.1** Figure S3.1 Multiple sequence alignment of annotated GT4 enzymes from *Bacteroides fragilis* and query sequences of monoglucosyldiacylglycerol synthases from *Streptococcus pneumoniae* and *Acholeplasma laidlawii* . Page 114
- Figure S3.2** Sequence alignment of GT4 enzymes. Page 119
- Figure 4.1** CPEC strategy followed for construction of pSF-BF3149 vectors. Page 127
- Figure 4.2** Vector map of pSF-TPI1-URA3 (A), pSF-Bf3149 (B), pSF- α MF-Bf3149-HDEL (C) and pSF-Kre2-Bf3149 (D). Page 129
- Figure 4.3** Vector map of pSF-BF3149-mKate2 (A), pSF- α MF-BF3149-mKate2-HDEL (B) and pSF-KRE2-BF3149-mKate2 (C). Page 130
- Figure 4.4** Agarose gel electrophoresis results of PCR amplification of pSF-BF3149 plasmids. Page 132
- Figure 4.5** Digestion analysis of pSF-BF3149 vectors. Page 133
- Figure 4.6** Agarose gel electrophoresis of restriction analysis of transformed colonies with pSF-Bf3149 plasmids. Page 134
- Figure 4.7** CPEC strategy followed for construction of pSF-BF3149-mKate2 vectors. Page 135
- Figure 4.8** Agarose gel electrophoresis results of PCR amplification of pSF-BF3149-mKate2 plasmids. Page 137
- Figure 4.9** Colony PCR analysis of pSF-BF3149-mKate2 vectors. Page 137
- Figure 4.10** Agarose gel electrophoresis results of restriction analysis of transformed colonies with pSF-Bf3149 -mKate2 plasmids. Page 138
- Figure 4.11** Agarose gel electrophoresis from PCR amplification of pSF-mKate2 plasmids. Page 140
- Figure 4.12** Agarose gel electrophoresis results of restriction analysis of transformed colonies with pSF-mKate2 plasmids. Page 141
- Figure 4.13** CPEC strategy followed for construction of p2a33-BF3149 vectors. Page 145
- Figure 4.14** Vector map of p2a33-mCherry (A), p2a33-Bf3149 (B), p2a33-Bf3149-mCherry (C), p2a33-Kre2-Bf3149 (4) and p2a33-Bf3149--mCherry (E). Page 146
- Figure 4.15** Agarose gel electrophoresis results of PCR amplification of p2a33-Bf3149 plasmids. Page 149
- Figure 4.16** Reaction evolution of Bf3149 from *E. coli* cultures. Page 152
- Figure 4.17** Confocal fluorescence microscopy of RH6082 + p2a33-mCherry strain. Page 155
- Figure 4.18** Confocal fluorescence microscopy of RH6082 + p2a33-BF3149-mCherry strain. Page 156
- Figure 4.19** Confocal fluorescence microscopy of RH6082 + p2a33-Kre2-BF3149-mCherry strain. Page 157
- Figure S4.1** Vector map of pSF-mKate2 (A), pSF- α MF-mKate2-HDEL (B) and pSF-KRE2-mKate2 (C). Page 163

List of figures





List of tables

Table 1.1 List of all the modification on sphingolipid -or fatty acid elongation- metabolism of *S. cerevisiae* cited in this section. Page 38

Table 1.2 List of metabolic engineering modifications for the new α -galactosylceramide production strain of *S. cerevisiae*. Page 41

Table 2.1 Table of primers and templates used to amplify each fragment for pTDH3-SUR2 plasmid. Page 50

Table 2.2 List of primers and templates used to amplify each fragment for pTEF2-ISC1 plasmid. Page 52

Table 2.3 Table of primers used in Colony PCR with the template (colonies) and the expected amplified fragment sizes. Page 54

Table 2.4 restriction enzymes and expected fragments after restriction analysis for pTDH3-SUR2 and pTEF2-ISC1 plasmids. Page 55

Table 2.5 List of primer design used for SCS7 knockout vectors. Page 63

Table 2.6 List of primers and templates used to amplify each fragment for pSF-BF3149 plasmids. Page 66

Table S2.1 List of primers used in Chapter 2 and their sequences. Page 70

Table S2.2 SUR2 gene and protein sequence. Page 71

Table S2.3 ISC1 gene and protein sequence. Page 71

Table S2.4 SCS7 gene and protein sequence, including 1 kb regions at upstream of start codon and downstream of stop codon. Page 72

Table S2.5 TEF2 gene sequence, including 1 kb regions at upstream of start codon and downstream of stop codon. Page 73

Table S2.6 PGK1 gene sequence, including 1 kb regions at upstream of start codon and downstream of stop codon. Page 75

Table S2.7 mCherry gene and protein sequence. Page 76

Table 3.1 List of primers and templates used to amplify each fragment for pET28a vectors for each *B. fragilis* candidate GT. Page 89

Table 3.2 List of primers and templates used to amplify each fragment for pET22b-Strep-SUMO-BF3149. Page 95

Table S3.1 List of primers used in Chapter 3 and their sequences. Page 112

Table S3.2 GT4 enzymes with solved crystal structure (October 2022). Page 115

Table S3.3 BF9343_3149 gene and protein sequence. Page 119

Table S3.4 BF9343_1306 gene and protein sequence. Page 119

Table S3.5 BF9343_3589 gene and protein sequence. Page 120

Table S3.6 BF9343_0585 gene and protein sequence. Page 120

Table S3.7 SUMO gene and protein sequence. Page 121

Table 4.1 List of primer design used for pSF-TPI1-URA3 plasmid construction. Page 128

Table 4.2 List of primers and templates used to amplify each fragment for pSF-BF3149 plasmids. Page 131

Table 4.3 Table of restriction enzymes and expected fragments for restriction analysis for plasmids pSF-TPI1-URA3 with only Bf3149 sequence. Page 133

List of tables



- Table 4.4** List of primers and templates used to amplify each fragment for pSF-Bf3149-mKate2 plasmids. Page 135
- Table 4.5** Table of primers used in Colony PCR for pSF-BfGT-mKate2 plasmids. Page 136
- Table 4.6** List of primers and templates used to amplify each fragment for pSF-mKate2 plasmids. Page 139
- Table 4.7** Restriction enzymes and expected fragments for plasmids pSF-mKate2. Page 140
- Table 4.8** List of primer design used for p2a33-BF3149 plasmid construction. Page 144
- Table 4.9** List of primers and templates used to amplify each fragment for p2a33-BF3149-mCherry plasmids. Page 148
- Table 4.10** List of strains used for cell-free extract ceramide-NBD activity assay. Page 150
- Table S4.1** List of primers used in Chapter 4 and their sequences. Page 160
- Table S4.2** mKate2 gene and protein sequence. Page 161
- Table S4.3** Truncated *KRE2* DNA (300 bp) and protein sequence (100 residues). Page 162
- Table S4.4** Alpha Mating Factor sequence (183 bp) and protein sequence (61 residues). Page 162
- Table S4.5** Linker DNA and protein sequence. Page 162



List of abbreviations

Abbreviation	Definition
BF3149	(Gene) α -GalactosylCeramide synthase/glycosyltransferase from <i>B. fragilis</i> gene BF9343_3149
Bf3149	(Protein/Enzyme) α -GalactosylCeramide synthase/glycosyltransferase from <i>B. fragilis</i> gene BF9343_3149
α -GalCer synthase	(Protein/Enzyme) α -GalactosylCeramide synthase/glycosyltransferase from <i>B. fragilis</i> gene BF9343_3149
α -GalCer GT	(Protein/Enzyme) α -GalactosylCeramide synthase/glycosyltransferase from <i>B. fragilis</i> gene BF9343_3149
GalCer	GalactosylCeramide
α -GalCer	α -GalactosylCeramide
GL	Glycolipid
GT	Glycosyltransferase
GSL	Glycosphingolipid
DHS	Dihydrosphingosine
DHC	Dihydroceramide
PHS	Phytosphingosine
PHC	Phytoceramide
Leu	Leucine
Ura	Uracil
Trp	Tryptophan
CEN	Centromeric sequence
ARS	ARS low copy number origin of replication
2 μ	2 micron origin of replication
YEp	Yeast Episomal plasmid
YCp	Yeast Centromeric plasmid
pb	Base pairs
kb	Kilo-base pairs
PCR	Polymerase Chain Reaction
CPEC	Circular Polymerase Extension Cloning
ISC1	(Gene) phosphoinositide phospholipase C (EC:3.1.4.12)
SUR2	(Gene) sphingosine hydroxylase (EC:1.14.18.5)
SCS7	(Gene) 4-hydroxysphinganine ceramide fatty acyl 2-hydroxylase (EC:1.14.18.6)
pTHD3	THD3 promoter
tADH1	ADH1 terminator

Abbreviations

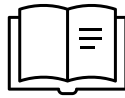


pTEF2	TEF2 promoter
tPGK1	PGK1 terminator
<i>ScoMG517</i>	(Gene) MG517 glycosyltransferase from <i>M. genitalium</i> , sequence codon-optimized for <i>S. cerevisiae</i>
<i>LEU2</i>	(Gene) Leucine auxotrophy selection marker,
<i>URA3</i>	(Gene)Uracil auxotrophy selection marker,
<i>KRE2</i>	(Gene) KRE2 from <i>S. cerevisiae</i> , alpha 1,2-mannosyltransferase (EC:2.4.1.-)
Kre2	(Protein) KRE2 from <i>S. cerevisiae</i> , alpha 1,2-mannosyltransferase (EC:2.4.1.-)
ER	Endoplasmic Reticulum
GA	Golgi Apparatus
ACN	Acetonitrile

Introduction



Introduction



Introduction



1.1 Glycolipids

Glycolipids (GL) are biological compounds with a wide range of diversity in their structures. They are produced by microorganisms, animals, and plants, although they vary greatly among species. Glycolipids are formed by two moieties (Figure 0.1): a sugar and a lipid. Due to the different polarity of each unit, glycolipids are amphipathic molecules. This feature confers GL the tendency to arrange in supramolecular structures^{1,2}.

Each of these two moieties may present a wide variety of structures. Lipids may vary in composition: fatty acids, fatty alcohols, fatty amino alcohols, sterols, carotenoids, etc. Each type can vary in chain length, desaturation, branching, etc. On the other hand, glycosides may comprise different types of structures -glucose, mannose, glucuronic acid, etc.- and may vary in their chain number, as they can be present as mono-, di-, oligo- or polysaccharides.

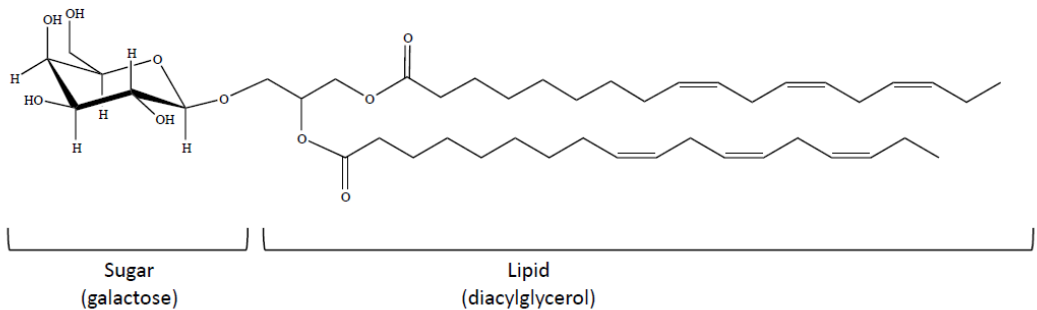


Figure 0.1 Example of glycolipid: chemical structure of β -galactosyldiacylglycerol. The galactosyl unit and the diacylglycerol are linked by an β -O-glycosidic bond.

Although many classifications for glycolipids have been proposed, the simpler categorization is the separation of glycolipids into two main groups³⁻⁶ depending on the way sugar and lipid parts are linked:

- **Simple glycolipids:** They also can be named saccharolipids. Sugar and lipid residues are directly linked to each other (Figure 0.2); therefore, they are composed of two parts. Simple glycolipids include:
 - Glycosylated fatty acids
 - Glycosylated fatty alcohols
 - Glycosylated carotenoids
 - Glycosylated hopanoids
 - Glycosylated sterols
 - Glycosylated paraconic acids

- **Complex glycolipids:** Glycosyl and lipid residues are linked through a linker group; therefore, they are composed of three parts. Complex glycolipids include:
 - Glycoglycerolipids (glycerol as linker group, Figure 0.2)
 - Glycosphingolipids (sphingosine/acylated-sphingosine as linker group, Figure 0.2)
 - Lipopolysaccharides (although there is no linker group, it is classified as complex glycolipids due to the complexity of the polysaccharide residue)
 - Phenolic glycolipids (phenol as linker group)
 - Glycopeptidolipids (peptide as linker group)
 - Nucleoside lipids (nucleoside as linker group)

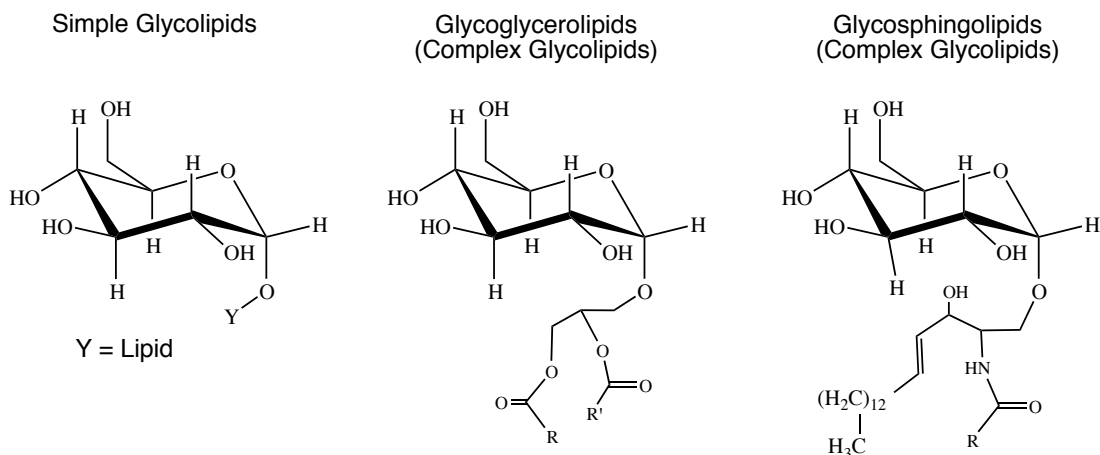


Figure 0.2 Chemical structure of different glycolipids. In simple glycolipids, lipid moiety (Y) is directly bound to the sugar moiety (glucose in this example). Complex glycolipids have linkers that bind the sugar and lipid moieties, these lipids are acyl chains of different structures (R or R').

Owing to the wide range of different structures of glycolipids, they are also implicated in many different biological roles. Due to its amphipathic property, glycolipids are present in cellular membranes, helping to maintain its structural integrity⁷. In addition, GL also partake in other roles that involve membrane anchoring for other biomolecules^{8,9} or cell recognition and signaling¹⁰. GL also participate in photosynthesis, as they have been found near the photosynthetic complexes and may play a role in electron transport¹¹.

Glycolipids, as natural products, are molecules with high biodegradability and biocompatibility¹², as well as low toxicity. Therefore, they can be used in a broad range of application substituting other synthetic components with similar properties. For instance, they can be used as biosurfactants, as they are an interesting candidate to replace non-ionic petroleum-based surfactants^{13,14}. Due to their capability to form supramolecular structures, GL can also be used as platforms for drug-delivery systems^{15–17}. Finally, since GL participate in cell signaling, they can also stimulate the immune system response because they bind to specific complexes of immune cells, therefore, they are potential candidates to be used as immunostimulators^{18,19}.

1.1.1 Glycosphingolipids

Glycosphingolipids (GSL) are a type of complex glycolipids which have sphingosine or acyl-sphingosine as linker group. GSL are a very wide group since they cover a great number of different structures and are present in many organisms. Variations can occur in the lipid, either in the sphingosine moiety or/and acyl chains, and in the sugar moiety. For instance, in mammals, different groups of GSL are present, being ganglio-, lacto- and globo-glycosphingolipids the most important classes. All GSL series start their synthesis by adding glucose to the ceramide acceptor in β configuration, followed by the union of galactose to the glucose with $\beta(1-4)$ bond, yielding lactosylceramide. Then, each line differs in its following residues: ganglio- series adds N-acetylgalactosamine in $\beta(1-4)$ and then another galactose in $\beta(1-3)$ at this point; lacto- series adds N-acetylglucosamine in $\beta(1-3)$ bond and then galactose in $\beta(1-3)$; and globo- series adds galactose in $\alpha(1-4)$ followed by N-acetylgalactosamine in $\beta(1-3)$ bond. Each series is tissue-specific, for example, ganglio-series are predominant in brain, lacto-series in secretory organs and globo-series in erythrocytes²⁰.

Glycosphingolipids play roles in membrane microdomain organization, regulation of membrane surface proteins and signal transduction. This allows them to be involved in different cellular functions such as immune response or cell adhesion²¹. Defective GSL patterns in these tissues can lead to several disorders in humans. For instance, Gaucher disease is caused by an accumulation of glucocerebroside (or glucosylceramide), synthesized by glucosylceramide synthase²² (coded by UGCG gene), in the cells. Glucocerebroside is particularly accumulated in macrophages and other white blood cells, which can cause enlarged spleen and liver, bone lesions or severe neurological disorders. This accumulation is caused by defective β -glucocerebrosidase²³, which hydrolyzes the glycosidic bond of glucocerebroside. Due to its impact on human health, β -glucocerebrosidase has been largely studied since it is a therapeutic target for pharma industry²⁴.

Another relevant function of GSL in humans is related to the immune system. GSL may boost the activation of invariant natural killer T (iNKT) cells. iNKT cells are involved in the suppression of autoimmune reactions, cancer metastasis, and the graft rejection response. Some GSL presented by CD1d (a MHC class I molecule) of dendritic cells can activate a T-cell antigen receptor expressed on the cell surface of iNKT cells to trigger the secretion of several cytokines (such as interleukin IL-2, IL-4 and interferon- γ) as important effectors in the subsequent immune response²⁵.

1.1.1.1 α -Galactosylceramide

iNKT cells can also be activated by nonmammalian α -GalCer, with an α -glycosidic linkage not found in mammals. Indeed, agelasphin, first isolated in 1993 as an antitumor agent in the marine sponge *Agelas mauritianus*, was identified as α -GalCer^{26,27}. Chemical optimization of the ceramide portion yielded a α -GalCer structure with C18-phytosphingosine and cerotic acid (26:0), named KRN700 (Figure 0.3), as

a potential candidate for clinical applications²⁸. More recently, α -GalCer was identified as a lipid antigen for iNKT cells when presented by CD1d^{29,30}. Furthermore, the biological importance of this molecule is in its α link which confers its unique properties³¹.

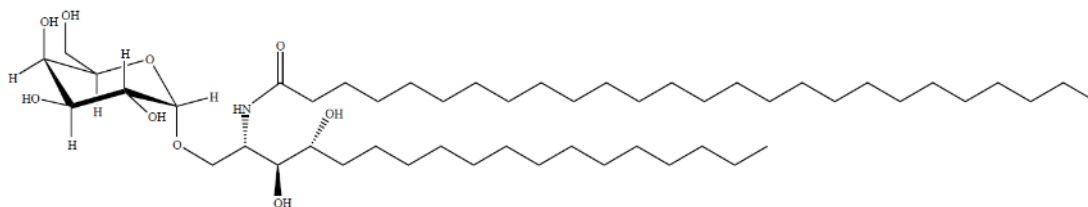


Figure 0.3 Chemical structure of KRN7000, α -galactosylceramide. The structure contains the galactose and the ceramide linked by an α -O-glycosidic bond.

Due to its medical applications, α -galactosylceramide is an interesting metabolite for industry. As an immunostimulating molecule, α -GalCer binds to CD1d receptor (Figure 0.4A), a member of the CD1 family of MHC-like molecules, from an antigen presenting cells -a resting dendritic cell (DC). CD1d- α -GalCer complex then interacts with TCR receptors of NKT cells, thus, initiating a cascade of signaling that boosts immune response (Figure 0.4B). Once NKT cell recognizes the CD1d- α -GalCer complex through TCR receptor, it induces NKT cells to upregulate Th1 and Th2 cytokines and chemokines release, activation of DC -that, in turn, enhances NKT activation-, CD8 killer cells, CD4 helper cells, NK cells and B cells -which will produce antibodies²⁵.

Owing to its immunostimulant attribute, α -GalCer has been broadly studied to be used as vaccine adjuvant^{32,33} and other applications such as in immunotherapy in cancer³⁴⁻³⁶.

α -GalCer has been traditionally obtained by chemical synthesis, however, protection steps are required in the synthesis process to maintain its stereoselectivity resulting in low yields of product. Finding natural sources to obtain these molecules is key for successful industry applications.

In the search of other natural sources, α -GalCer was isolated from *Bacteroides fragilis*, a gram-negative bacterium common of the human colon microbiota. Although the organism is generally commensal, it can become pathogenic causing infections if displaced into the bloodstream or surrounding tissues following trauma, surgery, or disease³⁷. However, the metabolic route of α -GalCer synthesis in the organism was not reported. The bacterial origin of α -GalCer supports the hypothesis that the identified congeners in *Agelas* were actually produced by symbiotic bacteria colonizing the sponge²⁵.

Finally, in a recent report by Okino *et al.*³⁸ the enzyme from *B. fragilis* that can catalyze the reaction of α -GalCer synthesis was identified: BF9343_3149. This publication confirmed our findings during this project that this enzyme was a retaining glycosyltransferase which uses UDP-galactose as sugar donor and ceramide as acceptor.

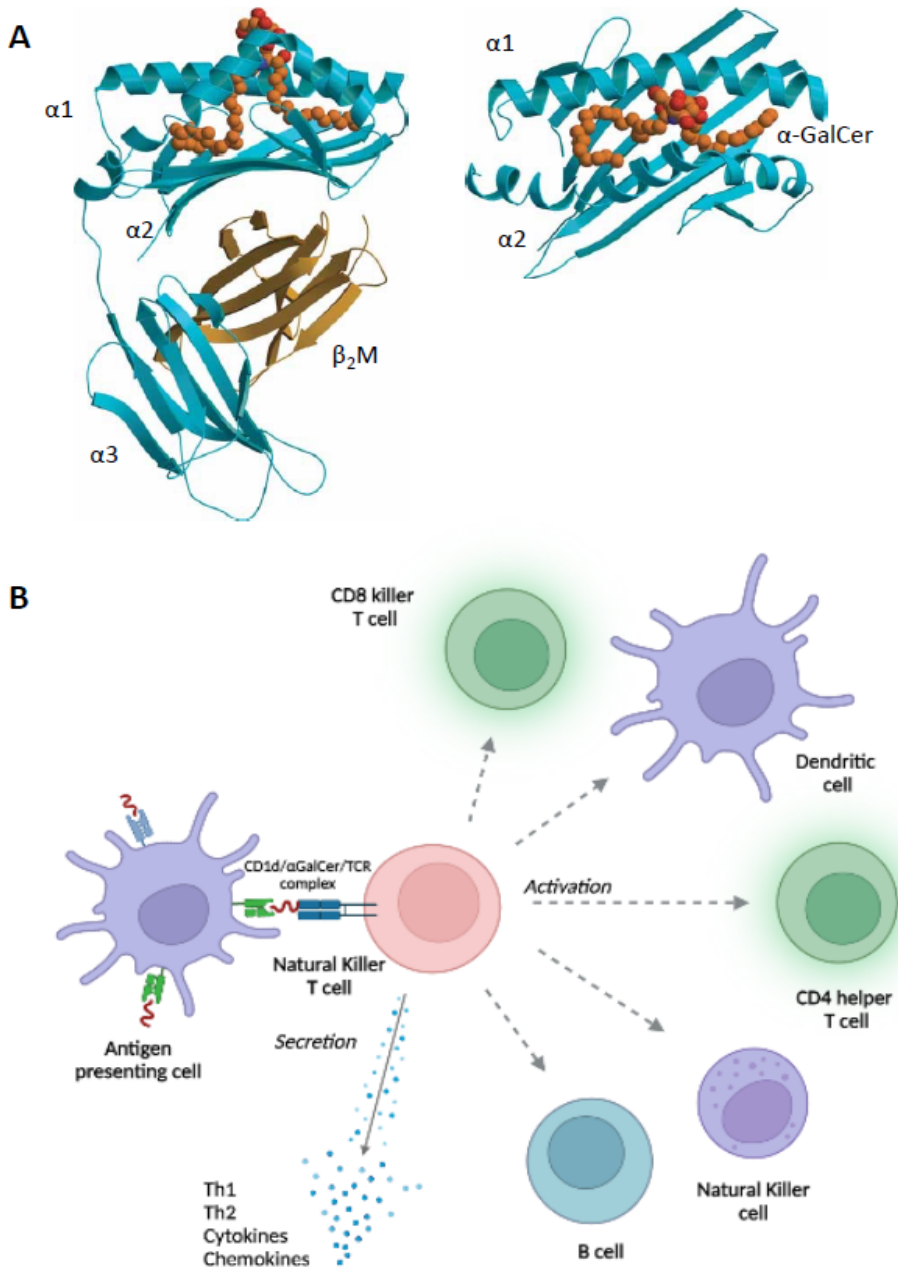


Figure 0.4 CD1d/α-galactosylceramide complex and its immune system activation scheme. A) 3D structure of CD1d/α-galactosylceramide complex, with CD1d α1, α2 and α3 domains (in blue), and β₂M domain (in yellow), represented as van der Waals spheres in front (left) and orthogonal top view (right). Figure extracted from Koch *et al.*¹⁵² B) Scheme of immune system activation through α-galactosylceramide binding to CD1d, from the antigen presenting cell, the activation of NKT cells, and the subsequent reactions of other immune cell activation and molecule secretion. Image created with BioRender.com.

1.1.2 Production of glycolipids

Owing to the wide variety of applications and industrial interest, there is a need for obtaining glycolipids using eco-friendly methods. Since glycolipids are widely present in cellular membranes of bacteria, fungi, plants or animals, they could be extracted. It has been reported that the amount of glycolipid that could be extracted from plants ranges from 8 to 700 mg per 100 g of biomass³⁹. However, extraction from plant presents some problems, being a low yield of extraction and, for animal tissues, it could rise ethical problems regarding the extraction from animal sources.

Another approach to obtain glycolipids could be by chemical synthesis. However, this synthesis can become difficult due to the stereoselectivity of the molecules. To obtain highly pure synthetic glycolipid of a specific type, many steps of protection and deprotection are necessary to maintain stereoselectivity, which lowers the final reaction yield. Moreover, chemical synthesis usually uses high quantities of organic solvents and reagents which opposes our society goals of reducing contaminant solvents and advancing towards a greener chemistry⁴⁰.

1.1.2.1 Enzymatic synthesis by glycosyltransferases

The use of enzymatic synthesis is another potential approach. The enzymes that carry out the transference of sugar moieties from an activated donor molecules (UDP-Gal in Figure 0.5) to a lipid acceptor (ceramide in Figure 0.5), forming a glycosidic bond, are glycosyltransferases (GT).

Enzymatic synthesis overcomes the problem of glycolipid stereoselectivity since enzymes always synthesize the desired stereoisomer with high efficiency. Moreover, enzymes can be used in mild conditions in aqueous solutions, so they avoid the necessity of using high amounts of organic solvents in the reaction.

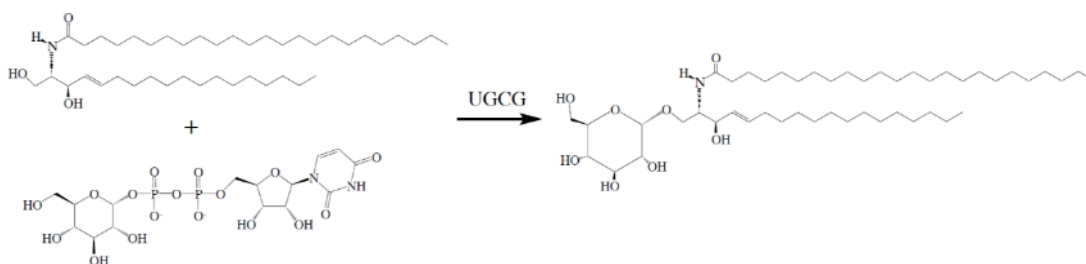


Figure 0.5 Enzymatic reaction of β -glucocerebroside synthesis by UGCG glycosyltransferase. Both substrates (left), UDP-glucose (bottom) and ceramide (top), are used by UGCG glycosyltransferase to synthesize β -glucocerebroside).

Classification of glycosyltransferases

There are several categorizations for GT: besides classification by families based on amino acid sequence similarities, there are two other classifications that split annotated GT into two larger groups depending on mechanism of action or three-dimensional structure.

Glycosyltransferases can be classified as inverting or retaining enzymes (Figure 0.4) based on their mechanism of action, which will determine the stereochemistry of the substrates and products⁴¹.

Inverting GT change the configuration of the anomeric carbon of the substrate after catalysis. Its mechanism of action consists of a single displacement in which the acceptor carries out a nucleophilic attack on the anomeric carbon (C-1) of the sugar donor⁴² (Figure 0.6A). In opposition, retaining GT were thought first to operate in a two-step mechanism that involves a glycosyl-enzyme intermediate, however, now another approach based on internal return S_Ni -like mechanism has been proposed, where the leaving group departure and nucleophilic attack occur in a combined but asynchronous way on the same face of the glycoside⁴³ (Figure 0.6B).

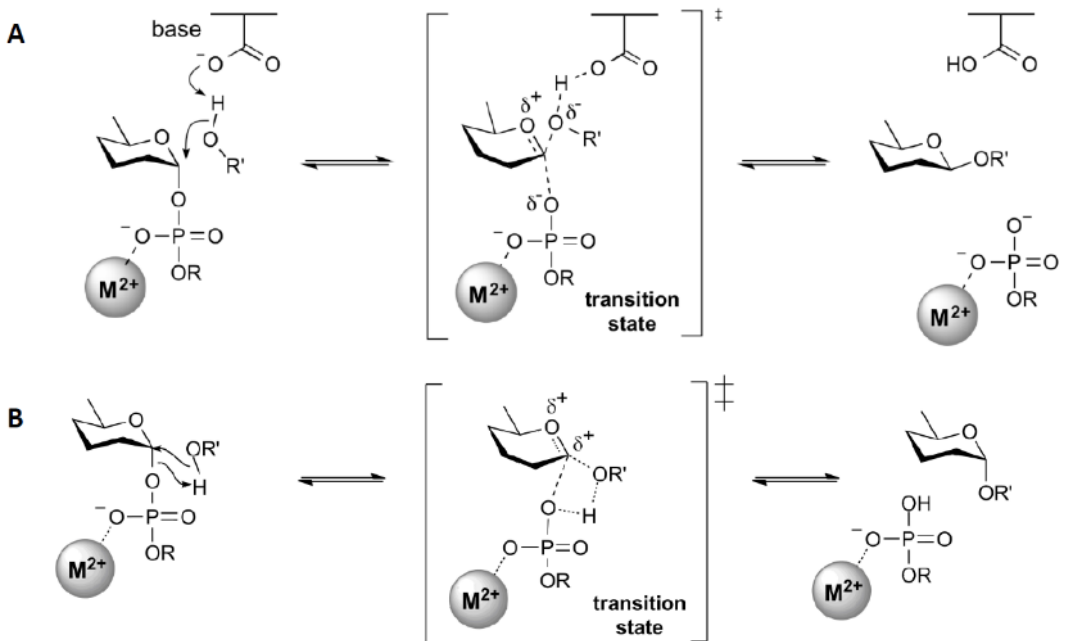


Figure 0.6 Glycosyltransferases catalytic mechanism of action. A) Inverting mechanism of action for a metal-dependent GT. A nucleophilic attack is carried out by the acceptor (H-O-R') to the anomeric carbon of the sugar donor (hexose ring). An intermediate product is formed in the transition state, then the final product is released. B) Retaining S_Ni -like mechanism of action. Nucleophilic attack and departure of leaving group happens at the same time through a unique transition state leading to the final product. Author: Spencer Williams, www.cazypedia.org¹⁵³

GT can also be classified in another two large groups by the 3D structure they present: GT-A and GT-B. GT-A protein topology consists of two closely adjacent $\beta/\alpha/\beta$ Rossmann domains. Due to the close proximity of both domains GT-A are usually considered as a single domain protein where both nucleotide- and acceptor-binding subdomains are present. On the other hand, GT-B topology also includes two separate $\beta/\alpha/\beta$ Rossmann domains that face each other for nucleotide and acceptor binding, although joint by a flexible link⁴².

Limitation of enzymatic synthesis

Enzymatic synthesis still has some limitations. First, although glycosyltransferases have a wide range of different reactions to produce a variety of different glycolipids, there are many unknown glycosyltransferases that synthesize other glycolipids. For example, until very recently, the enzyme that catalyzed the reaction of α -galactosylceramide synthesis was not identified³⁸. Furthermore, other than using known or newly discovered glycosyltransferases, engineered glycosyltransferases can provide a new set of products to be synthesized. Considering the promiscuity of some GT, rational design or directed evolution in protein engineering can use this promiscuity towards a donor or acceptor to shift the specificity for other non-preferred substrates⁴⁴. Protein engineering can also be used to create new GT⁴⁵, for instance, a glycosidase that was redesigned to have a novel mechanism of action, creating a glycosynthase, which synthesizes oligosaccharides⁴⁶. This could be carried out due to the similar mechanism of action between glycosyltransferases and glycosidases, which can also present inverting or retaining mechanism of action. Second, the cost of acquiring the reagents and the enzymes for industrial production may be very expensive, since GT enzymes use activated donors, such as UDP-galactose.

Finally, another approach that could also solve enzymatic synthesis problems is extraction from biological cultures of a metabolic engineered host. Metabolic engineering redirects and creates new metabolic pathways in a specific organism. Thus, allowing to use the host endogenous metabolites to synthesize the substrates of the desired reactions. For glycolipid production, a glycosyltransferase could be heterologous expressed in a host to use its metabolites to synthesize glycolipids, therefore, avoiding the necessity of purchasing expensive ingredients -activated sugars or enzymes- for the catalysis. Furthermore, in metabolic engineering not only the heterologous expression of an enzyme to create a new pathway is considered, but also improving endogenous pathways that will lead to increased pools of substrates and, thus, to yield higher amount of product⁴⁷.



1.2 Metabolic engineering

1.2.1 Definition and evolution

Metabolic engineering is the design and modification of metabolic pathways of a specific organism to produce a natural or non-natural added-value metabolite. We could set the origin of the metabolic engineering with the first genetic engineering attempts and recombinant DNA technology. Scientist managed to clone the DNA sequence of a protein into a vector that would be introduced into the host organism, which then expressed the protein.

After that, expression systems were improved, adding stronger or inducible promoters to regulate expression, vectors were optimized, and hosts were also selected and improved for an easier manipulation. Finally, protein expression was coupled to the metabolism of the host, to increase the efficiency of a pathway, or to add a new pathway that led to the production of a new metabolite⁴⁸.

Advancements in technology helped metabolic engineering: PCR increased DNA obtention and manipulation, assembly technologies made it easier to clone DNA, and genome editing technologies allowed to manipulate metabolic pathways at genome level. All this allowed to improve the engineering process of microorganism to be used as cell factories.

However, currently metabolic engineering is limited to known pathways and enzymes, which limits the range of metabolites that can be obtained. Synthetic biology could help widening this range of products by creating new synthetic circuits, improved enzymes that catalyze novel reactions, etc.

Erb *et al.*⁴⁹ described the stages of metabolic engineering and its evolution. Figure 0.7 shows the evolution of solutions -or products available through metabolic engineering- depending on pathway modification and enzyme modification. Erb *et al.* defined 5 levels of metabolic engineering, each level having an advanced knowledge on enzymes and pathways that increases the number of available solutions -reactions and products. In the figure, the enzyme solution space means the number of possible enzyme reactions available and the pathway solution space, the number of possible pathways that can be designed. Levels 1, 2 and 3 (Existing pathways, 'Copy, paste and fine-tuning' pathways, and 'Mix and match' pathways, respectively) have the same enzyme solution space, since are based only on known enzyme. However, levels 4 and 5 (Synthetic pathways with novel reactions, and Synthetic pathways with novel chemistries, respectively), include novel enzyme either designed by protein engineering with known enzyme mechanisms or novel mechanisms.

Not only enhanced pathways are available, but also novel pathways can be obtained through combining known enzymes. However, the number of known enzymes will limit the number of metabolic pathways that can be engineered. Thus, the synthesis of novel enzymes will also increase the number of pathways that can be created in a host. As new synthetic enzymes appear, either with novel reaction or novel mechanism of action, the number of pathways that are available for metabolic engineering increase.

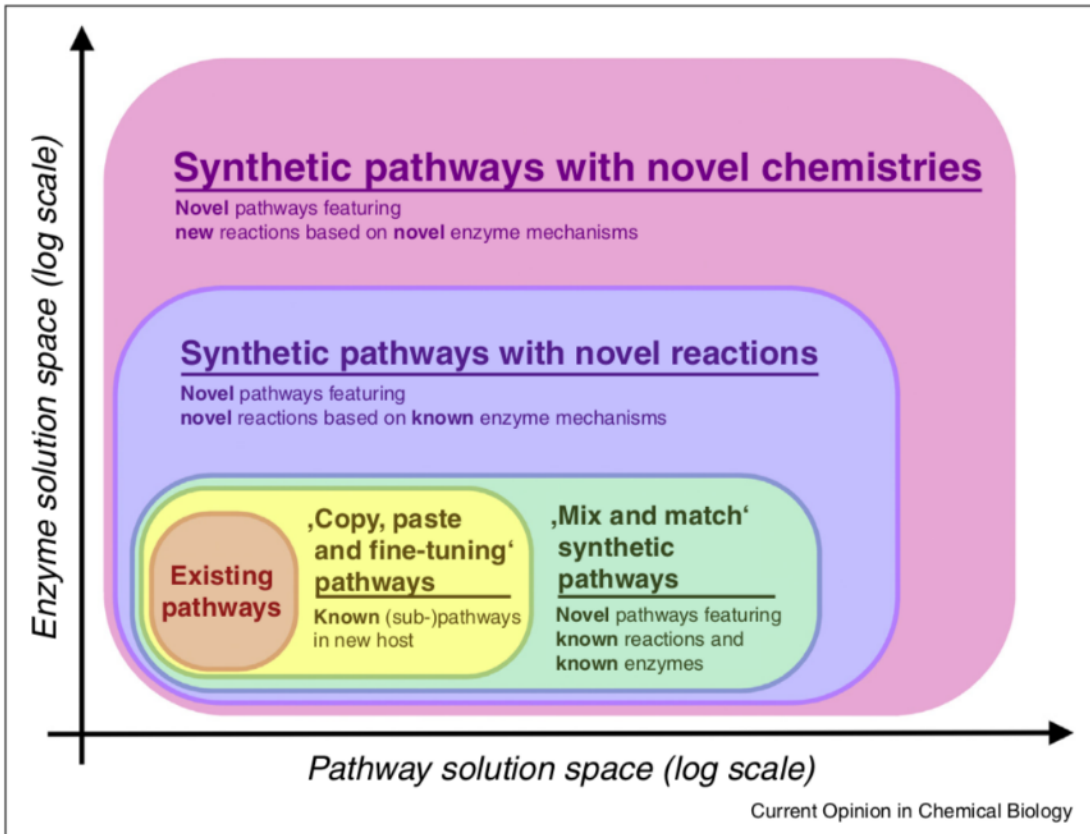


Figure 0.7 The five different levels of metabolic engineering as identified by Erb *et al.*⁴⁹ Image extracted from Erb *et al.*



1.2.2 Examples of metabolites obtained by metabolic engineering

The purpose of metabolic engineering is to find solutions to today's numerous challenges. Because of the high adaptability and versatility of metabolic engineering, many solutions have already been designed. However, they are still increasing and developing as the knowledge is progressing.

Nowadays, the advances in manipulation of DNA in different organisms allows to use them as metabolic engineering hosts, which expands the number of pathways that are available to be exploited. For instance, *E. coli* has been used to produce several high-added value products such as D-glucaric acid⁵⁰, carotenoids -Lycopene⁵¹, β -carotene⁵², zeaxanthin⁵³ or astaxanthin⁵⁴, or bioactive compounds -such as polyketides⁵⁵. Other bacteria have also been engineered, such as *Pseudomonas putida* or *P. aeruginosa* to produce rhamnolipids^{13,56}. Yeast have also been metabolic engineered to synthesize a wide range of metabolites: fatty acid in *Starmerella bombicola*⁵⁷, limonene⁵⁸ and lipids -from growth on lignocellulosic materials- in *Yarrowia lipolytica*⁵⁹ or rhamnolipids in *S. cerevisiae*¹⁴. Nevertheless, microalgae are becoming more important in metabolic engineering due to its photosynthetic capability and ability to produce high-value chemicals⁶⁰).

As well as high-added value products, metabolic engineering can also be used to approach other current challenges such as biofuel production or environmental pollution. Many organisms have been used to produce fatty acids for biofuel conversion, for instance, the already mentioned *S. bombicola* and *Y. lipolytica*, but also *E. coli* for isoprenoid-based biofuels⁶¹ and fatty acid conversion⁶², *S. cerevisiae*⁶³ and *Eulena gracilis*, which is a known biofuel-producer microalgae species⁶⁴. As for bioremediation, solutions are challenging to implement due to the difficulties associated to non-optimal conditions, different from laboratory scale, high concentration of pollutants, presence of simple carbon sources and antibiotics from other microorganisms, or high competition for nutrients⁶⁵. However, several approaches have been made, for example with *Pseudomonas* for degradation of PAHs (Polycyclic Aromatic Hydrocarbons)⁶⁶, 1,2,3-trichloropropane⁶⁷ and 2-chlorotoluene^{65,68}. Besides, new advances are being reported, for example CO₂ capture with *E. coli*⁶⁹ or toxic metals and metalloids with rhizobacteria⁷⁰. Furthermore, this field is continuously expanding, and metabolic engineering tools help designing new applications⁷¹.

1.3 Yeast as host for metabolic engineering

One of the most important steps to carry out a metabolic engineering project is to choose the host. Although metabolic reactions are numerous in all organisms, some organisms already have the enzymes necessary for reactions to catalyze a step for the final product or its precursors, but other organisms do not. Moreover, some organisms are well known and already have many technologies for its manipulation, which makes them more suitable for engineering as the modification of DNA and expression in the host will be more straightforward.

To produce α -galactosylceramide KRN7000 by metabolic engineering, the enzyme (BF9343_3149 GT from *B. fragilis*) and two substrates are required: UDP-galactose as the sugar donor and phytoceramide as the lipid acceptor. The enzymatic reaction is presented in Figure 0.8:

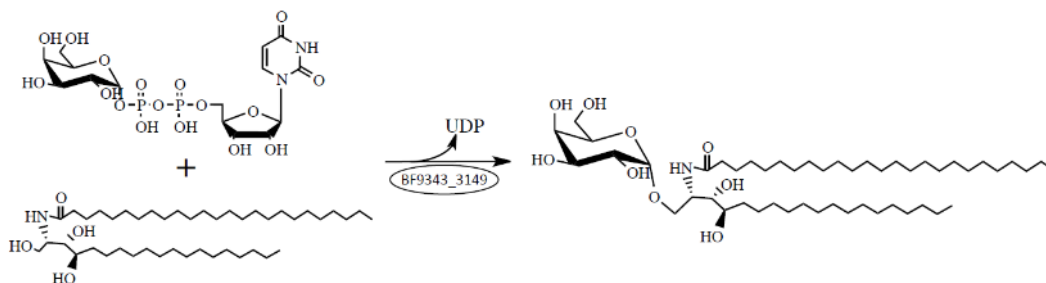


Figure 0.8 α -galactosylceramide synthesis reaction catalyzed by BF9343_3149 glycosyltransferase from *B. fragilis*. Enzyme (BF9343_3149, α -galactosylceramide synthase, in circle) uses UDP-galactose (top-left) as donor and transfers the sugar to the phytoceramide acceptor (bottom-left), releasing UDP (uridyl-diphosphate), to produce α -galactosylceramide.

Saccharomyces cerevisiae can also synthesize glycosphingolipids, even though, with different structure from α -galactosylceramide. Therefore, yeast metabolism contains both metabolic pathways of synthesis of the sugar donor, UDP-galactose, and acceptor, phytoceramide.

In the lines below a brief description of *S. cerevisiae* is explained, along its cellular organization, the importance of glycosphingolipids in the cell, the description of the most important metabolic routes for the purpose of this thesis, and finally, some examples of metabolic engineering approaches using yeast as host for production of lipid-derived products.

1.3.1 Organism description

S. cerevisiae is a eukaryotic unicellular organism, an *Ascomycota* from the fungi kingdom. It has been used from ancient times for many applications, such as beer brewing, baking or winemaking. Its importance in human manufacturing industry along the ages has made it one of the most thorough studied organisms, being a model organism for molecular and cell biology.

Regarding its morphology, *S. cerevisiae* cells have ovoid shape, with their distinctive buds, which are its form of reproduction, and their size is 5-10 μm in diameter.

As a eukaryotic cell, yeast cell is organized in organelles. Briefly explained and shown in Figure 0.9: the **nucleus** contains the DNA of the cells, and gene expression and DNA replication takes place; in **mitochondria** the tricarboxylic acid (TCA) cycle for aerobic respiration occurs; lipids are synthesized in the **endoplasmic reticulum** and cytoplasm; lipid β -oxidation is also taking place in cytoplasm and **peroxisome**; **cytoplasm** contains many reactions, such as glycolysis, and the solutes of the cell (that are not in an organelle) and the organelles, as well as the cytoskeleton; **Golgi apparatus** is a place where

proteins and molecules are modified and sorted to other parts of the cell; the **vacuole**, which helps maintaining pH and ion-homeostasis, serving as ion storage and, in nutrient deprivation, as a degradation compartment for proteins or even other organelles; and finally, **cell membrane** contains the cytoplasm and regulates the uptake and secretion of molecules, and **cell wall** surrounds the cell membrane, consisting predominantly of β -glucan, confers a strong protection to mechanical stress or osmotic pressure, and determines cell shape and integrity.

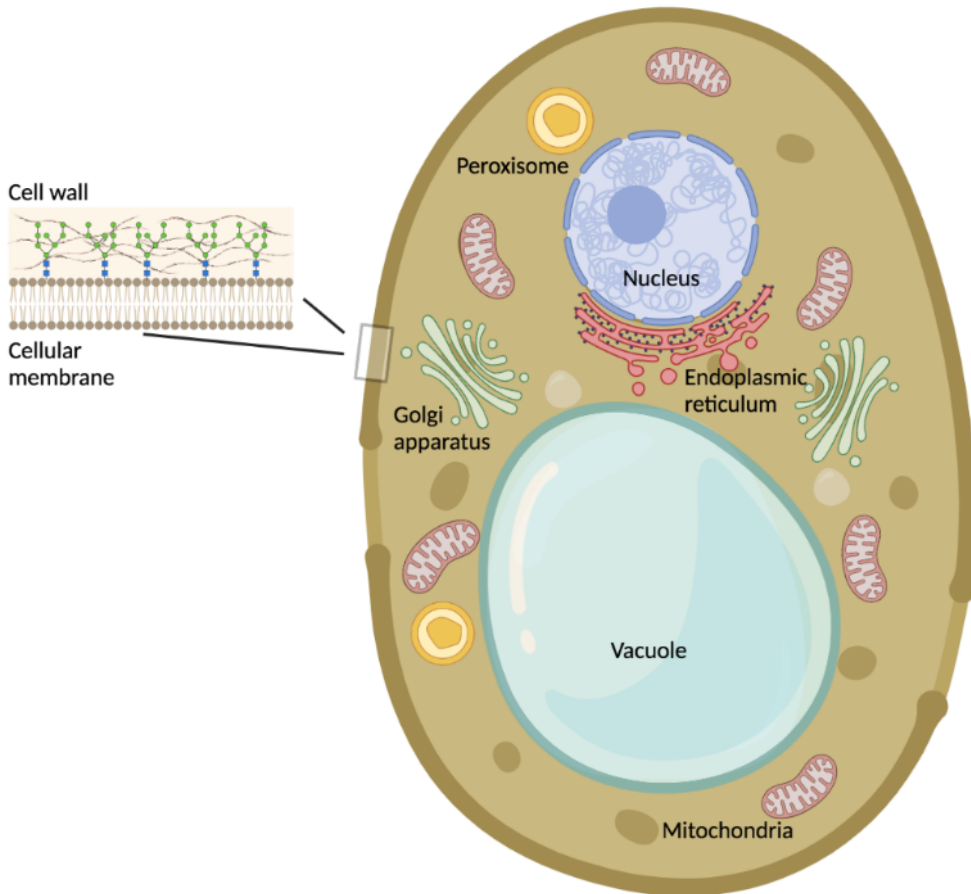


Figure 0.9 *S. cerevisiae* cellular organization with organelles. Yeast cell contains different organelles that have different function: Nucleus (blue), Endoplasmic reticulum (red), Golgi apparatus (green), Mitochondria (brown), peroxisome (yellow), and vacuole (light blue). Cell membrane and wall are presented in a close-up picture at the left: cellular membrane consist of a lipid bilayer, and cell wall is formed of β -glucan with other biomolecules as depicted. Image created with BioRender.com.



1.3.2 Metabolism

Three metabolic pathways are especially important in this project since they are specially linked to glycosphingolipid synthesis and the synthesis of the desired metabolite α -galactosylceramide: nucleotide-sugar synthesis (for UDP-galactose), ceramide synthesis and glycosphingolipid synthesis (for endogenous GSL synthesis that will compete with α -GalCer synthesis). Whereas nucleotide-sugar synthesis occurs in the cytoplasm, ceramide and GSL synthesis are in specific compartments: endoplasmic reticulum and Golgi apparatus, respectively.

Furthermore, it is important to identify reactions that are linked to these pathways, since they can compete for the same substrate needed for the synthesis of α -GalCer, 'sinking' the metabolites and reducing the yield of the desired product.

Figure 0.10 shows the metabolic routes of these three pathways (plus glycolysis), including enzyme and metabolites that participate in them. Other metabolic routes that are linked to these pathways are also indicated, though they are only named. Finally, the different compartments where the different routes take place are also drawn in different colors.

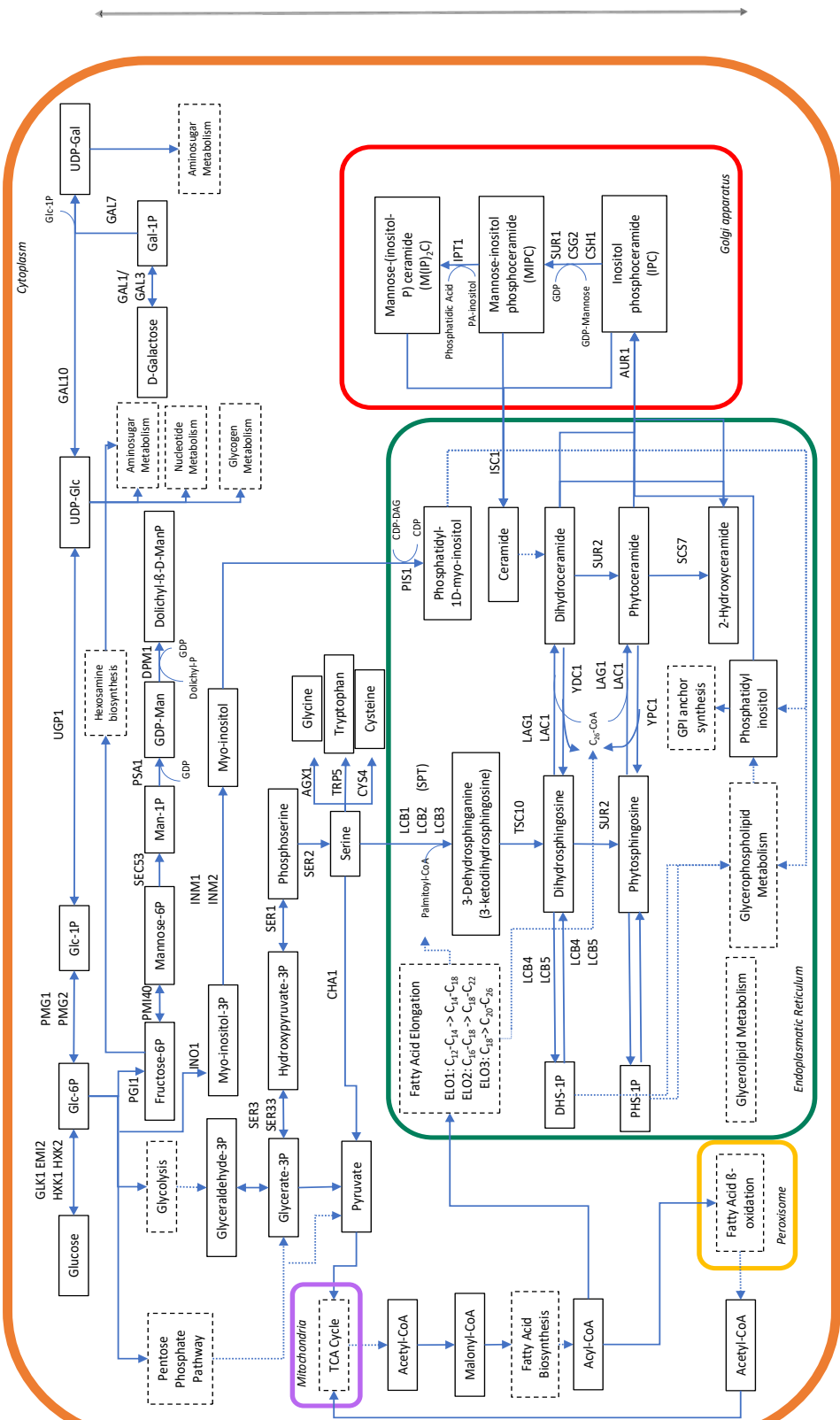


Figure 0.10 *S. cerevisiae* metabolism map. The picture includes detailed pathways for ceramide, UDP-nucleotide and GSL synthesis in yeast, glycolysis is also described. Other pathways are included but only their names appear in a dotted-line box. Arrows indicate the direction of the reaction. Genes that encode the enzymes catalyzing the reaction appear on top or next to the arrow. Metabolites are indicated in straight-lined boxes. Smaller letters with arrows linked to reactions indicate other metabolites required for the reaction. Colored lines indicate the compartments where reaction takes place: cell membrane (orange), mitochondrion (purple), endoplasmic reticulum (green), peroxisome (yellow) and Golgi apparatus (red).

1.3.2.1 Nucleotide-sugar synthesis pathway

One of the substrates needed for the α -galactosylceramide synthesis is UDP-galactose. This metabolite is synthesized through the nucleotide-sugar pathway and can be synthesized from glucose or/and galactose (Figure 0.11). In the case of glucose, it is first converted to glucose-6-phosphate by glucokinases (*GLK1*). This metabolite connects UDP-sugar pathway with glycolysis. Glucose-6P is then converted to glucose-1P by action of phosphoglucomutase (*PGM1/2*), which follows the route of nucleotide-sugar synthesis and is converted to UDP-glucose by *UGP1* UTP-glucose-1-phosphate uridylyltransferase. Finally, UDP-glucose is converted to UDP-galactose by UDP-glucose 4-epimerase *GAL10*.

Alternatively, UDP-galactose can be formed starting from a galactose molecule, which is phosphorylated to galactose-1-phosphate by galactokinase *GAL1* (regulated through Gal3p). Finally, galactose-1P is converted to UDP-galactose by *GAL7* UDPglucose-hexose-1-phosphate uridylyltransferase.

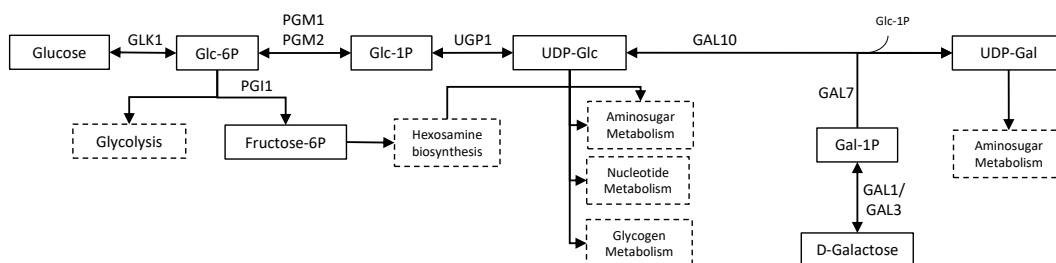


Figure 0.11 Nucleotide-sugar synthesis pathway in *S. cerevisiae*. The metabolic route starting point can be considered glucose, although glucose-6-phosphate (Glc-6P), that is also used for glycolysis, can also follow the pathway. Arrows indicate the direction of the reaction. Genes that encode the enzymes catalyzing the reaction appear on top or next to the arrow. Metabolites are indicated in straight-lined boxes. Smaller letters with arrows net to reaction arrows indicate other metabolites required for the reaction. Pathways that are linked to this pathway are included in a dotted-line box. Legend: *GLK1*, Glucokinase (EC 2.7.1.1); *PGM1/2*, phosphoglucomutase PGM1/2 (EC 5.4.2.2); *UGP1*, UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9); *GAL10*, bifunctional UDP-glucose 4-epimerase / aldose 1-epimerase (EC 5.1.3.2 / EC 5.1.3.3); *GAL7*, UDPglucose-hexose-1-phosphate uridylyltransferase (EC 2.7.7.12); *GAL1*, Galactokinase (EC 2.7.1.6); *GAL3*, Transcriptional regulator *GAL3*; Glc-1P, Glucose 1-phosphate.

1.3.2.2 Ceramide synthesis pathway

Biosynthesis of ceramide (Figure 0.12) derivatives starts with two substrates: serine and palmitic acid (bound to coenzyme A: palmitoyl-CoA). These substrates are condensed to form 3-dehydrosphinganine by serine palmitoyltransferase (SPT), which is composed of different subunits, *LCB1*, *LCB2* and *LCB3*. Then, 3-dehydrosphinganine is reduced to dihydrosphingosine by 3-dihydrosphinganine reductase *TSC10*.

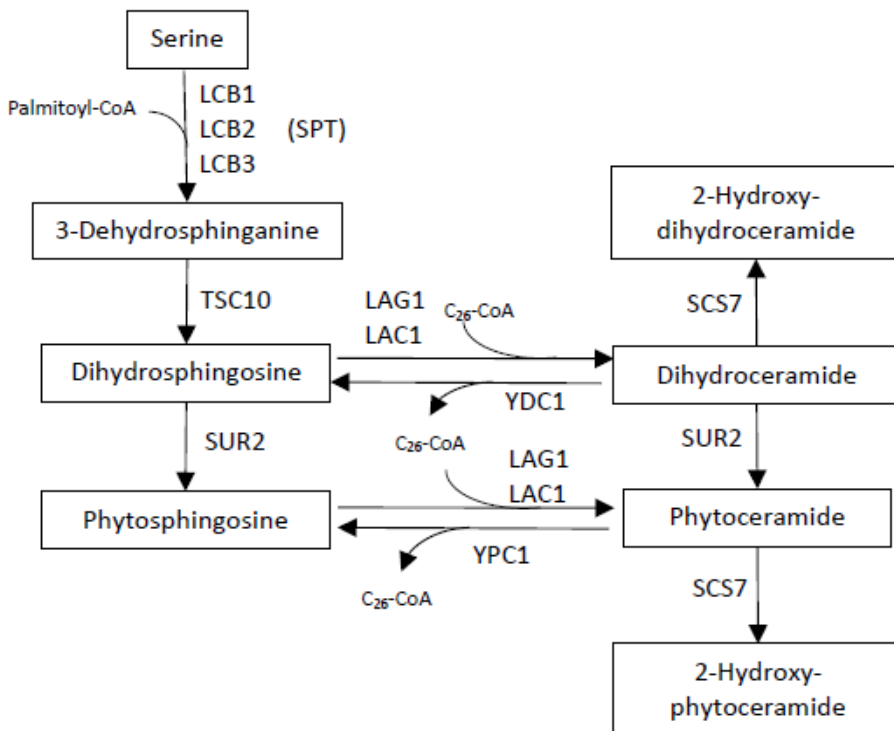


Figure 0.12 Ceramide synthesis pathway in *S. cerevisiae*. The metabolic route starting point is the condensation of Serine with Palmitoyl-CoA. Arrows indicate the direction of the reaction. Genes that encode the enzymes catalyzing the reaction appear on top or next to the arrow. Metabolites are indicated in straight-lined boxes. Smaller letters with arrows net to reaction arrows indicate other metabolites required for the reaction. Pathways that are linked to this pathway are included in a dotted-line box. Legend: *LCB1*, Serine C-palmitoyltransferase LCB1 (EC 2.3.1.50); *LCB2*, Serine C-palmitoyltransferase LCB1 (EC 2.3.1.50); *LCB3*, dihydrosphingosine 1-phosphate phosphatase; *TSC10*, 3-dehydrosphinganine reductase (EC 1.1.1.102); *SUR2*, sphingosine reductase (EC 1.14.18.5); *LAG1*, very-long-chain ceramide synthase (EC:2.3.1.297); *LAC1*, very-long-chain ceramide synthase (EC:2.3.1.297); *YDC1*, dihydroceramidase (EC 3.5.1.-); *YPC1*, phytoceramidase (EC 3.5.1.-); *SCS7*, 4-hydroxysphinganine ceramide fatty acyl 2-hydroxylase (EC 1.14.18.6); CoA, Coenzyme A; C₂₆, acyl carbon chain of

Dihydrosphingosine can follow two reactions: (1) acylation by very-long-chain ceramide synthases (*LAG1* or *LAC1*) using an acyl-CoA (usually 26-carbon chains); (2) hydroxylation by sphingosine hydroxylase *SUR2*. The first reaction forms dihydroceramide, which is later hydroxylated to form phytoceramide (which can be also hydroxylated forming hydroxy-phytoceramide). The second reaction yields phytosphingosine, which is then acylated to synthesize phytoceramide.

Ceramides can be hydrolyzed to release its respectively sphingosine and an acyl chain by action of *YDC1* or *YPC1* ceramidases. Although both ceramidases can accept both ceramide species, *YDC1* has preference for dihydroceramide, and *YPC1* for phytoceramide.

Finally, ceramide species can be further hydroxylated in their acyl chains by very long chain fatty acid hydroxylase *SCS7*. This enzyme can hydroxylate both dihydroceramide and phytoceramide to produce 2-hydroxy-ceramides. Although the predominant ceramide species is phytoceramide in yeast (also its 2-hydroxylated form), dihydroceramide could also be found (Figure 0.13).

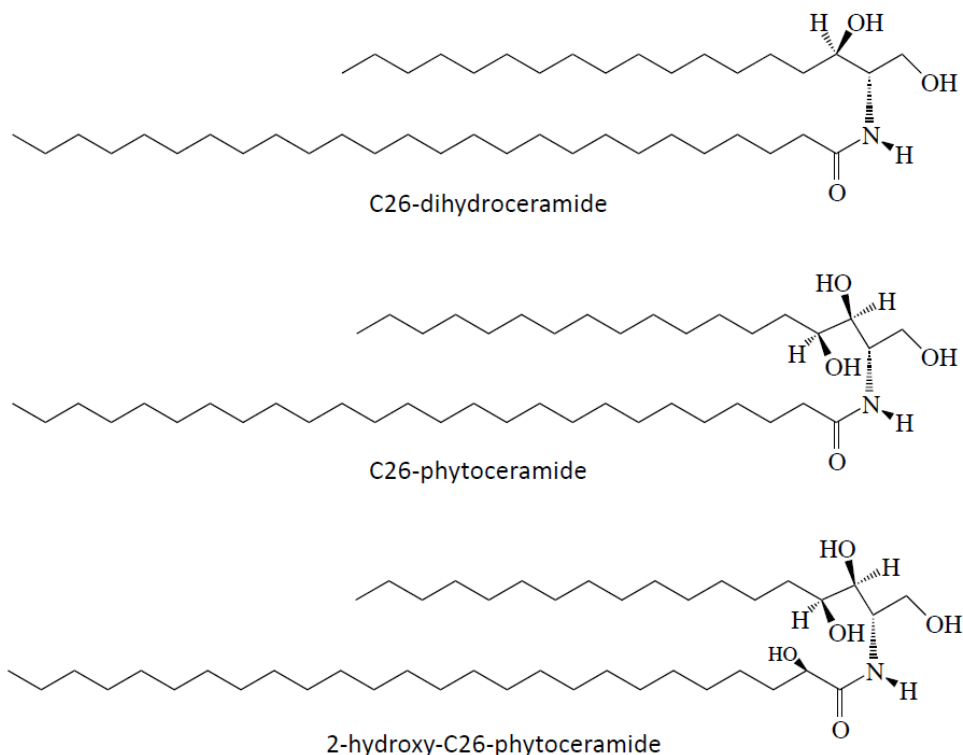


Figure 0.13 Yeast ceramide structures. The image represents the most predominant species in yeast (C26-dihydroceramide, C26-phytoceramide, and 2-hydroxy-C26-phytoceramide).

1.3.2.3 Glycosphingolipid synthesis pathway

Glycosphingolipids are synthesized from ceramide. Therefore, GSL may contain different structures of ceramides. In yeast, there are three types of GSL: inositol phosphorylceramide (IPC), mannosyl phosphorylinositol ceramide (MIPC) and mannosyl di-(inositolphosphate) ceramide (M(IP)₂C).

As Figure 0.14 shows, *AUR1* (inositol phosphorylceramide synthase) is the first enzyme of the reaction catalyzing the synthesis of IPC by transferring inositol-P from inositol-phosphatidic acid (PA-inositol) and releasing phosphatidic acid. Then IPC is mannosylated by action of mannosyl phosphorylinositol ceramide synthase *SUR1* (*CSG2* regulates *SUR1* activity, *CSH1* encodes for another MIPC synthase), MIPC. In the last step, MIPC is the substrate of inositolphosphotransferase 1 (*IPT1*), which catalyzes the addition of inositol-P to MIPC, producing M(IP)₂C.

Finally, all GSL can be hydrolyzed by phosphoinositide phospholipase C (*ISC1*), releasing ceramide again.

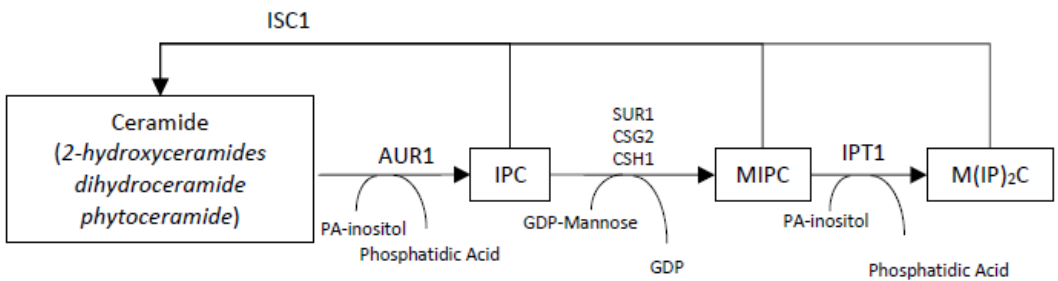


Figure 0.14 Complex sphingolipids (GSL) synthesis pathway in *S. cerevisiae*. The metabolic route starting point is ceramide (which can include different species). Arrows indicate the direction of the reaction. Genes that encode the enzymes catalyzing the reaction appear on top or next to the arrow. Metabolites are indicated in straight-lined boxes. Smaller letters with arrows next to reaction arrows indicate other metabolites required for the reaction. Legend: *AUR1*, inositol phosphorylceramide synthase catalytic subunit (EC 2.7.1.227); *SUR1*, inositol phosphorylceramide mannosyltransferase (EC 2.4.1.370); *CSG2*, Mannosyl phosphorylinositol ceramide synthase regulatory protein; *CSH1*, inositol phosphorylceramide mannosyltransferase (EC 2.4.1.370); *IPT1*, mannosyl-inositol-phosphoceramide inositolphosphotransferase (EC 2.7.1.228); *ISC1*, phosphoinositide phospholipase C (EC 3.1.4.11). PA, phosphatidic acid; GDP, guanosine diphosphate; IPC, inositol phosphorylceramide; MIPC, mannosyl phosphorylinositol ceramide; M(IP)₂C, mannosyl di-(inositolphosphate) ceramide.

1.3.3 Glycosphingolipids in yeast cell biology

In yeast, GSL are implicated in many cell functions, which are related to their presence in the cell membrane. GSL can be observed in organized microdomains called 'rafts', which also contain ergosterol to stabilize these lipids in the membrane⁷². These sphingolipid-enriched rafts which have the property to aggregate proteins, which in turn help stabilize the raft. This aggregation is involved in protein trafficking and sorting to the membrane^{8,9}. The rafts are used for some proteins as an anchor to form the eisosome¹⁰, an immobile rigid subcellular structure of the membrane which is essential for signal transduction, response to stress, cell homeostasis and proton flux^{8,10,73}.

Some of these roles have been elucidated through several yeast mutants with defective synthesis of complex sphingolipids. For instance, inhibition or deletion of *AUR1* protein was found to be lethal, however, some mutants showed that the lethality could be avoided by replacing C26 ceramides with C18 ceramides⁷⁴. MIPC was found to be essential for cell integrity⁷⁵. Moreover, IPC accumulation was found to be associated with regulation of Ca²⁺ homeostasis, since *csg1/csg2* mutants could not grow in high calcium concentration, and this growth inhibition was reversed by deleting IPC synthesis⁷⁶.

1.3.4 Metabolic engineering of yeast *S. cerevisiae*

Considering the previous knowledge on yeast metabolic pathways, enzymes, and DNA manipulation, *S. cerevisiae* has been established as a major host for metabolic engineering applications. Furthermore, yeast total lipid content can range between 3.5 and 10.7% of total cell dry weight^{77,78}, which makes the organism suit for production of lipids and lipid derivatives.

Biofuels are one of the metabolic engineering applications with an important impact on our society. Many attempts on engineering yeast lipid metabolism have been tried to enhance fatty acid production. Some examples have been the reversal of the β -oxidation cycle in *S. cerevisiae*⁷⁹ for medium-chain fatty acids and medium-chain fatty acid ethyl esters and n-butanol production. Expression of heterologous enzymes with higher affinity towards specific substrates combined with deletion of reversal reactions also yielded high production of free fatty acids and fatty alcohols^{63,80}, which are basic products of biofuels and other oleochemical compounds.

Other products of industrial interest had been obtained through engineering of a yeast platform. Isoprenoids which are the basis of fine chemicals, such as Paclitaxel, vitamin K or tetrahydrocannabinol (THC), and of bulk chemicals such as limonene, farnesene or isoprene. Yeast presents the native mevalonate (MVA) pathway and different strategies such as improvement of acetyl-CoA availability, control of MVA pathway competition, redesign of prenyl phosphate metabolism for C10 and C20 (rather than C15) isoprenoids or modification of cytochromes P450 have been used⁸¹. In addition, an innovative approach on engineering the endoplasmic reticulum compartment of yeast to produce triterpenoids has also been developed⁸².



Furthermore, engineered yeast has been developed to obtain aromatic compounds such as shikimic acid, coumaric acid, vanillin, or resveratrol⁸³, or glycolipids such as rhamnolipids¹⁴.

In conclusion, *S. cerevisiae* can be a versatile host since its metabolism can be enhanced and expanded through metabolic engineering. A wide range of products can be synthesized and, due to the adaptability of yeast to different nutrient sources, renewable feed can also be used for the obtention of such high added-value products.

1.4 Framework of the project

The importance of vaccinating the population to improve public health has been highlighted in the past recent years with the Covid-19 pandemic. Pharma industry is continuously growing and searching for new bioactive molecules that can increase vaccine efficiency. Therefore, the necessity of the industry to improve vaccines requires new adjuvants that can boost the response of the immune system.

The studies on α -galactosylceramide effect on immune system point out that it is an interesting metabolite to be used as vaccine adjuvant. The synthesis of this molecule is difficult, thus, new methods for its obtention are needed. A potential approach would be metabolic engineering to obtain a strain capable of producing α -galactosylceramide.

The Laboratory of Biochemistry of the Institut Químic de Sarrià has a prolonged experience in working with glycosyltransferases that catalyze glycolipid synthesis and metabolic engineering of glycolipids in *E. coli* as a model. For instance, MG517, which catalyzes the synthesis of β -glucosyldiacylglycerol in *M. genitalium*, was identified and characterized⁸⁴. Furthermore, this enzyme was also cloned and expressed in engineered *E. coli* strains, to create a biological platform for production of glycolipids using *E. coli* as host^{47,85}. For these complex glycolipids, the pathways of the sugar and lipid metabolism were both engineered, and several modifications were observed to be crucial. Due to this experience in glycosyltransferases and metabolic engineering, it was decided to expand this knowledge to produce α -GalCer using yeast as a host since both precursors are presented in the cells.

Different strategies are developed in this thesis to produce the complex glycosphingolipid. Yeast is an ideal candidate since it can synthesize both precursors of the α -GalCer synthesis: UGP-galactose and phytoceramide. However, yeast lacks the glycosyltransferase that catalyzes the synthesis of this metabolite since α -GalCer is not present in yeast. Although α -GalCer synthase was identified recently³⁸, when this project started it had not been discovered yet. Therefore, the identification of the enzyme has also been carried out in this project, together with its biochemical characterization. Then, the enzyme could be expressed in yeast to evaluate the production of α -GalCer in the host.

Introduction



Objectives



Objectives



Objectives





Objectives

The main goal of this thesis project is to create a novel biological platform for production of α -galactosylceramide by metabolic engineering using *S. cerevisiae* as a host for the cell factory. Therefore, the focus of the project is to redesign of the metabolic pathways for the synthesis of the sugar and lipid precursors, to find and characterize the glycosyltransferase that catalyzes α -galactosylceramide production and generate the new engineered yeast strains.

Therefore, the specific objectives of this thesis are:

1. To redesign the metabolic pathways to increase the pool of the precursors of the complex glycolipid α -galactosylceramide considering the nucleotide sugar and lipid pathways.
2. To generate the engineered strains with the overexpression and knock-out of identified key enzymes involved in the ceramide and complex sphingolipid biosynthesis.
3. To identify the glycosyltransferase enzyme in *Bacteroides fragilis* that is responsible of the synthesis of α -galactosylceramide and perform the biochemical characterization.
4. To evaluate the expressions of the glycosyltransferase in different cellular compartments in the host yeast strain.

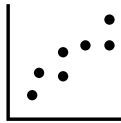
Objectives



Results



Results



Results





Chapter 1

Design of a metabolic engineering strategy to produce α -galactosylceramide



Chapter 1. Design of a metabolic engineering strategy to produce α -galactosylceramide





1.1 Optimizing host metabolism

1.1.1 Ceramide and glycosphingolipid metabolic pathways: a bibliography study

To produce α -galactosylceramide using *S. cerevisiae* as a host not only is necessary to express the heterologous *B. fragilis* α -galactosylceramide synthase (BF3149), but also to optimize the host metabolism to maximize product synthesis.

Then, the pathway of ceramide and glycosphingolipid (GSL) metabolism is analyzed. Previous approaches on yeast ceramide pathway modification are presented to design the new host and, finally, perform the modifications in the organism.

1.1.1.1 Bibliography study

The first step on the design of a new metabolically engineered yeast strain is to explore previous works on the metabolism pathways to be modified. This will allow to know: (1) the effects of desired modifications that have already been studied, (2) byproducts that can form, and (3) compromised cell viability due to modifications.

The target product of this project is α -galactosylceramide, then the reaction between both precursors UDP-Gal and phytoceramide is the transfer of galactose from the donor UDP-Gal to the acceptor lipid in α configuration. The biosynthesis of ceramide and derivatives has been studied in yeast largely since it is a pathway (See Introduction for more details of this pathway) that shows homologies to the pathway of human cells. Yeast ceramide differs from human/mammals in the conversion of the precursor dihydrosphingosine (DHS) into phytosphingosine (PHS) through a hydroxylation, whereas DHS is desaturated and converted into sphingosine; and the use of predominantly C26 acyl chains while humans/mammals' ceramides show different acyl chains depending on the tissue they are located - although unsaturated C24 acyl chain are common^{86,87}.

In Figure 1.1 it can be observed the mentioned differences among ceramide and sphingosine species in both human and yeast. PHS has a hydroxylation at position C4 of dihydrosphingosine (DHS), catalyzed by an hydroxylase encoded by *SUR2* gene. Sphingosine (human/mammals) has a desaturation at the same position. In humans, the desaturation of this position is catalyzed by the desaturase encoded by *DES1* gene. Besides, yeast Scs7 α -hydroxylase catalyzes the synthesis of ceramides containing α -hydroxy fatty acids. Therefore, main yeast sphingolipids contain two hydroxylation and a C26 acyl chain.

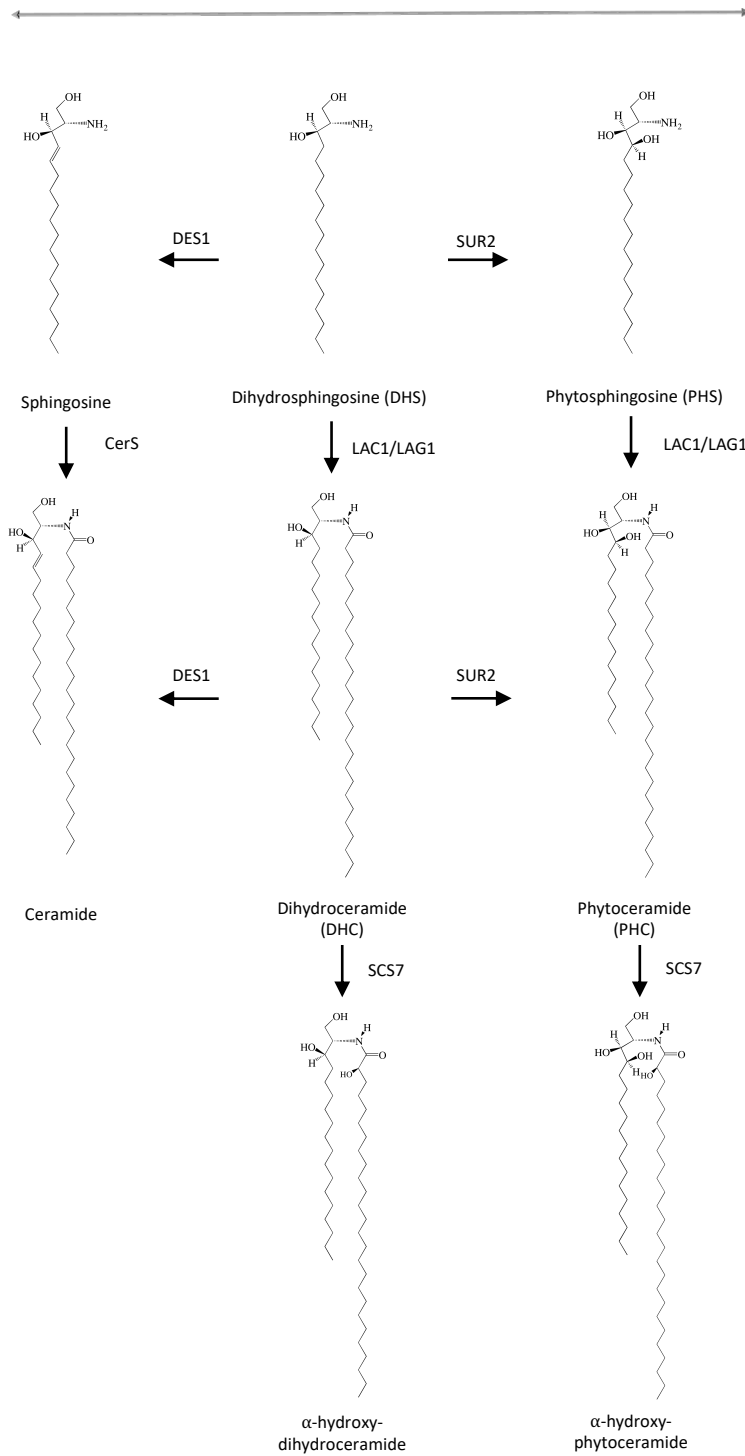


Figure 1.1 Sphingosine and ceramide species in yeast *S. cerevisiae* and *Homo sapiens*. Here it is shown ceramide-NS referred in Murakami *et al.*⁸⁸, which uses a common acyl chain length of 24 carbons. Arrows indicate the transformation from one species to another, on top/right of the arrows it is indicated the enzymes that catalyze the reaction: DES1 (human), SUR2 (yeast) and SCS7 (yeast).

An interesting metabolic engineering study that has been closely followed in this project was the work of Funato and collaborators work⁸⁸, who reported the use of *S. cerevisiae* as host for production of humanized ceramide.

Murakami and collaborators approached two strategies: first, they redirected the ceramide metabolism towards production of dihydroceramide (DHC), which then would be converted to (human-)ceramide by using *sur2 Δ scs7 Δ* strain expressing human *DES1* desaturase; second, they tried to increase ceramide pool by reducing its hydrolyzation to sphingosine by inactivation of *Ydc1* ceramidase and the overexpression of *Isc1* phospholipase which converts GSL to ceramide (Figure 1.2).

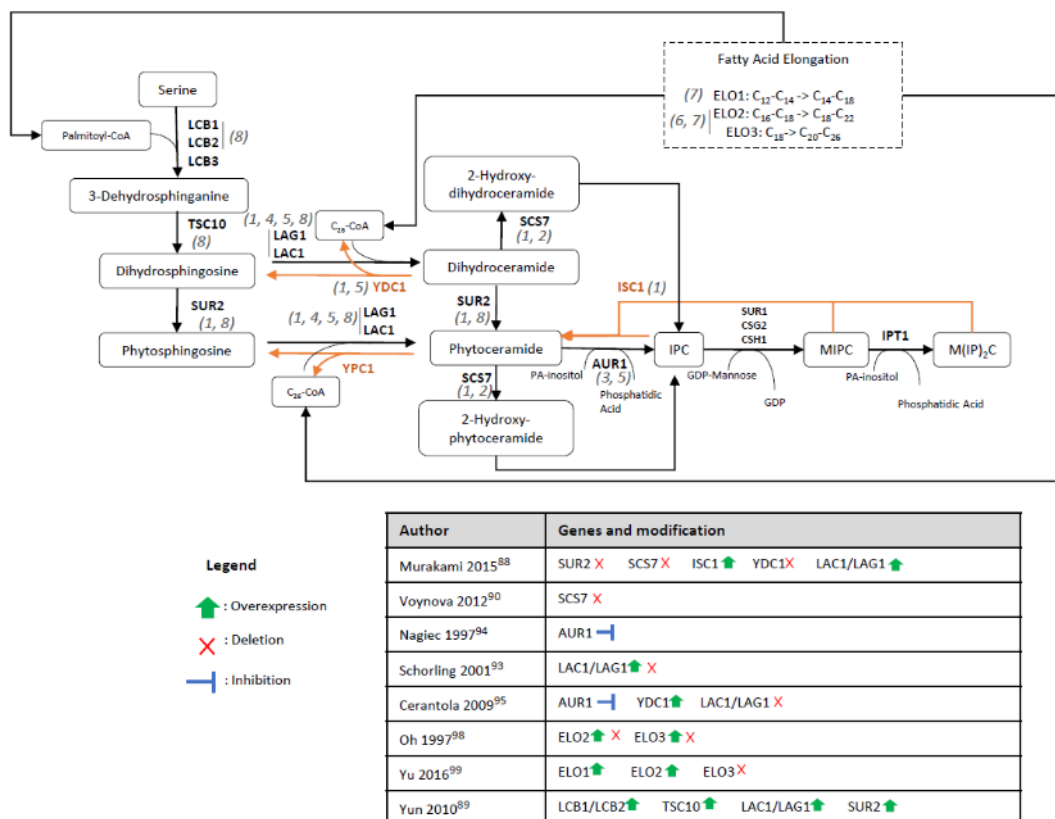


Figure 1.2 Metabolic pathway map of modifications from bibliography. Figure includes both sphingolipid synthesis map and fatty acid elongation enzymes that link to the sphingolipid pathway. Orange arrows and names indicate catabolic metabolism. Numbers in grey next to gene names indicate the reference in the table below. Legend indicates which modification was performed (overexpression, deletion, or inhibition), and table contains the references and the modifications on each gene for each study. Legend: LCB1/LCB2, Serine C-palmitoyltransferase LCB1/LCB2 (EC 2.3.1.50); TSC10, 3-dehydrosphinganine reductase (EC 1.1.1.102); SUR2, sphingosine reductase (EC 1.14.18.5); LAG1/LAC1, very-long-chain ceramide synthase (EC:2.3.1.297); YDC1, dihydroceramidase (EC 3.5.1.-); SCS7, 4-hydroxysphinganine ceramide fatty acyl 2-hydroxylase (EC 1.14.18.6); AUR1, inositol phosphorylceramide synthase catalytic subunit (EC 2.7.1.227); ISC1, phosphoinositide phospholipase C (EC 3.1.4.11); ELO1, fatty acid elongase ELO1 (2.3.1.199); ELO2, fatty acid elongase ELO2 (2.3.1.199); ELO3, fatty acid elongase ELO3 (2.3.1.199).

This study showed that deletion of *SUR2* and *SCS7* genes resulted in the accumulation of DHS and DHC without the α -hydroxyl group, which could then be converted into human-ceramide by Des1 desaturase. Moreover, *YDC1* deletion and *ISC1* overexpression also yielded higher levels of ceramide. In conclusion, these results showed that: (1) it is possible to redirect metabolic flux towards the production of a specific type of ceramide species; (2) *SCS7* and *YDC1* gene deletion does not compromise cell viability; and (3) *ISC1* overexpression and *YDC1* deletion could increase ceramide pool that could be then transformed into a final glycosphingolipid product. They also reported the increase on aureobasidin A (AbA) sensitivity when overexpressing *LAC1* and *LAG1* ceramide synthases.

Another more recent approach on increasing ceramide levels was carried out by Yun *et al.*⁸⁹ analyzing the effects of modification of genes involved in sphingolipid pathway in *S. cerevisiae*. They modified 6 key genes in the pathway separately: *LCB1* and *LCB2* (Serine palmitoyltransferase 1 and 2), *TSC10* (3-ketodihydrosphingosine reductase), *LAC1* and *LAG1* (ceramide synthases), and *SUR2* (sphingolipid C4-hydroxylase). Results showed that, although overexpression of all genes led to some increased ceramide levels, only *TSC10* led to an increase of almost 5-fold levels of ceramide compared to wild type strain.

Additional important results from other studies show that knockout of the *SCS7* gene leads to viable cells⁹⁰. Furthermore, it has been reported that *SCS7* deletion enhances resistance to syringomycin E. The hypothesis is that α -hydroxylation of complex sphingolipids may be key in the lipid-lipid interaction in the membrane, that in turn forms lipid rafts that may serve as syringomycin E binding sites and channel formation⁹¹. However, the lack of these lipid rafts may also induce hypersensitivity for other compounds⁹². These indicates that although *SCS7* deletion⁹² does not affect cell viability directly, the deletion may affect the sensitivity to some drugs if used for strain selection, even more, altering the ceramide and sphingolipid species and ratios which may also influence the cell.

Another key gene that has been studied largely is *AUR1*. This gene encodes for the inositol phosphorylceramide synthase, which catalyzes the first step of the complex sphingolipid synthesis which is the synthesis of inositol phosphorylceramide (IPC). It is largely known that inhibiting the activity of Aur1 synthase is lethal for the cell, however, the reasons of this toxicity have also been argued. First it was thought that ceramide accumulation in the cell was toxic^{93,94}, however, later studies suggested that it is the lack of complex sphingolipids (IPC, MIPC and M(IP)₂C) that can cause the lethality, since these molecules have important roles in cell membrane integrity and others⁹⁵. Aur1 synthase activity inhibition has been achieved either by gene deletion or supplementing growth media with antifungal aureobasidin A (AbA).

Sensitivity to AbA may vary depending on genetic background of the yeast strain since accumulation of different species of ceramide may increase or decrease this sensitivity to the antifungal^{88,95} or other drugs⁹⁶. Other studies showed that *SCS7* and *SUR2* deletion can even change the toxicity of the ceramide accumulation due to *AUR1* suppression^{87,97}. This study suggested that the hydroxylation state of the ceramide influences ceramide toxicity due to its accumulation -caused by the inhibition of *AUR1*. However, each hydroxylase inhibition had a different effect. Thus, it is unknown whether the



overexpression of *SUR2* and the deletion of *SCS7* could attenuate the effect of ceramide accumulation due to *AUR1* deletion or by other means.

Increasing ceramide synthase activity was also studied out by Schorling *et al.*⁹³, who tried to increase ceramide levels by overexpressing *LAC1* and *LAG1*, however, they were unable to obtain a strain with ceramide overproduction. This could be explained due to the toxicity of increased ceramide levels, so cells with increased ceramide levels caused by the overexpression of these genes would not grow and be recovered.

It is also interesting the role of the fatty acid elongases for VLCFA. Since yeast sphingolipids contain mostly VLCFA in their structure, modifying the fatty acid elongation pathway could also help obtaining ceramides with C26 fatty acids.

The first attempt on elucidating how yeast elongases work on VLCFA synthesis and how its modification influences on yeast sphingolipids was made by Oh *et al.*⁹⁸. In this study, they modified *ELO2* and *ELO3* genes, combining overexpression and deletion of both genes and analyzing acyl chain profile in yeast sphingolipids. They observed that by deleting *ELO3*, ceramide did not contain C26-chains, thus indicating that this enzyme is essential on the synthesis of C26 fatty acids. Furthermore, when deleting *ELO2*, C26 chains were present. Another result of the study was that disruption of either *ELO2* or *ELO3* reduced the content of sphingolipid, which could be associated with lower ceramide synthase activity, besides deletion of both caused lethality. This study proved that modifying fatty acid chain in ceramides is possible and cells are still viable.

Overexpressing either *ELO2* or *ELO3* could also boost a specific VLCFA presence in the cell. This was reported more recently by Yu *Et al.*⁹⁹ who demonstrated that overexpressing *ELO2* in *elo3Δ* strain yielded a 1.5-fold level increase of C22 VLCFA and, on the other hand, overexpressing *ELO3* led to a 2-fold level increase of C26 VLCFA. However, this study did not focus on sphingolipid but in fatty acid content overall. Even so, it could be inferred that this could also led to an increase on ceramide with C26 acyl chains. Furthermore, this study also worked on overexpressing *Lag1* and *Lac1* ceramide synthases, however, this was done only to study aureobasidin A sensitivity in strains with *LAC1/LAG1* overexpressed.

To summarize this bibliography study, Table 1.1 contains the information on genes and modifications carried out in each cited study.



Table 1.1 List of all the modification on sphingolipid -or fatty acid elongation- metabolism of *S. cerevisiae* cited in this section. Table includes name of the author with its reference, gene names, modification, and results obtained from the study. NOTE: Not all genes of each study may be included, only the ones that have been mentioned in the previous text.

Reference	Gene	Modification	Results
Murakami 2015 ⁸⁸	<i>SUR2</i>	Deletion	<i>SUR2</i> and <i>SCS7</i> deletion, with the expression of hDES1, led to de production of humanized ceramide (no hydroxylations). Deletion of <i>YDC1</i> led to a 2-fold increase production of ceramide, when <i>ISC1</i> was overexpressed in the same strain, 4-fold increase on ceramide levels was obtained. <i>LAC1/LAG1</i> overexpression led to a higher sensitivity to AbA.
	<i>SCS7</i>	Deletion	
	<i>ISC1</i>	Overexpression	
	<i>YDC1</i>	Deletion	
	<i>LAC1/LAG1</i>	Overexpression	
Voynova 2012 ⁹⁰	<i>SCS7</i>	Deletion	Deletion of <i>SCS7</i> led to the production of ceramide with non-hydroxylated C26.
Nagiec 1997 ⁹⁴	<i>AUR1</i>	Inhibition by AbA	AbA inhibits <i>AUR1</i> , which prevents the cells to produce IPC, which leads to cell death.
Schorling 2001 ⁹³	<i>LAC1/LAG1</i>	Deletion + Overexpression	<i>LAC1</i> and <i>LAG1</i> overexpression did not lead to an overproduction of ceramide. This could be explained that ceramide accumulation is toxic, which is consistent with results when applying AbA and stopping IPC synthase activity. Cells without <i>LAC1/LAG1</i> can live in presence of AbA.
Cerantola 2009 ⁹⁵	<i>LAC1/LAG1</i>	Deletion	Deletion of <i>LAC1/LAG1</i> does not arrest growth in mutant strain. Overexpression of <i>YDC1</i> may lead to IPC synthesis in <i>LAC1/LAG1</i> mutant strains, which could indicate reverse ceramidase activity of Ydc1 ceramidase, which allows growth in presence of AbA.
	<i>YDC1</i>	Overexpression	
	<i>AUR1</i>	Inhibition by AbA	
Oh 1997 ⁹⁸	<i>ELO2</i>	Deletion + Overexpression	Deleting <i>ELO3</i> leads to no presence of C26-chains, thus indicating that this enzyme es essential on the synthesis of C26 fatty acids. When deleting <i>ELO2</i> , C26 chains were present, which indicates <i>ELO3</i> is essential in the C26 fatty acid synthesis. Disruption of either <i>ELO2</i> or <i>ELO3</i> reduced the content of sphingolipid, which could be associated on lower ceramide synthase activity. Deletion of both caused lethality.
	<i>ELO3</i>	Deletion + Overexpression	
Yu 2016 ⁹⁹	<i>ELO1</i>	Overexpression	Deletion of <i>ELO3</i> and overexpression of <i>ELO1</i> and <i>ELO2</i> led to the production of C22 fatty acids (docosanol).
	<i>ELO2</i>	Overexpression	
	<i>ELO3</i>	Deletion	
Yun 2010 ⁸⁹	<i>LCB1</i>	Overexpression	All gene overexpression showed slight increased ceramide levels, except <i>TSC10</i> overexpression which had the most effect on ceramide levels since it led to a 5-fold level increase.
	<i>LCB2</i>	Overexpression	
	<i>TSC10</i>	Overexpression	
	<i>LAC1</i>	Overexpression	
	<i>LAG1</i>	Overexpression	
	<i>SUR2</i>	Overexpression	



1.1.2 Strain design

New engineered *S. cerevisiae* strains could be designed to obtain an improved biological platform to produce α -galactosylceramide (this molecule would contain the phytoceramide structure, therefore, if not indicated otherwise, α -galactosylceramide refers to the structure containing said structure). This means to express the BF3149 GT enzyme in a yeast platform and redirect metabolic flux towards the production of phytoceramide (PHC), the lipidic precursor, increasing phytoceramide pool to be more available to BF3149 enzyme.

First, phytoceramide is the main target of the metabolic redirection approach since it is most similar to the synthetic α -galactosylceramide, with unsaturated acyl chains and hydroxylation at position 4 of the sphingosine backbone. To increase the phytoceramide pool, different strategies are proposed: The first strategy is based on *SUR2* gene overexpression to favor the reaction of conversion of DHS/DHC into PHS/PHC. Although it would not be expected phytoceramide to be the only ceramide species in the new host, it was hypothesized that this approach could help to increase PHS/PHC pools. The second strategy is focused on preventing side-reactions such as α -hydroxylation of the ceramides by *Scs7* α -hydroxylase by deleting the *SCS7* gene. This would eliminate this hydroxylation step, and DHC and PHC would be the only ceramide species in the host.

The third strategy is to increase ceramide availability. One approach could be deleting *AUR1* gene, involved in the first step of complex sphingolipids synthesis, however it is known that it suppresses cell viability^{93,94}. Although cell lethality through *AUR1* inhibition can be avoided^{87,97}, it would imply deleting *SUR2*, which needs to be overexpressed for the purpose of this project. For this objective, Murakami and collaborators described two approaches: the overexpression of *ISC1* phospholipase gene to increase hydrolysis of glycosphingolipids to yield ceramide, and deletion of *YDC1* gene that is responsible of ceramide hydrolysis. Yun and collaborators⁸⁹ also reported that the overexpression of *TSC10* gene, which performs the synthesis of dihydroceramide, had a 5-fold change on final ceramide levels.

From all these possibilities, it was decided to discard *YDC1* deletion due to the complexity of performing a second knockout in the strain. Then, it was decided to carry out only one more gene overexpression to avoid increasing metabolic burden. Thus, *ISC1* overexpression was selected over *TSC10*. Although *TSC10* has been reported to have a higher increase on ceramide levels, *ISC1* can reduce the levels of complex sphingolipids, increasing ceramide, which would make it more available for BF3149 and increase α -GalCer production.

Finally, expression of BF3149 synthase in the yeast is expected to glycosylate the ceramide species and convert it to the α -galactosylceramide derivatives. Expression of a bacterial glycosyltransferase in yeast can be challenging, since in eukaryotic organism these enzymes that perform glycosylations are usually anchored to the membrane of Golgi apparatus. Therefore, the enzyme should be redirected towards this compartment.

On the one hand, it is not known whether the synthesis of a non-natural glycosphingolipid in yeast could compromise cell survival. On the other hand, if α -galactosylceramide produced in *S. cerevisiae* proves to be successful, it would also be interesting if cell lethality caused by the deletion of *AUR1* could be avoided by the synthesis of α -galactosylceramide since endogenous complex glycolipids would be replaced by a new glycosphingolipid.

To summarize all modifications that will be included in the final engineered strain, Figure 1.3 and Table 1.2 show a list of these modification and the representation in a metabolic pathway map.

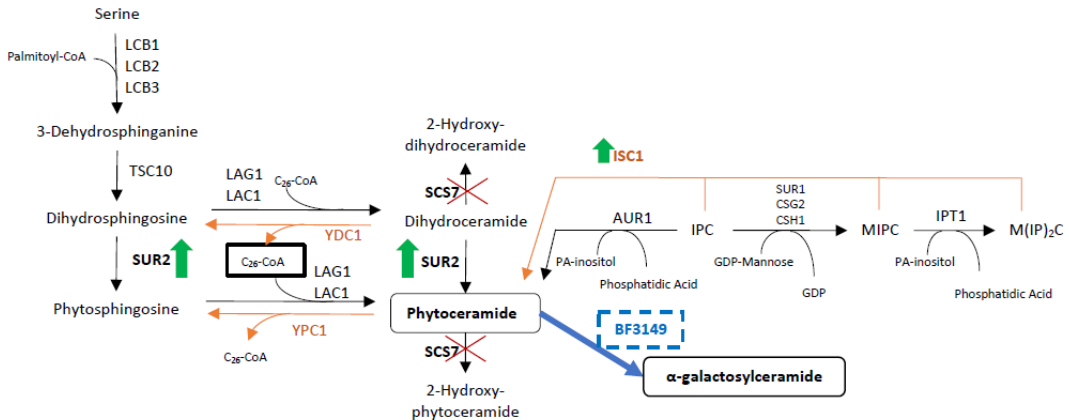


Figure 1.3 Scheme of metabolic flux redirection approach on new *S. cerevisiae* strain. Black box indicates targeted products (phytoceramide as the specific structure and α -galactosylceramide as final product). Black arrows and names indicate anabolic metabolism. Orange arrows and names indicate catabolic metabolism. Blue discontinue box indicates expression of a heterologous gene. Bold blue arrow indicates the heterologous reaction inserted in the metabolism. Bold green arrows next to gene names indicate overexpression to favor desired reaction. Red crosses indicate pathways that are deleted by gene knock-out. Legend: LCB1/LCB2, Serine C-palmitoyltransferase LCB1/LCB2 (EC 2.3.1.50); TSC10, 3-dehydrosphinganine reductase (EC 1.1.1.102); SUR2, sphingosine reductase (EC 1.14.18.5); LAG1/LAC1, very-long-chain ceramide synthase (EC:2.3.1.297); YDC1, dihydroceramidase (EC 3.5.1.-); YPC1, phytoceramidase (EC 3.5.1.-); SCS7, 4-hydroxysphinganine ceramide fatty acyl 2-hydroxylase (EC 1.14.18.6); AUR1, inositol phosphorylceramide synthase catalytic subunit (EC 2.7.1.227); ISC1, phosphoinositide phospholipase C (EC 3.1.4.11); SUR1, inositol phosphorylceramide mannosyltransferase (EC 2.4.1.370); CSG2, Mannosyl phosphorylinositol ceramide synthase regulatory protein; CSH1, inositol phosphorylceramide mannosyltransferase (EC 2.4.1.370); IPT1, mannosyl-inositol-phosphoceramide inositolphosphotransferase (EC 2.7.1.228); BF3149, α -galactosylceramide glycosyltransferase BF3149 from *B. fragilis*; PA, phosphatidic acid; GDP, guanosine diphosphate; IPC, inositol phosphorylceramide; MIPC, mannosyl phoshporylinositol ceramide; M(IP)₂C, mannosyl di-(inositolphosphate) ceramide; CoA, Coenzyme A; C₂₆, acyl carbon chain of 26 carbon.



Table 1.2 List of metabolic engineering modifications for the new α -galactosylceramide production strain of *S. cerevisiae*. Table includes gene names, modifications, the approach that will be followed to apply the modification, and the expected results.

Gene	Modification	Expected results
<i>SUR2</i>	Overexpression	Increase of C4-hydroxylase activity that will increase synthesis of phytoceramide/phytosphingosine over dihydroceramide/dihydrosphingosine
<i>ISC1</i>	Overexpression	Increase phospholipase C activity that will increase overall ceramide pool
<i>SCS7</i>	Deletion	Remove ceramide acyl chain hydroxylation activity
BF3149	Heterologous expression	Addition of α -galactosylceramide synthase activity in yeast strain

1.1.3 Engineering approach

As it is mentioned in the Introduction section, metabolic engineering includes a wide range of techniques to modify the organism metabolic pathways. These techniques are based on genetic engineering and molecular biology. For instance, overexpression of an endogenous gene could be achieved by either transforming with a plasmid with the desired gene, inserting this construct in the genome, or even modify endogenous promoter to increase its expression.

Modifying organism's genome is something that is more complex and requires some effort. Therefore, it was decided that overexpression of *SUR2* and *ISC1* would be done by synthesizing two yeast expression vectors with *SUR2* and *ISC1* genes so phenotypes could be studied separately at first. Later on, they could then be cloned in a single vector to study the phenotype of co-expression of both genes. Since they are endogenous genes, both genes will be obtained through PCR gene amplification from *S. cerevisiae* genome.

As for Bf3149 expression, it was also decided to use an expression vector to study the performance of the enzyme in the yeast cell environment. This gene will be obtained through codon optimization after nucleotide translation; thus, the synthetic gene will be then cloned into the final vectors.

Although other techniques that do not need to modify the genome can be used to reduce the expression at transcription or even post-translational level, removing completely the product of a gene requires modifying the genome. CRISPR/Cas9 technology -see Introduction- will be used for *SCS7* gene knockout to obtain the new $\Delta scs7$ mutant strain.

Selection of the *S. cerevisiae* strain is another decision for the final engineered strain. Although many strains of *S. cerevisiae* are available, it was agreed to use strain RH6082¹⁰⁰ which was already used in several studies on ceramide metabolism^{88,100,101}, so it was decided to use this as the host for our metabolic engineering approach. After request, Dr. Howard Riezman from University of Geneva kindly provided the strain.

1.2 Summary

As conclusions, the metabolic engineering approach on RH6082 yeast strain will consist of the next steps (see Figure 1.4):

- Design and construction of *SUR2* and *ISC1* overexpression plasmids
- Analysis of *SUR2* and *ISC1* overexpression in the engineered strains
- Design and construction of CRISPR/Cas9 knockout vectors for *SCS7*
- *SCS7* gene knockout
- Analysis of the *scs7* Δ engineered strain
- Design and construction of BF3149 expression vector
- Analysis of the different engineered yeast strains overexpressing Bf3149 galactosyltransferase

These steps will compose the overall project, which will result with the production of the α -galactosylceramide derivatives of the new engineered strains reporting the key modifications to forward the lipid pathway to phytoceramide and to synthesize the glycolipid.

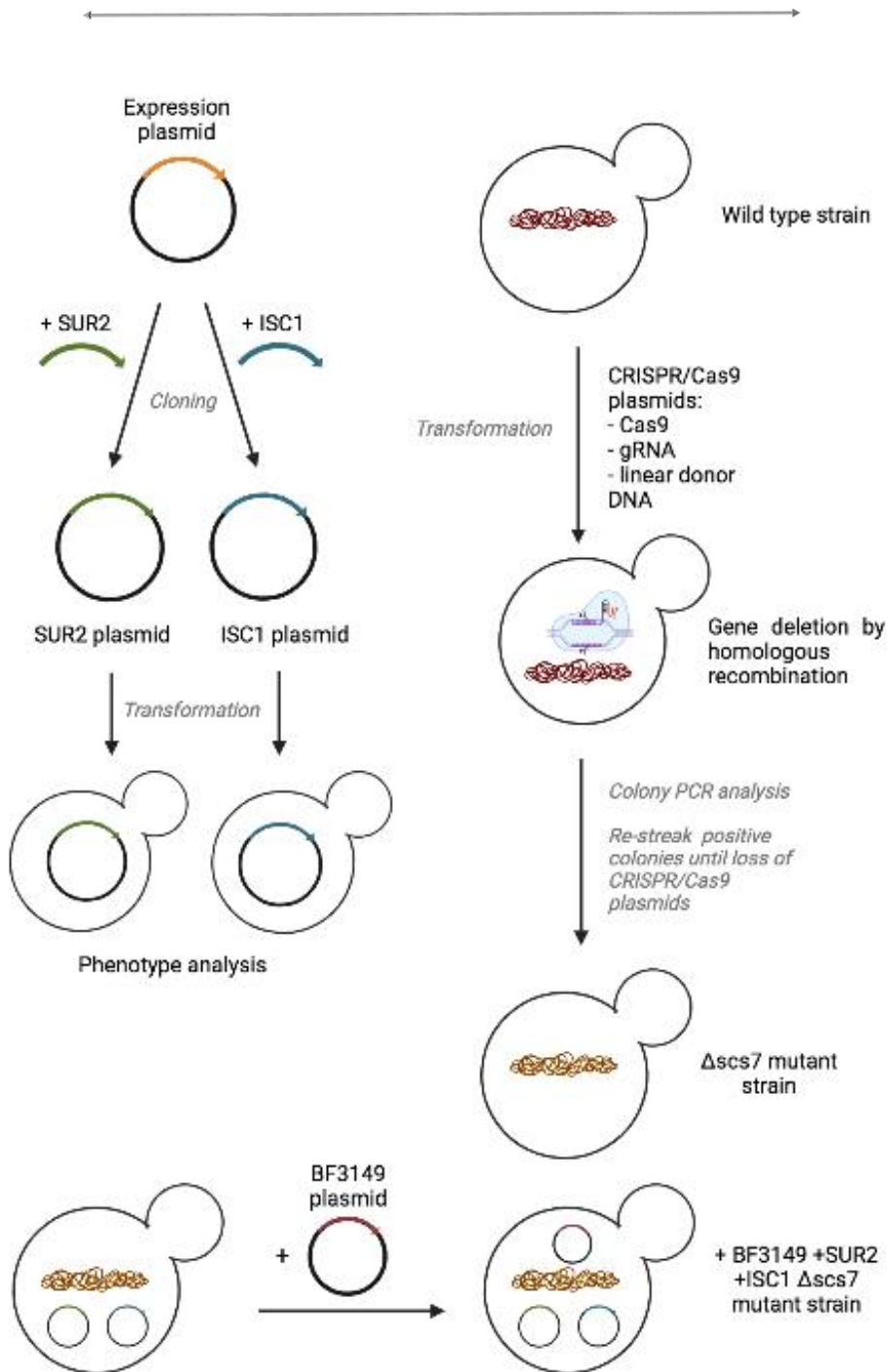


Figure 1.4 Scheme of metabolic engineering strategy on creating a new *S. cerevisiae* strain for production of α -galactosylceramide. Circular forms with colors arrows indicate DNA plasmids with the different genes. Circular form with another smaller circle (bud) attached represents yeast morphology. Arrows indicate succession of steps. Image created with BioRender.com.

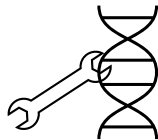
Chapter 1. Design of a metabolic engineering strategy to produce α -galactosylceramide





Chapter 2

Engineering a *S. cerevisiae* strain metabolism to optimize phytoceramide production







2.1 Redirection of metabolic flux on sphingolipid pathway

2.1.1 Design and construction of *SUR2* and *ISC1* yeast expression vectors

The strategy of metabolic flux redirection implies the overexpression of two key genes: *SUR2* and *ISC1* (the sphingosine reductase and phosphoinositide phospholipase C genes, respectively), in order to increase phytoceramide conversion and recycling of ceramide from complex sphingolipids. Two different constructs were made for each of the genes. However, the same vector was used as backbone. The strategy was to study the effects of overexpression of each gene in a similar genetic environment: strong constitutive promoters of similar expression levels, as well as the same vector. This would avoid an important difference in expression due to different copy number of plasmid and expression level of the promoter.

2.1.1.1 Design of *pTDH3-SUR2* and *pTEF2-ISC1* vectors

The first step to design *SUR2* and *ISC1* plasmids was to select a vector that could be used as background for cloning both genes. There are several factors to take into account when choosing a vector for gene expression in yeast.

Plasmid copy number

Plasmid copy number will determine the number of copies of the plasmid that will be present in the cell. The higher the number of copies, the higher the expression levels, since more copies of the gene are available for the RNA-polymerase to bind and transcript the DNA sequence. Two types of plasmids could be differentiated on their copy number: high copy number and low copy number plasmids. In yeast, YEp (Yeast Episomal plasmid) is a high copy number plasmid that is based on the native 2 μ (2 micron) episomal plasmids native of *S. cerevisiae* and can maintain high numbers of plasmid copies (>10); YCp (Yeast Centromeric plasmid) is a low copy number plasmid with an origin of replication -ARS- and a centromeric sequence -CEN-, which help to maintain only 1-2 copies in the cell¹⁰². Although more plasmid copies can be translated into more gene expression, this is not always proportional. Product toxicity can affect cell growth and viability, which makes high copy number plasmids less suitable for metabolic engineering proposals. Furthermore, YEp stability is also much lower and plasmid can be lost after several generations since it is not equally distributed during cell division. YCp, in contrast, due to the ARS origin of replication and the centromeric sequence, is split after cell division as they are treated as a chromosome by the cell division mechanism.



Promoter

Promoter determines the expression level of the gene and is flanking at 5'. There are several promoters. First, based on the expression level -or strength- they show, there are strong and weak promoters and also a range between them. Second, they can be constitutive or inducible promoters. Constitutive promoters are constantly active whereas inducible promoters are only active when a specific stimulus is present, e.g. galactose presence for GAL promoters.

In metabolic engineering applications, increased expression as a result of a strong promoter is not always optimal as it can compromise cell viability. Besides, promoter strength can also affect plasmid copy number¹⁰².

Terminator

Terminator is present in the flanking 3' region of the gene and affects the mRNA stability, thus, affecting protein synthesis. Considering this, a good terminator will enhance gene expression by improving mRNA half-life, that in turn will increase protein output. Modified or even synthetic terminators have also been created to enhance their strength^{103,104}.

Selection marker

In yeast, selection markers are usually auxotrophic markers, which complement an essential metabolic pathway of the strains. These strains can only grow when either the marker or the essential metabolite is present in the medium. For instance, *LEU2* marker allows a strain, which is *LEU2* gene product defective and cannot synthesize leucine, to grow in a medium without supplemented Leu. The use of one selection marker may be due to convenience, however, it is also important to consider the impact on plasmid copy number due to the plasmid burden¹⁰². For example, *LEU2* selection marker yields lower copy numbers due to plasmid burden, which can also affect cell growth.

Considering all these plasmid characteristics, it was decided to use a low copy number plasmid with a strong constitutive promoter to maintain plasmid stability and avoid compromising cell survival while keeping high expression levels. Since *SUR2* and *ISC1* are also expressed in all cell cycle phases, a constitutive promoter had to be used. In the Laboratory of Biochemistry from IQS and in collaboration with the group of Prof. Marjan de Mey of the Center for Synthetic Biology from Ghent University, the plasmid pTDH3-ScoMG517 was synthesized. The plasmid pTDH3-ScoMG517 contains the gene Mg17 glycosyltransferase which was of interest in a previous metabolic engineering project. This vector would be used as backbone to construct next plasmids with *SUR2* and *ISC1*. Figure 2.1 shows the design of pTDH3-ScoMG517 plasmid and the new designed vectors for *SUR2* and *ISC1* expression.

For *SUR2*, *TDH3* promoter (pTDH3) and *ADH1* terminator (tADH1) would be maintained in the final vector, and ScoMG517 coding sequence would be substituted by the coding sequence of *SUR2*. For *ISC1*, the whole cassette -i.e., promoter, coding sequence and terminator- would be substituted for



TEF2 promoter (pTEF2), *ISC1* coding sequence and *PGK1* terminator (tPGK1). *TEF2* promoter is also a strong constitutive promoter and *PGK1* terminator is also a commonly used terminator similar to *TADH1*¹⁰⁵. Although *LEU2* was not the preferred selection marker, it was already present in the vector, so it was decided to keep it for convenience.

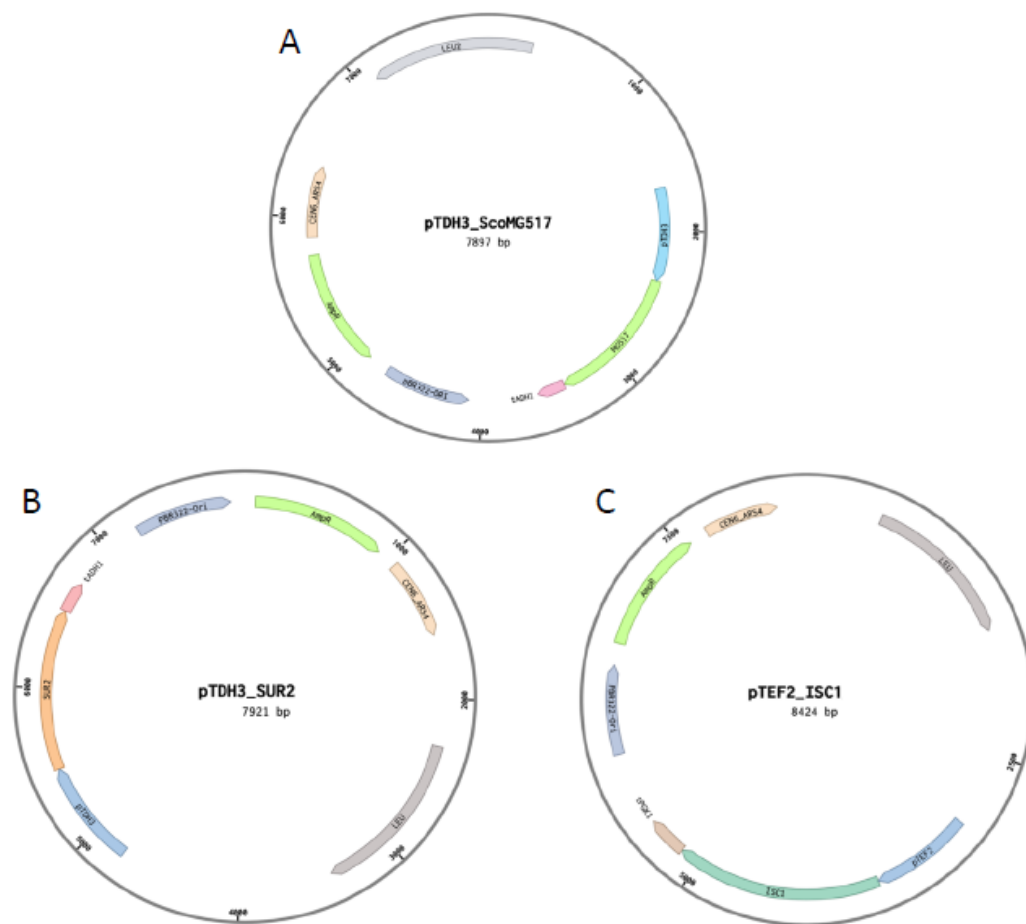


Figure 2.1 Vector map of pTDH3-ScoMG517 (A), pTDH3-SUR2 (B) and pTEF2-ISC1 (C). Vector elements are marked in colored arrows inside the circle of the plasmid that indicate sequence direction. Legend: Light grey: *LEU2*. Light brown: *CEN6-ARS4*. Light green: *Amp^r*. Strong grey: *pB322-Ori*. Pink/red: *tADH1*. Light green (A): *MG517*. Light blue (A): *pTDH3*. Strong blue (B): *pTDH3*. Orange: *SUR2*. Strong blue (C): *pTEF2*. Strong green: *ISC1*. Brown (C): *tPGK1*.



2.1.1.2 Construction of vectors

Vector construction was carried out using CPEC assembly (Figure 2.2). First, vector fragments were needed: backbone and insert/s. These fragments had a 20 bp sequence with homology sequences with the adjacent fragment. Backbones were linearized by PCR and inserts were obtained by amplification with PCR from genomic DNA (extracted from *S. cerevisiae* strain Sc1170).

pTDH3-SUR2

Vector backbone was obtained from pTDH3-ScoMG517 by linearizing the vector without including MG517 gene by PCR. *SUR2* insert would be amplified from genomic DNA. Table 2.1 shows the list of primers and templates used, and fragments obtained from each PCR.

Table 2.1 Table of primers and templates used to amplify each fragment for pTDH3-SUR2 plasmid. Table indicates the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences can be found in Table S3.1 from this chapter's Annexes.

Fragment	Template	Primers	Length of fragment (bp)
pTDH3 backbone	pTDH3-ScoMG517	BB_FS	6871
		BB_RS	
<i>SUR2</i> insert	Genomic DNA	SUR2_F	1050
		SUR2_R	

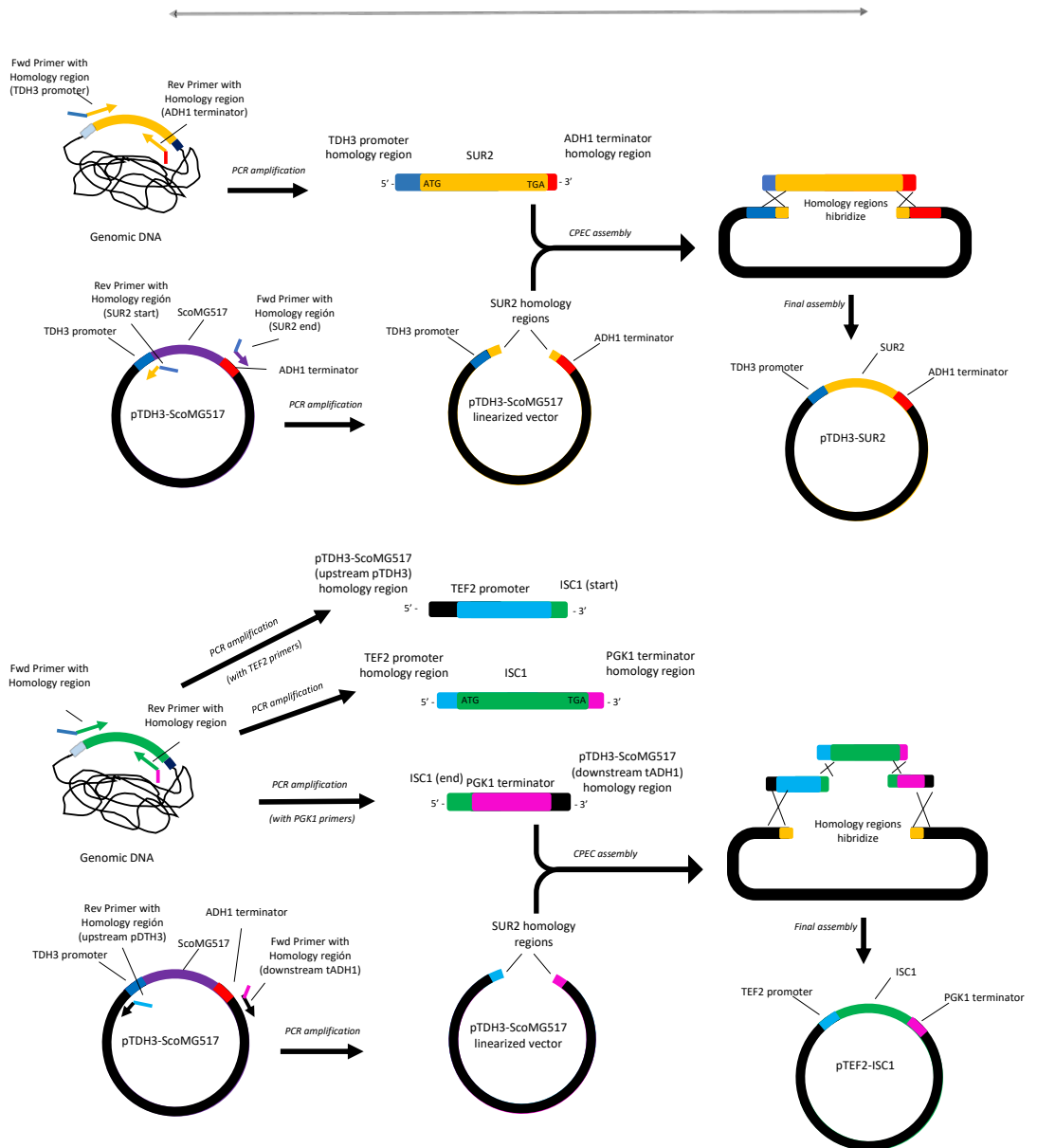


Figure 2.2 CPEC strategy followed for *SUR2* (Top) and *ISC1* (Bottom) cloning. *SUR2*, *ISC1* genes, and *TEF2* promoter and *PGK1* terminator sequences were amplified from genomic DNA. pTDH3-ScoMG517 was used as backbone for the construction. Each fragment was amplified with 20 bp sequences homologous to the adjacent regions.

pTEF2-ISC1

Vector backbone was obtained from pTDH3-ScoMG517 by linearizing the vector without including expression cassette. *TEF2* promoter, *PGK1* terminator and *ISC1* coding sequence, pTEF2 and tPGK1 would be amplified from genomic DNA. Table 2.2 shows the list of primers and templates used, and fragments obtained from each PCR.

Table 2.2 List of primers and templates used to amplify each fragment for pTEF2-ISC1 plasmid. Table indicates the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences can be found in Table S3.1 from this chapter's Annexes.

Fragment	Template	Primers	Length of fragment (bp)
Vector backbone	pTDH3-ScoMG517	BB_FI	5990
		BB_RI	
TEF2 promoter	Genomic DNA	TEF2_F	740
		TEF2_R	
ISC1 insert	Genomic DNA	ISC1_F	1474
		ISC1_R	
PGK1 terminator	Genomic DNA	tPGK1_F	340
		tPGK1_R	

DNA fragments were amplified by PCR using PrimeStar GXL polymerase (TakaraBio) as described in the Material and methods chapter. PCR products were then analyzed by agarose gel electrophoresis. Figure 2.3A shows that vector backbone for *SUR2* plasmid was amplified but backbone for *ISC1* plasmid was poorly amplified (Figure 2.3B). Also, *SUR2* and *ISC1* amplification genes were successfully obtained with the expected bands of 1050 and 1474 bp, respectively. *TEF2* promoter was also successfully amplified, however, *PGK1* terminator was not. *SUR2* gene, *ISC1* vector backbone and tPGK1 fragments were amplified again (Figure 2.4). Although *SUR2* gene still showed low yields of amplification, all fragments were present. PCR products were purified -two PCR reactions of *SUR2* gene were pooled together- and quantified.

Cloning was carried out using CPEC assembly. For each reaction, an equimolar amount of DNA fragments: linearized pTDH3 backbone and *SUR2* gene for the pTDH3-*SUR2* plasmid, and linearized vector backbone for *ISC1*, pTEF2, *ISC1* gene and tPGK1 for the pTEF2-*ISC1* plasmid. Reaction products were used to transform DH5 α *E. coli* cells.

Transformant colonies of pTDH3-*SUR2* were analyzed by Colony PCR (primers pTDH3_fw and tADH1_rv). Expected results for Colony PCR (1931 bp for pTDH3-*SUR2*) are shown in Table 2.3. 12 colonies were analyzed for pTDH3-*SUR2* (Figure 2.6) and 6 of the colonies show the expected fragment of 1931 bp. Since *SUR2* and ScoMG517 were sequences of similar length, it was decided that a further restriction analysis should be made to ensure *SUR2* and not ScoMG517 was present. BsaI was used for

digestion (Table 2.4), and fragments of 5320 and 2601 bp were expected for pTDH3-SUR2, but pTDH3-ScoMG517 would show a unique fragment of 7897 bp. Figure 2.7 shows that colonies 2, 3, 4 and 5 matched the expected fragment sizes.

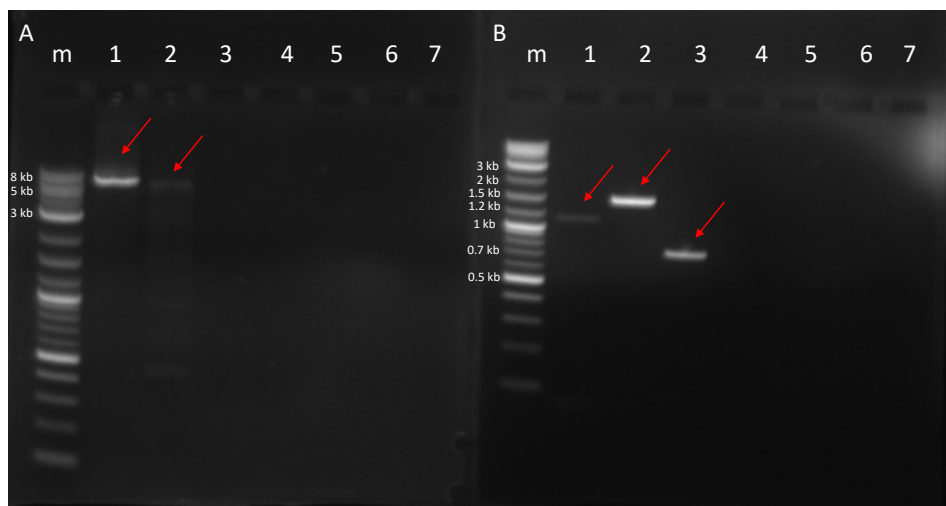


Figure 2.3 Agarose gel electrophoresis results of PCR for pTDH3-SUR2 and pTEF2-ISC1 vector synthesis. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. A) Lane 1: pTDH3-SUR2 backbone. Lane 2: pTEF2-ISC1 backbone. B) Lane 1: *SUR2* gene. Lane 2: *ISC1* gene. Lane 3: *TEF2* promoter. Lane 4: *PGK1* terminator.

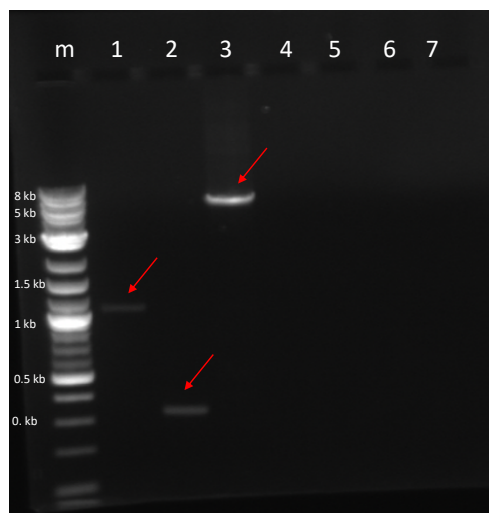


Figure 2.4 Agarose gel electrophoresis with different amplifications. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. Lane 1: *SUR2* gene. Lane 2: *PGK1* terminator. Lane 3: *ISC1* backbone.

As for pTEF2-ISC1 plasmid, Colony PCR showed no amplification in any of the colonies. This might be due to the long amplification fragment that was expected (2355 bp). Usually, for DNA fragments longer than 2 kb, Colony PCR Taq polymerase is not able to amplify them. Therefore, it was decided to extract plasmid from 8 of the colonies and analyze the vector by restriction analysis (Figure 2.5). The restriction with EcoRI and HindIII enzyme on plasmid pTEF2-ISC1 yields 5597, 1365 and 1063 bp fragments (Table 2.4). Figure 2.6 showed that 6 out of 8 colonies matched the expected fragments. Two different patterns appeared (indicated with arrows in Figure 2.8), thus, both colonies were sent for sequencing analysis.

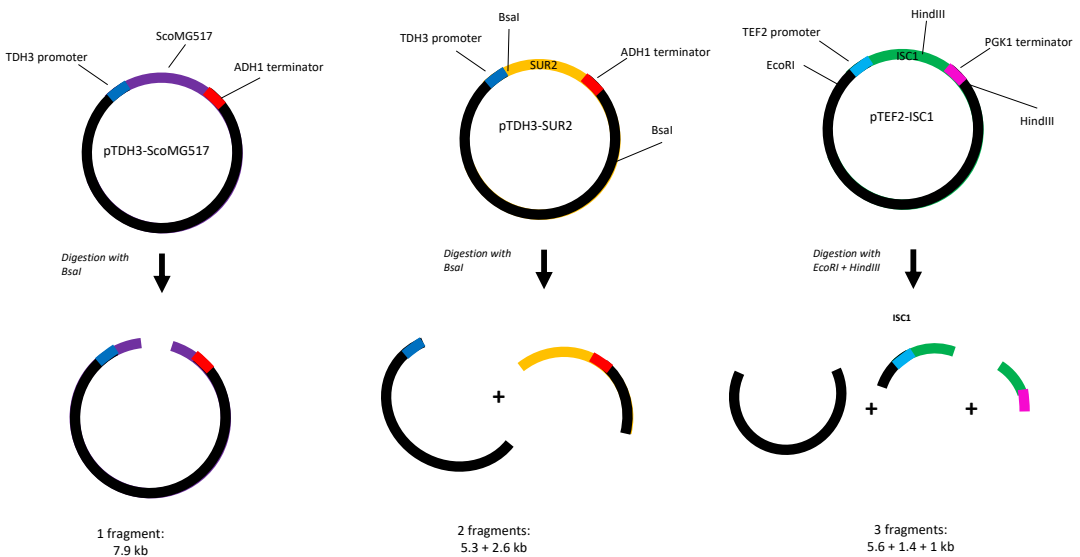


Figure 2.5 Digestion analysis approach of pTDH3-SUR2 and pTEF1-ISC1 vectors. pTDH3-ScoMG517 was used as control and digested with BsaI with pTDH22-SUR2. pTEF2-ISC1 was digested with EcoRI and HindIII. Each plasmid was expected to yield different fragments, indicated in Table 2.4.

Table 2.3 Table of primers used in Colony PCR with the template (colonies) and the expected amplified fragment sizes. Primer sequences can be found in Table S3.1 from this chapter's Annexes.

Fragment	Template	Primers	Length of fragment (bp)
SUR2	Colony 1-12 with pTDH3-SUR2	pTDH3_fw tADH1_rv	1931

Table 2.4 restriction enzymes and expected fragments after restriction analysis for pTDH3-SUR2 and pTEF2-ISC1 plasmids.

Plasmid	Restriction enzymes	Length of fragments (bp)
pTDH3-SUR2	BsaI	5320+2601
pTEF2-ISC1	EcoRI + HindIII	5597+1365+1063

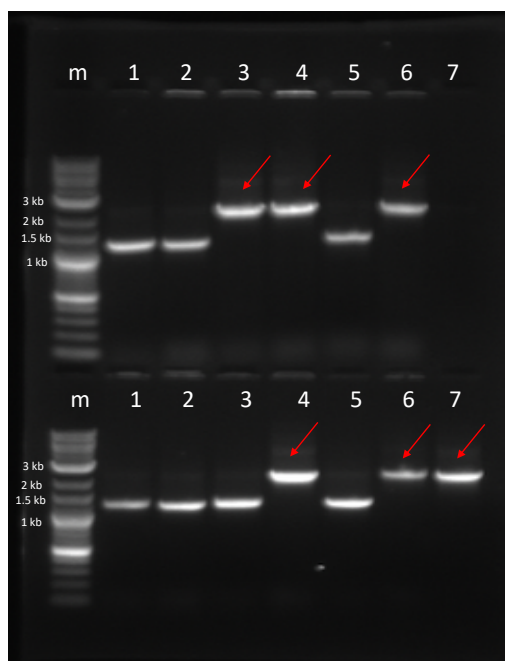


Figure 2.6 Agarose gel electrophoresis results of Colony PCR for pTDH3-SUR2. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. Top: Lanes 1-6: Transformed colonies 1 to 6 of product of pTDH3-SUR2 assembly. Bottom: Lanes 1-7: Transformed colonies 7 to 13 of product of pTDH3-SUR2 assembly.

Finally, for each plasmid, two colonies that matched the expected results after restriction analysis were sent for sequencing analysis. Expected sequences from designed constructs, obtained from *in silico* cloning, were compared to the sequencing results. For pTDH3-SUR2, only *SUR2* gene was sequenced, but for pTEF2-ISC1, the whole construct was sequenced, including *TEF2* promoter, *ISC1* gene and *PGK1* terminator. Colonies that matched with 100% homology with the sequencing results were successful and were used for following experiments.

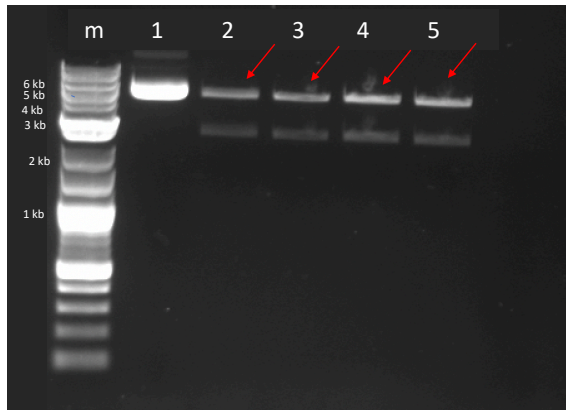


Figure 2.7 Agarose gel electrophoresis results of restriction analysis of positive colonies of transformation with **pTDH3-SUR2**. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. Lane 1: non-digested plasmid from Colony 3. Lanes 2-5: digestion product of plasmids from colonies 3, 4, 5 and 9, in that order.

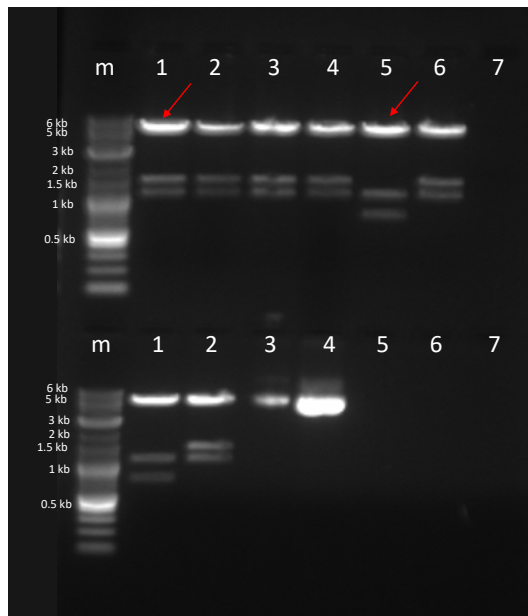


Figure 2.8 Agarose gel electrophoresis results of restriction analysis of transformed colonies with **pTEF2-ISC1**. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. Top: Lanes 1-6: Transformed colonies 1 to 6 of product of pTEF2-ISC1 synthesis. Bottom: Lanes 1-4: Transformed colonies 7 to 10 of product of pTEF2-ISC1 synthesis.



2.1.2 Analysis of pTDH3-SUR2 and pTEF2-ISC1 strains

After successfully constructing pTDH3-SUR2 and pTEF2-ISC1 expression vectors and using them to transform the yeast strain, the phenotype of the resulting strain was analyzed by qPCR for *SUR2* and *ISC1* transcription levels and TLC for ceramide production. In addition, two new plasmids were constructed replacing *SUR2* or *ISC1* genes by mCherry gene in pTDH3-SUR2 and pTEF2-ISC1 plasmids. The mCherry gene encodes for a red fluorescent protein. This new vectors, pTDH3-mCherry and pTEF2-mCherry were used as controls for the phenotype analysis of the new strains.

2.1.2.1 qPCR analysis

Quantitative PCR (qPCR) or real-time PCR (RT-PCR) is based on the comparison of the threshold cycle (Ct) value, which is the number of cycles when fluorescence starts to be detected. Therefore, a low Ct value indicates high concentration of the target gene mRNA in the sample. For qPCR, mRNA is first converted to cDNA with reverse transcriptase (RT). This cDNA is then used as template in the qPCR reaction and, since reverse transcriptase does not amplify DNA but converts all mRNA into DNA, it will indicate the concentration of the mRNA in the sample. It is important to use housekeeping genes when analyzing gene expression. These genes are expressed continuously and with no expected variations, therefore, they are expected to show minimal differences in expression levels between samples and experiments. In qPCR, they are used to normalize expression values between samples and experiments. The gene used in this analysis was *ACT1*, an already tested housekeeping gene¹⁰⁶.

For RNA extraction and qPCR analysis, yeast strain YM4271 was transformed with both plasmids. It was used this strain instead of strain RH6082 due to a complication when transforming the last strain: control plasmids with fluorescent protein did not show colored colonies, which were observed before. For that reason, it was decided to use a new strain. Since qPCR protocol for yeast was optimized for first time in the Laboratory of Biochemistry, it was not considered important to use a different strain to implement the protocol.

For qPCR analysis, extraction of high-quality RNA is a requirement. The RNA extraction was carried out treating cells with lyticase (commercial lysing enzyme from *Arthrobacter luteus* purchased from Merck) followed by TRIzol (commercial reagent for ARN isolation from ThermoFisher), the protocol is described in Materials and methods chapter. After treatment, the mixture was then used to follow the RNeasy Mini Kit extraction protocol. Extracted RNA was then treated with Dnase I to remove genomic DNA that may contaminate the sample and would interfere with qPCR. RNA is quantified, RNA concentration was critical, therefore, after quantification, samples with low levels of RNA (less than 5 ng/mL) were discarded, and, if needed, a new culture was set up to repeat the experiment.

Extracted RNA was then converted into cDNA using reverse transcriptase (RT), following protocol from First-Standard cDNA Synthesis kit (NZYTech). As a control, a sample with no RT was also included and

the protocol of cDNA synthesis was also followed. The control was carried out adding mQ water instead of RT (Control was named as NO RT control). Another control included was with no template: all components were added but with no extracted RNA (NO T control). In both controls, no signal is expected to appear, since NO RT control would have no cDNA and RNA would be digested with Rnase, and NO T control would not contain the template for the reverse transcriptase to convert into cDNA. For *SUR2* expression analysis, designed primers were SUR2_RT_N_Fw + SUR2_RT_N_Rv (see Table S3.1 from this chapter's Annexes). For *ISC*, primers ISC1_RT_Fw + ISC1_RT_Rv (see Table S3.1 from this chapter's Annexes). And also, primers ACT1_RT_N_Fw + ACT1_RT_N_Rv to analyze the housekeeping gene *ACT1*. Figure 2.9 shows the scheme of qPCR analysis carried out for *SUR2* and *ISC1* expression analysis.

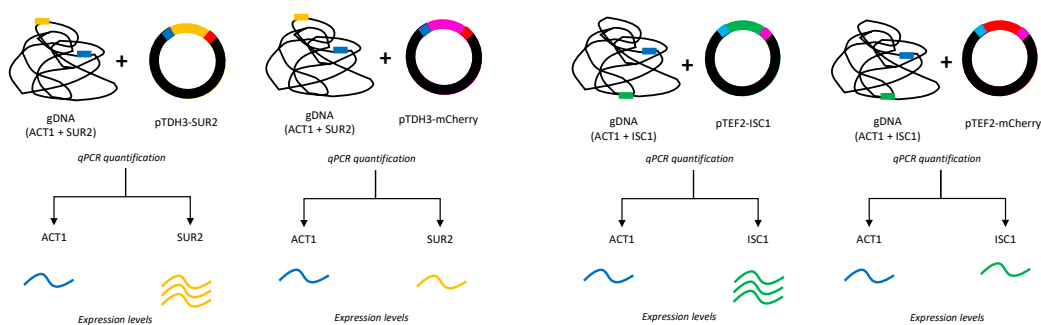


Figure 2.9 qPCR scheme for *SUR2* and *ISC1* expression analysis. Each cell contains its genomic DNA (gDNA) and a plasmid. Genomic DNA contains copies of the housekeeping gene *ACT1*, and *SUR2* and *ISC1* genes. These genes are expressed at a basal level. *ACT1* allows to normalize the basal levels and compare the strains overexpressing *SUR2* or *ISC1* to the basal expression of these genes. Strains without the plasmid containing *SUR2* or *ISC1*, only the basal level of expression of these genes will be detected. Normalizing *SUR2/ISC1* expression levels by the expression levels of *ACT1* (which is considered as the same in all conditions), *SUR2/ISC1* overexpression of strains with pTDH3-*SUR2* or pTEF2-*ISC1* can be compared.

New expression vectors using backbones of pTDH3-*SUR2* and pTEF2-*ISC1* containing mCherry (fluorescent protein) gene replacing *SUR2* or *ISC1* genes, respectively, were used as controls and analyzed. The pTDH3-mCherry and pTEF2-mCherry vectors were already constructed in the laboratory and used as a control for expression and new primers (mCherry_RT_fw and mCherry_RT_rv, see Table S3.1 from this chapter's Annexes) were also designed for qPCR analysis of mCherry.

Gene expression analysis was carried out by culturing the YM4271 yeast strains with: pTDH3-*SUR2*, pTEF2-*ISC1*, pTDH3-mCherry and pTEF2-mCherry. RNA extraction was carried out as described previously, and final cDNA was obtained and used for the qPCR analysis. Three colonies were analyzed for each strain, and two technical replicas for each colony was also applied. Biological replicas and technical replicas data for each strain were analyzed. The mean of the expression level for each strain was calculated and analyzed by t-test.

Figure 2.10 shows that pTDH3-SUR2 strain has higher expression of *SUR2* than the corresponding endogenous *SUR2* in the pTDH3-mCherry yeast strain. The average *SUR2/ACT1* fold between all biological replica showed that the fold change of expression level was 144 ± 36 in the strain overexpressing *SUR2*, confirming the overexpression of this gene.

Results from *ISC1* overexpression showed that a signal was always detected even in the controls (NO RT and NO T). These results were discarded since that would indicate dimerization of the primers.

Finally, a new set of primers for *ISC1* analysis should be designed in future work and used in the qPCR. Once overexpression of *ISC1* is confirmed, the phenotype of the new strains should be analyzed. Ceramide production in each strain should be determined and characterized to determine differences in ceramide overall content and ceramide structure. Due to the difficulty of separating and quantifying different structures of ceramide from a culture sample, use of HPLC-MS/MS or ESI-MS/MS methods would be more suitable than TLC to obtain results with higher accuracy and sensitivity.

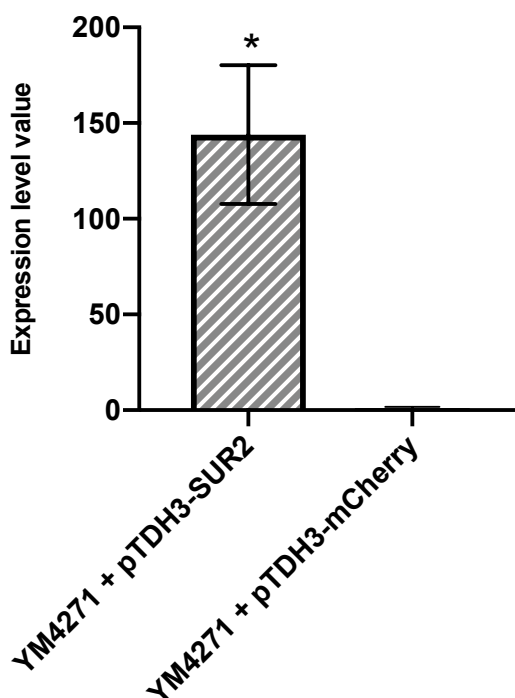


Figure 2.10 Graph bar of the normalized qPCR results of *SUR2* expression. Biological replica and technical replica data of each strain (YM4271 + pTDH3-SUR2 and YM4271 + pTDH3-mCherry) were merged. Y-axis represents expression level value. * indicates difference in expression level after t-test analysis with 95% significance with (p-value: 0.0024).

2.1.3 *SCS7* gene knock-out

The third approach on yeast metabolic engineering was to delete *SCS7* gene to avoid the hydroxylation of the fatty acid chain of the ceramide. Gene deletion will be carried out using CRISPR/Cas9 technology. To knock-out a gene using CRISPR/Cas9 tool, three constructs are required: (1) a plasmid for the expression of Cas9 nuclease, (2) a plasmid that will produce the guide-RNA for genome-targeting, and (3) a donor DNA that will replace the target gene by the homologous recombination repair system.

Cas9 expression plasmid

Plasmid with Cas9 (Figure 2.11) is the p414-TEF-Cas9 that was obtained by Prof. Marjan de Mey group from the Center for Synthetic Biology, University of Gent. P414-TEF-Cas9 is a low copy number plasmid (with CEN6/ARS4 origin), with Cas9 gene under the control of *TEF1* promoter and *CYC1* terminator. Yeast selection marker is *TRP1*, which confers growth under tryptophan deprivation. The *E. coli* origin of replication is pBR322 and selection marker is ampicillin.

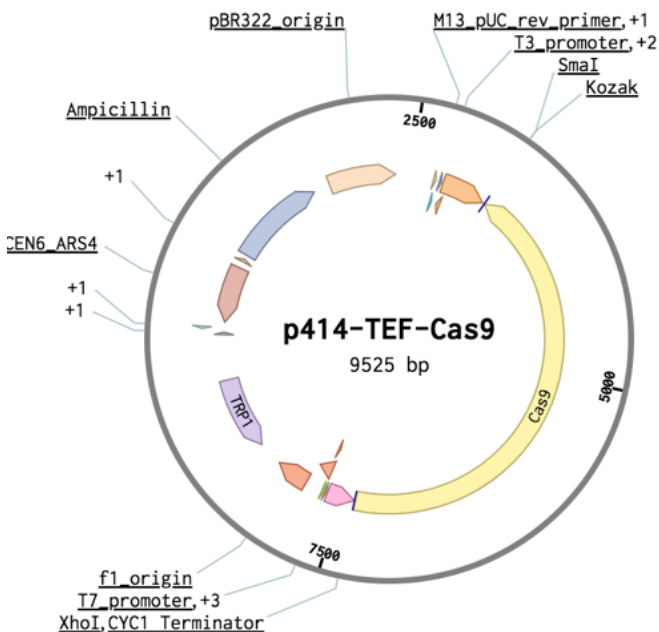


Figure 2.11 Vector map of p414-TEF-Cas9 plasmid. Vector elements are marked in colored arrows in the center, the name of *the* vector and its size. Legend: Yellow, Cas9. Pink: *TEF1* promoter. Orange (downstream of Cas9): *CYC1* terminator. Light brown: pBR22 origin. Blue: Ampicillin resistance. Brown: CEN6-ARS4. Purple: *TRP1* selection marker. Orange (downstream TRP1): *f1* origin.

Guide-RNA plasmid

Guide-RNA is a 20 bp sequence that is required to target the gene to be removed. The sequence is followed by an unchanged sequence that contains spacer, a JAD designed RNA 3' UTR, and gRNA structural component sequence. This construct is under the control of SNR52 promoter and SUP4 terminator. The plasmid has a selectable marker different from Cas9 expression plasmid (tryptophan auxotrophy) which is URA3, for uracil auxotrophy. It also contains a 2 μ origin, which indicates it is a high-copy number plasmid, necessary to ensure Cas9 has the guide RNA to be targeted to the specific region of the genome.

Guide-RNA plasmid (gRNA) was constructed from TES1-gRNA plasmid also from Prof. Marjan de Mey and was used as scaffold to insert the 20 bp gRNA sequence. This sequence (5'-TGGGATTTGGCGTCACATCC-3') targets the Cas9 to SCS7 gene (89 bp upstream of SCS7 stop codon), promoting a cut in the genomic DNA, which in turn activates the Homology-Directed Repair (HDR) system. Since the desired sequence to insert is of such a short length (20 bp), replacing the previous 20 bp with the insertion of the new gRNA sequence is done by splitting and linearizing the vector in two fragments of an approximately length of 3 kb (Figure 2.12). Linearization is performed with primers that introduce homologous regions with the new 20 bp sequence for gRNA. When PCR is finished, both fragments of the vector include the gRNA sequence and, after CPEC assembly, they will hybridize to form the final plasmid with the SCS7 gRNA sequence. Table 2.5 includes the primers designed for the construction of this vector, and Figure 2.14 shows the vector map of the final vector.

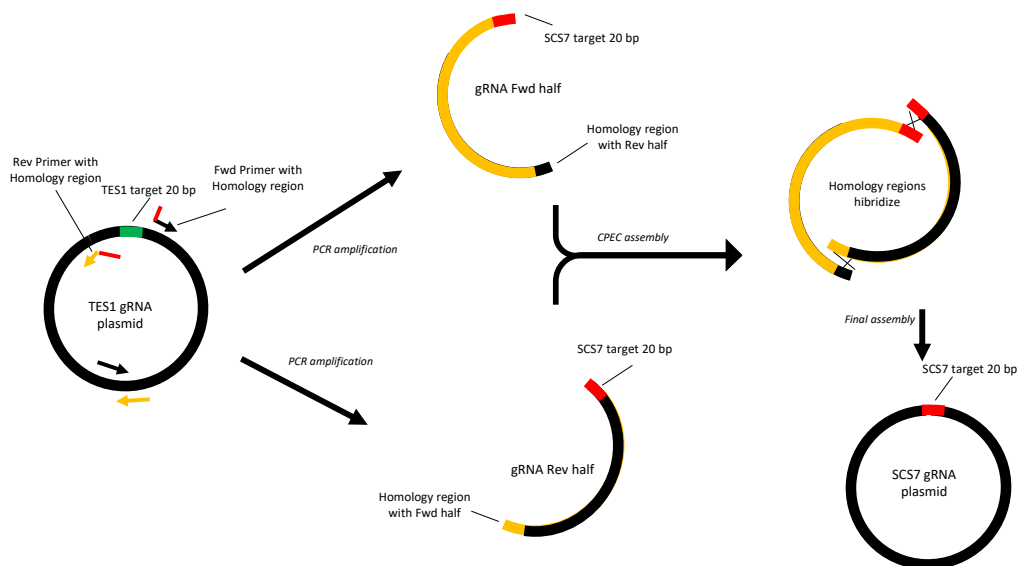


Figure 2.12 CPEC strategy followed for SCS7 gRNA plasmid construction. TES1 gRNA plasmid is used as backbone to insert the 20 bp sequence that target SCS7. Vector is amplified in two halves (Fwd and Rev), using primers with an overlapping sequence corresponding to the 20 bp of SCS7 gRNA. Assembly will be carried out hybridizing gRNA sequence and a sequence that was amplified with the primers at 'half' vector.

Donor DNA plasmid

Donor DNA is a linear DNA that is used with gRNA plasmid to co-transform the yeast expressing Cas9. This DNA is constructed from the pET22b-VcCDA-Strep plasmid available in the Biochemistry laboratory, which contains the pET22b plasmid and the VcCDA gene (of interest in another project). In this project, pET22b-VcCDA-Strep is used as template to amplify and obtain linear donor DNA by PCR. Although donor DNA is a linear sequence that is introduced into the strain to be modified, it is first cloned into a plasmid (pET22b-VcCDA) that then will be used as template for the amplification of the linearized fragment. Donor DNA contains 500 bp of the upstream of the start codon of *SCS7*, which is fused to the 500 bp downstream of the stop codon of the gene (Figure 2.13). This 1 kb fragment will replace the actual gene and, since it does not contain the coding sequence of the gene, will result in the deletion of *SCS7* from the genome.

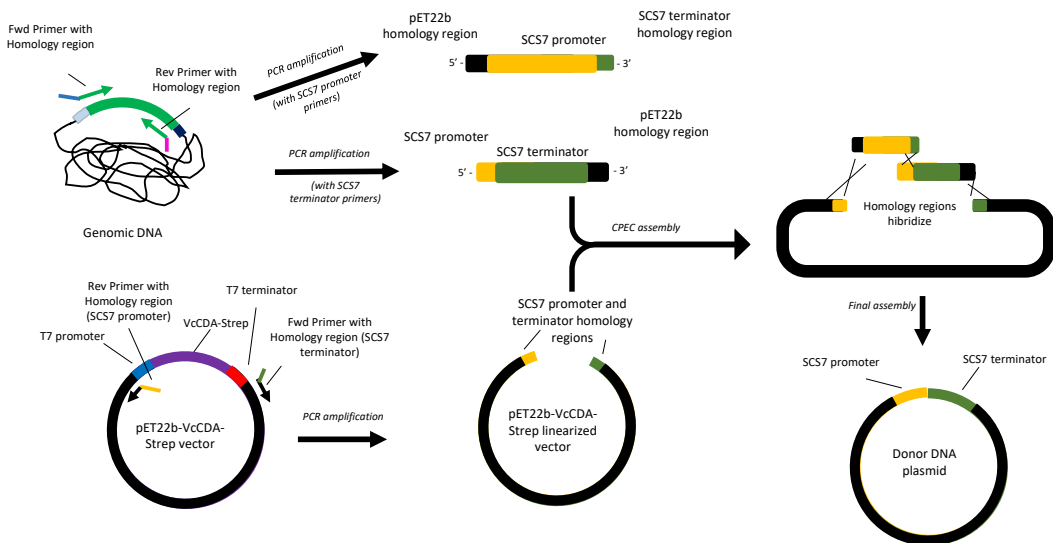


Figure 2.13 CPEC strategy followed for donor DNA plasmid construction. 500 bp upstream of *SCS7* start codon (*SCS7* promoter, yellow) and 500 bp downstream of Stop codon (*SCS7* terminator, green) sequences are amplified from gDNA. pET22b-VcCDA-Strep was used as backbone and linearized, removing T7 promoter, VcCDA and T7 terminator. Each fragment contains 20bp sequence with homology to the adjacent fragments.

For the construction of the gene, pET22b-VcCDA-Strep was used as vector backbone. The gene construct -promoter, gene, and terminator- of the vector was removed and replaced by the donor DNA sequence. Both upstream and downstream regions of *SCS7* were obtained from genomic DNA. The vector also has ampicillin selection and a bacterial origin for a high-copy number plasmid.

Plasmid assembly was done following CPEC protocol, therefore, all fragments contained homologous regions for hybridization in the assembly. Table 2.5 contains the design of the primers used for the assembly, and Figure 2.14 shows the vector map of this plasmid.



Table 2.5 List of primer design used for SCS7 knockout vectors. In the list it is specified the name of the gene, the fragment obtained after amplification, the direction of the primer and the homologous and binding regions of the primers. Primer sequences can be found in Table S3.1 from this chapter's Annexes.

Redirection	Fragment	Primer	5' homologous region	3' binding region
SCS7 gRNA plasmid	gRNA fw half	Forward	gRNA sequence	JAD designed RNA 3' UTR
		Reverse		half vector
	gRNA rv half	Forward		half vector
		Reverse	gRNA sequence	SNR52p
SCS7 donor DNA plasmid	pET22b vector backbone	Forward	SCS7 downstream sequence	20 bp downstream of pET22b terminator
		Reverse	SCS7 upstream sequence	20 bp upstream of pET22b promoter
	SCS7 500 bp upstream	Forward	20 bp upstream of pET22b promoter	480-500 bp upstream start codon
		Reverse	0-20 bp downstream SCS7 stop codon	0-20 bp upstream SCS7 start codon
	SCS7 500 bp downstream	Forward	0-20 bp upstream SCS7 start codon	0-20 bp downstream SCS7 stop codon
		Reverse	20 bp downstream of pET22b terminator	480-500 bp downstream SCS7 stop codon

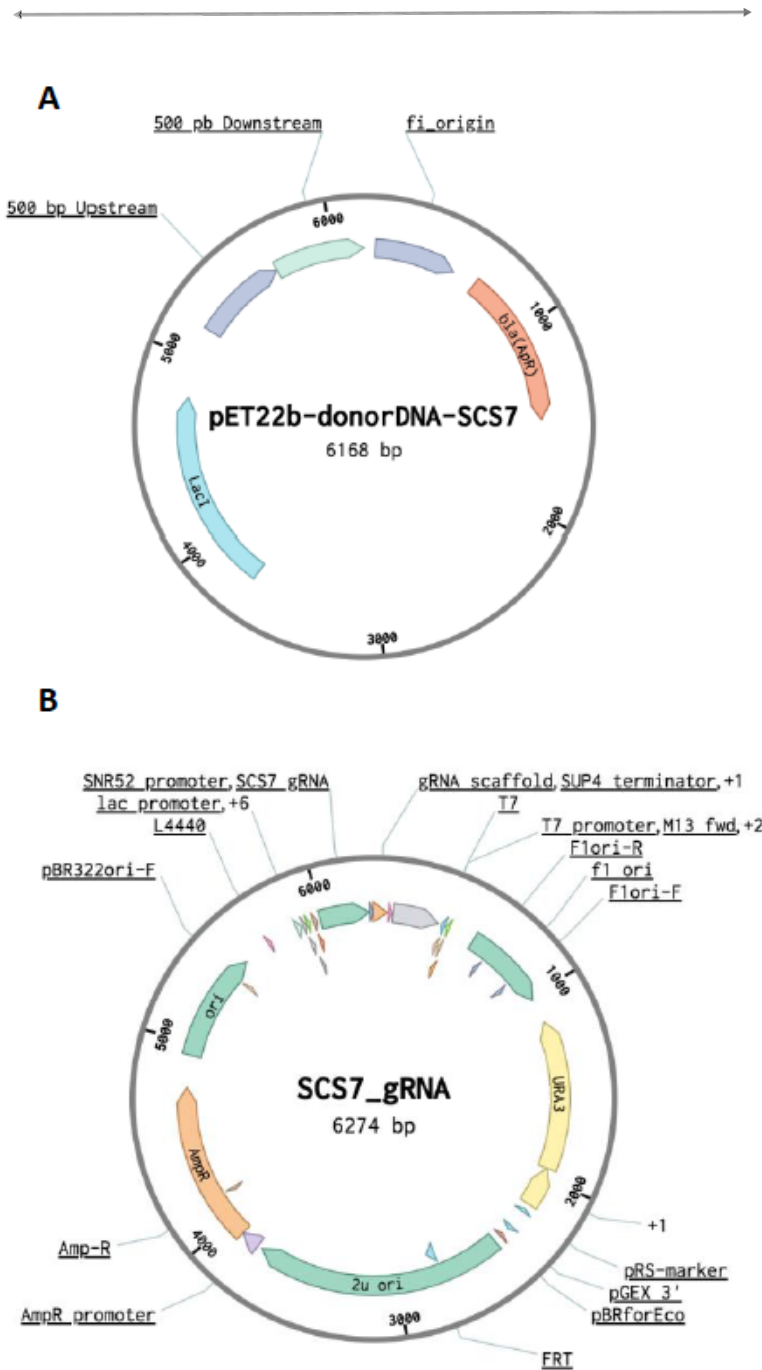


Figure 2.14 Vector map of SCS7 donor DNA plasmid (pET22b-donorDNA-SCS7) (A) and SCS7 gRNA plasmid (B). Vector elements are marked in colored arrows inside the circle of the plasmid that indicate sequence direction. In the center, the name of the vector and its size. Legend: A) donor DNA. Light blue: LacI. Strong blue: SCS7 500 bp upstream of start codon. Green: SCS7 500 bp downstream stop codon. Strong blue (downstream donor DNA.): f1 origin. Orange: Ampicillin resistance. B) SCS7 gRNA. Orange: Ampicillin resistance. Green: pBR322 origin. Short sequences (in that order): SNR52 promoter, SCS7 gRNA, gRNA scaffold, SUP4 terminator. Green: f1 origin. Yellow: URA selection marker. Green: 2 μ origin.



2.1.3.1 Construction of knock-out vectors

Guide-RNA plasmid

As it has been described previously, gRNA plasmid was split in two halves that included homologous regions. Each half had homologous regions, and one of them contained the gRNA sequence. PCR amplification was done using primers listed on Table 2.6. Once fragments were amplified and checked with the expected 3.5 and 2.9 kb bands (see Figure 2.15), they were purified and quantified. CPEC assembly was carried out with an equimolar amount of each fragment. Assembly product was used to transform DH5 α *E. coli* cells.

Due to both plasmids -TES1 gRNA and new SCS7 gRNA plasmids- being the same size, it was impossible to distinguish between the two of them by either colony PCR or restriction analysis. Therefore, two colonies after transformation of DH5 α *E. coli* with PCR product were cultured for plasmid extraction and analyzed by Sanger sequencing, which confirmed the 20 bp insertion of SCS7 gRNA sequence.

Donor DNA plasmid

Although donor DNA is a linear sequence that is introduced into the strain to be modified, it is first cloned into a plasmid that then will be used as template for the amplification of the linearized fragment. This vector is constructed with three fragments obtained with primers listed on Table 2.6: vector backbone, 500 bp of SCS7 upstream region, and 500 bp of SCS7 downstream region after stop codon. Figure 2.15 shows the results of the amplification. After amplified fragments matched the expected sizes of 500 bp for donor DNA and 5.4 kb for vector backbone, they were purified and quantified. CPEC assembly was done with an equimolar amount of each fragment. *E. coli* DH5 α cells were transformed with assembly product and colonies were analyzed by colony PCR using primers SCS7p_fw and SCS7t_rv (see sequence in Table S3.1 from this chapter's Annexes), which amplified a fragment of approximately 1.3 kb. Once colony PCR results showed positive colonies that matched the expected size (Figure 2.16), one of the samples was sent for sequencing analysis. Sanger sequencing results confirmed the successful construction of donor DNA plasmid with the expected fragments of 500 bp upstream and 500 bp downstream of SCS start and stop codons, respectively.

Table 2.6 List of primers and templates used to amplify each fragment for pSF-BF3149 plasmids. Table indicates the name of the final vector to be constructed, the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. SCS7p and SCS7t are references for the 500 bp region that would be the promoter and terminator of SCS7.

Vector	Fragment	Template	Primers	Length of fragment (bp)
SCS7 gRNA plasmid	gRNA fw half	TES1-gRNA plasmid	gRNA_SCS7_fw gRNA_half_rv	3503
	gRNA rv half	TES1-gRNA plasmid	gRNA_half_fw gRNA_SCS7_rv	2859
SCS7 donor DNA plasmid	pET22b vector backbone	pET22b-VcCDA-Strep plasmid	pET22b_donorBB_fw pET22b_donorBB_rv	5240
	SCS7 500 bp upstream	Genomic DNA	SCS7p_fw SCS7p_rv	540
	SCS7 500 bp downstream	Genomic DNA	SCS7t_fw SCS7t_rv	540

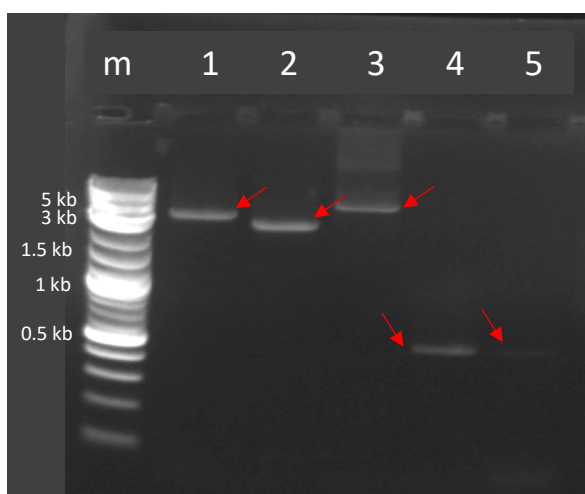


Figure 2.15 Agarose gel electrophoresis results of PCR amplification of SCS7 gRNA and donor DNA plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows indicate the fragments that matched the expected sizes. Lane 1: gRNA fw half. Lane 2: gRNA rv half. Lane 3: pET22b vector backbone. Lane 4: SCS7 500 bp upstream. Lane 5: SCS7 500 bp downstream.

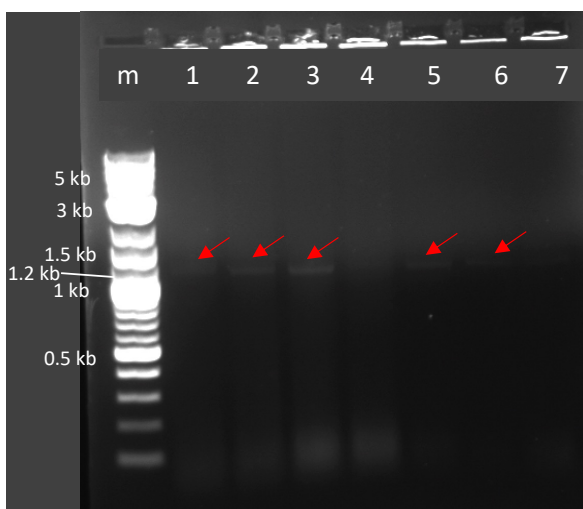


Figure 2.16 Agarose gel electrophoresis results of Colony PCR for SCS7 donor DNA plasmid. *m* contains DNA lane marker (1 kb plus DNA ladder from NEB). Arrows show fragments that match the expected results. Lanes 1-7: Colonies 1-7 transformed with CPEC assembly product of SCS7 donor DNA plasmid.

2.1.3.2 Yeast transformation and results

After all plasmids were successfully constructed and analyzed by Sanger sequencing, gene knockout was performed.

First, *SCS7* donor DNA was amplified using the donor DNA plasmid. PCR product was analyzed, matching the 1 kb size that was expected, and finally it was purified. RH6082 *S. cerevisiae* strain was co-transformed with linear donor DNA, *SCS7* gRNA plasmid and p414-TEF-Cas9 plasmid. Yeast transformation was carried out following the modified Gietz protocol¹⁰⁷. Transformed cells were plated on selective plates for each selection marker: SD (Synthetic Defined medium) -Ura, SD -Trp and SD -Ura -Trp. Plates with only one selection marker were used as control for positive transformation of each plasmid.

Colonies from double selection plate (SD -Ura -Trp) were analyzed by colony PCR. If gene deletion was successful, a band of 1 kb would appear, otherwise, the DNA fragment containing the *SCS7* gene would be 2.2 kb (Figure 2.17). However, after agarose gel electrophoresis, no bands appeared. This could mean that *SCS7* knockout was not successful.

Although new transformation experiments were performed, no colonies appeared. Probably the plasmids were not successfully assimilated, and cells could not grow on selective medium. New experiments are required in the future in order to achieve the *SCS7* knockout.

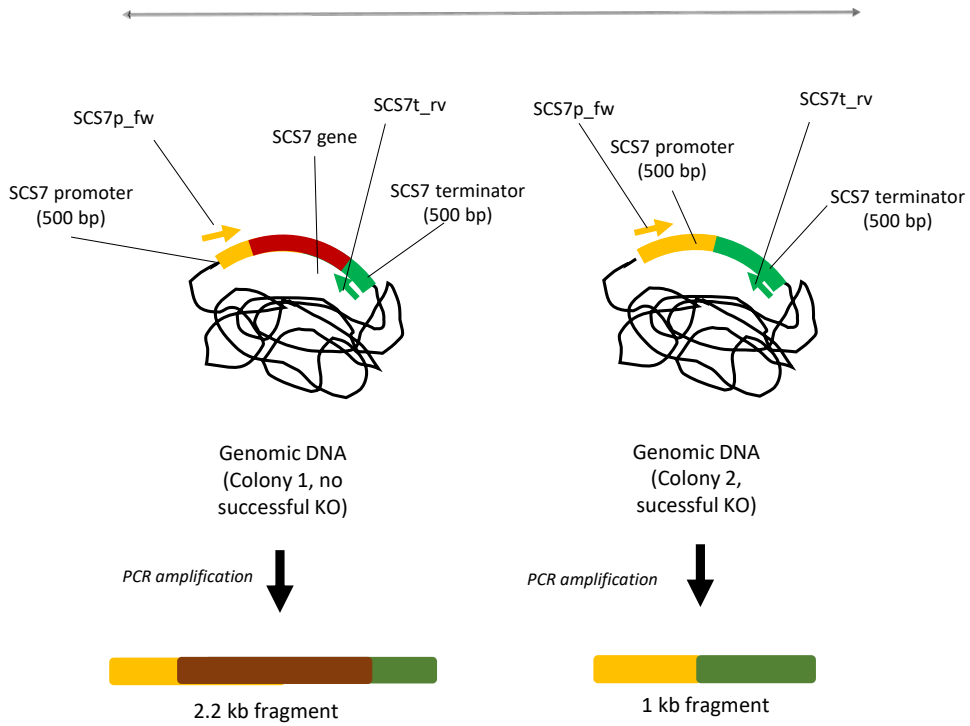


Figure 2.17 Colony PCR analysis of *SCS7* knockout colonies. Colonies after transformation are analyzed by Colony PCR using primers used for *SCS7* donor DNA plasmid construction (*SCS7p_fw* and *SCS7t_rv*). If the KO has not been successful, primers will hybridize and amplify the 500 bp of promoter region, *SCS7* and 500 bp of terminator region (2.2 kb). Otherwise, if KO is successful, *SCS7* gene has been removed, and only 500 bp of promoter and 500 bp of terminator regions will be amplified (1 kb).



2.2 Summary

The design of *SUR2* and *ISC1* expression vectors was achieved, although some difficulties appeared in the construction of both vectors regarding the amplification of the vectors. However, the fragments for their assembly were finally amplified. After construction, both vectors showed the expected sequences, and transformation was also successful.

Expression analysis by qPCR for *SUR2* gene was successful, and a fold-change of 144 was achieved. Even so, qPCR analysis for *ISC1* was unsuccessful and must be repeated since results indicated that the primers used were dimerizing. Therefore, *ISC1* expression should be analyzed again using a new set of primers for qPCR.

The plasmids for *SCS7* gene knockout were successfully synthesized, however, after transformation with the plasmids and donor DNA, no positive colonies were obtained. It remains for future work to carry out the knockout of *SCS7* gene.



2.3 Annexes Chapter 2

Table S2.1 List of primers used in Chapter 2 and their sequences. Table indicates name of the primer as mentioned in the chapter, sequence, and length (Lngh). Underlined nucleotides indicate the added homology region for assembly.

Name	Sequence	Lngh
BB_FS	<u>AAGTAAAGAAAGAGAAATAAACTCTCGAGGCGAATTTCTTATG</u>	43
BB_RS	<u>GCATTCGATGTTACGTTCA</u> TTTTGTTTGTATGTGTGTTTATTC	45
SUR2_F	<u>AACACACATAAACAAACAAAATGAACGTAACATCGAATG</u>	39
SUR2_R	<u>AAGAAATTCGCCTCGAGAGT</u> TTATTTCTCTTTCTTACTTCATTTTCAG	49
BB_FI	<u>TTTCGAGAAGGATATTATTTCCGAGCAAATGCCTGCAAATCG</u>	42
BB_RI	<u>CATTGATCTTGACCTATCAA</u> ACTAGTAGACTGAATTCGAGTGAG	45
TEF2_F	<u>TCGGAATTCAGTCTACTAGT</u> TTGATAGGTCAAGATCAATGTAAAC	45
TEF2_R	<u>CTGTCTTTTCTGTTGTACATG</u> TTTAGTTAATTATAGTTGTTGAC	45
ISC1_F	<u>CGAACTATAATTA</u> ACTAAACATGTACAACAGAAAAGACAGAGATG	45
ISC1_R	<u>CGATTTCAATTC</u> AATTCATTTCTCGCTCAAGAAAGTTTG	44
tPGK1_F	<u>CTTTCTTGAGCGAGAAATGA</u> ATTGAATTGAATTGAAATCGATAGATC	47
tPGK1_R	<u>ATTTGCAGGCATTTGCTCGG</u> AAATAATATCCTTCTCGAAAG	41
pTDH3_fw	CTGAAATTATCCCTACTTGAC	23
tADH1_rv	TGACCTACAGGAAAGAGTTAC	21
SUR2_RTN_Fw	GGAATGTCCGCTTTGAAGAT	20
SUR2_RTN_Rv	TTCATGGTGGCAAAGGTA	20
ISC1_RT_Fw	GTTCCGTGGCGACTGGTACG	20
ISC1_RT_Rv	CACCCTGCTTGGCGTATGGG	20
ACT1_RTN_Fw	GTTACGTGCCTTGGACTTC	20
ACT1_RTN_Rv	CGGACATCGACATCACACTT	20
mCherry_RT_Fw	AAGCTGTCCTTCCCGAGGG	20
mCherry_RT_Rv	AGTCCTGGGTCACGGTCACC	20
gRNA_SCS7_fw	<u>TGGGATTTGGCGT</u> CACATCCGTTTTAGAGCTAGAAATAGCAAG	43
gRNA_half_rv	CGAGAGCGCTATTTTCTAA	20
gRNA_half_fw	GCATTATAGAGCGCACAAAGG	21
gRNA_SCS7_rv	<u>GGATGTGACGCCAAATCC</u> CAGATCATTTATCTTTCACTGCGGAGAAG	47
pET22b_donorBB_fw	<u>CCTCATACCTAATCATTTCT</u> TCTAAACGGGTCTTGAGGG	39
pET22b_donorBB_rv	<u>TGTTGTTCTTTGTCGTGGG</u> CATTTTCGCGGGATCGAGATC	39
SCS7p_fw	GCCCACGACAAAGAACAAC	19
SCS7p_rv	<u>TGGACGAGGCTGACCAATA</u> ACTTCGTAAGTTTACTAGTTTACTTACC	43

SCS7t_fw	<u>AAGCTAAA</u> ACTAGTACGAAGTTATTGGTCAGCCTCGTC	38
SCS7t_rv	AGAAATGATTAGGTATGAGGTTACTC	25

Table S2.2 *SUR2* gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	<i>SUR2</i>		
GenBank Accession number	NM_001180605		
DNA sequence	Length	1050	bp
<p>ATGAACGTAACATCGAATGCAACTGCAGCCGGTTCCTTCCACTAGCATTGGTCTCAAGACCTCATTGGGTTTATGCACTATGCAAGGCCCTGCCATTAATTTACGCCCCAAGGAATCCTTGCTGCCGAAATGAGTGATGGTGTGCTGGCCTTGTTGCGCCGGTGTGGCTACTGGCGTGTCTGGTATATCCATGTAATAGACACTTCCATCTGGCTGAGAAGTACA GAATTCATCCGAGCGAAGAGGTTGCCAAGAGGAACAAGGCGTCGAGAATGCATGTTTTCTTGAAGTGATTCTACAAC ATATCATACAGACCATTGTTGGCCTTATCTTTATGCACCTCGAGCCGATCATGACTGGGTTTGAAGAAAAATGCCATG TGAAGCTTCGTGCAGACCTTCTCGGATTATCCAGATGCCGCTATTTATTACGGCTATATGTACGGAATGTCGCTTT GAAGATCTTGCAGGCTTTTTATTCGTTGATACATGGCAACTTTTTGCATAGATTGATGCATATGAATAAGACCTTAT ACAAATGGTTCCTCTGTTTCATCATGAACTATACGTGCCATATGCTTACGGTCTCTTTCAACAATCTGTTGAGGGC TTCTGTTAGATACTTTGGGAACCGGATTGCCATGACGTTAACTCATTGACTCACAGAGAGCAAATCATTCTTTTAC CTTTGCCACCATGAAGACTGTCGATGACCACTGTGGGTATGCTTGGCACTTGACCCATTCCAATGGCTTTCCCTAATA ACGCTGTCTATCAGATATCCACCACCAGCAATTTGGTATCAAGACGAACTTTGCTCAACCATTTTTCACTTTCTGGGAC AATTTGTTCCAACTAACTTTAAAGGGTTTGAAGAATATCAAAAAGAAGCAAAGACGTGTACCATCGACAAGTACAAA GAGTTTTTGCAAGAGAGAGAATTGAAAAGAAGGAGAACTCAAAAACCTCAAAGCTATGAATGCTGCTGAAAATGA AGTAAAGAAAGAGAAATAA</p>			
Protein sequence	Length	349	aa
<p>MNVTSNATAAGSFPLAFLKTSFGFMHYAKAPAINLRPKESLLPEMSDGVLALVAPVVAYWALSGIFHVIDTFHLAEKYRIHP SEEVAKRNKASRMHVFLEVLQHIQIVGLIFMHFEPIYMTGFEENAMWKLRLADLPRIIPDAIIYGYMYGMSALKIFAGFLF VDTWQYFLHRLMHMKNKTLYKWFHSHVHHELYVPYAYGALFNNPVEGFLDLTGTGIAMTLHLTHREQIILFTFATMKTVDD HCGYALPLDPFQWLFNNAVYHDIHQQFGIKTNFAQPFFTFWDLNFQTNFKGFEEYQKKQRRVTIDKYKEFLQERELEKK EKLNFKAMNAEENEVKEK*</p>			

Table S2.3 *ISC1* gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	<i>ISC1</i>		
GenBank Accession number	NM_001178910		
DNA sequence	Length	1434	bp
<p>ATGTACAACAGAAAAGACAGAGATGTTACAGAGAGGAAGGAAGATGGTCAATCTGAGTTTGAAGCACTGAACGGGAC CAACGCAATTATGTCGATAATAGTAAAGCGTATTCCATAAAGTTTCTGACCTCAATACATGGGGGTTAAAATACGTC TCAAACACCGTAAAGAAAGACTCAGAGCAATTGCTGATAAATTGGCGGGCCACTCAATGCTTACGCCAATATCTGAC GAGTTGTTGCCAATGGTGGAGATAGTAATGAAAACGAAGATTACGACGTGATTGCCTTACAAGAAATCTGGTGTGTG GAAGACTGGAAGTATCTAGCTTCTGCGTGTGCCTCAAAGTATCCGTATCAGCGTTTGTCCATTCTGGTATTCTGACGG GGCCTGGTTGGCCATACTGTCCAAGGTCCCGATAGAGTCACTTTCTTACCAGTTCCCGATAAACGGTAGACCGA GTGCGGTGTTCCGTGGCGACTGGTACGTAGGGAAATCTATAGCAATCACCGTATTGAACACAGGAACCCGCCCATTTG CAATAATGAACAGTACATGCACGCCCATACGCCAAGCAGGGTGTGCGGCTACTTGTGCCACAGATCTTGTGACGG CCTGGGATTTACGACGGCTCATTAAAGCTTACAGGCAGGCCGGTATGCGGTGATTGTGGTGGGTGACTTAAACTCCA GACCGGGCTCACTGCCCACAAAATTTCTCACGCAGGAGCCGGCCTGGTCTGACTCTGGGAGCAATTGCATGGGAAGC AAGACTGGCGGTGATCGCTCGTCTGCTCCATTGCAACAATTGCTTAAAGGGCTGTACCACGTGCGCATTCGCTGCTCAA ACATGGAGGGCCCAAGACAACCCGATGAGGCATGCAGTTGGATTATGCTTTATCGACCTGATTTCTTGCATACTCATGC AGTAGACGACGGTGTCAAGTCACTGAACGGATCCCTCACCTGGACTGCAAGTGTCTGACCAATTTGCATACTCATGC ACCCTTAACATCGTCCACAGGGCACAGAGTCCCGTCCATCCACTCCGTTAAGCGTGCGAAGACTCATGATAGAGAGC</p>			

TACTTGGCAACTGACGAAGAGGCAGCGAGATTGTTGACTAACAGAACCATAAGGTTGAAGTGCAGTTGTCAGCTGAC GGTACTGAGTTTGACTCCACTACTTTTGTAAAGGAGTTGCCGCCGAGGAGAAAATAAGTATTGCTACGGACTACAGT AACGACTACAAAAAGCATAAAATTTTGGATCTGAACCGTCTTTGCTGATGCAGATTCTGCGTAGTGATTTCAGAAAG ATTTTACGTTGACCAAATCCATAGACCAAGACATTACGGTAAGGGGTCTGCCCGCTATTTGGTAATTTCTTGAACC ATTAATAAACAGCTTGGTGGGTTGTTCCAGTTGCTTGGTTGCTGTAGTTGTGTACCACATGGGTGTTGCTTTGAAG AACATGAACCAGCTATTTGCATGTTTCTGTTCTGTGTCGGTGTCTTTGTTGGACTTTGATTGAATACGGTCTTCACCGT TTCCTATTTCAATTCGATGATTGGTACCTGAAAGTAACATCGCATTGCCACACATTTTCTACTACATGGTTGCCATCAT TACTTGCCCATTGGACAAGTACCGTTTAGTTATGCCACTACTCTGTTCTGTCATCCTTTGTGCTCCATTTTACAAGTTGGTA TTTGCTCTGCTGCCACTTTATTGGGCTTACGCTGGTTTTGCTGGCGGTCTTTTCGGTTATGTCTGTTATGACGAATGTCAT TTCTTCTGCACCACTCTAAATTGCCTCCCTTCATGCGTAAGTTGAAAAATATCACCTGGAACATCATTATAAAAACTAC CAACTGGGATTTGGCGTCACATCCTGGTTTTGGGACGAAGTTTTGGCACCTACTTAGGCCCCGATGCCCATTTGTCCA AAATGAAATATGAATAATTATTGGTCAGCCTCGTCCAAAATTGCAACCTAGGAAAAACAACATTACGAGCAATGAAAT CTTGAGCAAAGTTCTTTCTGCGACATATTTATTTCAAATCTTATTTATATTTATATGTAACACTAGTATATATACATAC ACATGGAATGGAATCATAGATGAAATCGATGAACAAGCAGCTTATTTCTTAAACGAATCCAATCGATGACGACTATA CGAAAAAGCGGCTCAAAAAAGAACCGAATTGACAGGTACATTTTTCTTTCACAGATCAGGTTGGCAAAAAAATACTA ACCGTATCCTGTGTTACTGTCGTTAGAACAGGAATCAAGATGAGCACAGTGAAGTTATCTGATAGTCCCTAGACTCTCA TTGACTTTTTGTAATATTAGGACACAGAATACTAAGGGTCTTCTCGAGAATAAATGAACACATAAGCAAATCAGAA ATTCTACATACTATAAGAGTACCTCATACTAATCATTCTTAAATCCATTCCGTTATCAGTAATTGCATTGCGAGGTTCCAT TAACTGGTGCCGACATTTATATAGTTATATTGGTCCAATGATCTTATCATCTCCACACCGTATACAACATATCTTAGT AAACAACGGCTAGTAAGTGAACGATGGTCCATTTAGGCACGAAGAATAAAGTAAAACTGTATCTGCGTTGCGCAA ATCATCAATTGCATTTATGTTGGAACGAGAGTAATTAATAGTGACATGAGTTGCTATGGTAAACATCTAATGCTTACATC GTATATTAATGTACACCTCGTATACGTTTAAAGTGTGATTGCACCTATTGCAGAAGGAATGTTAAACGAGAAGCTCAGAC AATACTGAAGCTGTGTTAAAGACCTATTAGTTGAACATGTTATGGTAGGTACATATATGAGGAATATGAGTCGTCACAT AAATGTATAGTAACTACCGGAATCACTATTATATTGGTCATGATTAATATGACCAATCGGCGT			
Protein sequence	Length	384	aa
MSTNTSKTLELFSKTKVQEHNNTANDCVVYQNRKIYDVTRFLSEHPGGDESILDYAGKDITEIMKDSVDVHEHSDSAYEILEDE YLIGLATDEEAARLLTNKNHKVEVQLSADGTEFDSTTFVKELPAEEKLSIATDYSNDYKXKHKFLDLNRPLLMQILRSDFKDFY VDQIHRPRHYGKSAPLFGNLEPLTKTAWWWVPVAVLPPVVVYHMGVALKNMNLQFLACFLFCVGVFVWTLIEYGLHRFL FHFDDWLPESNIAFATHFLLHGCHHYLPMDKYRLVMPPTLFFVILCAPFYKLVFALLPLYWAYAGFAGLFGVVCYDECHFFLH HSLKPPFMRKLLKXHYLHEHYKYNQLGFGVTSWFVDFVGTLYLGPDAPLSKMKYE*			

Table S2.5 *TEF2* gene sequence, including 1 kb regions at upstream of start codon and downstream of stop codon. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters.

Gene Name	<i>TEF2</i>		
GenBank Accession number	NM_001178466		
DNA sequence	Length	1377	bp
ATGGGTA AAGAGAAGTCTCACATTAACGTTGTCGTTATCGGTCATGTCGATTCTGGTAAAGTCTACCACTACCGGTCATT TGATTTACAAGTGTGGTGGTATTGACAAGAGAACCATCGAAAAGTTCGAAAAGGAAGCCGCTGAATTAGGTAAGGGT TCTTCAAGTACGCTTGGGTTTTGGACAAGTTAAAGGCTGAAAGAGAAAGAGGTATCACTATCGATATTGCTTTGTGGA AGTTCGAAACTCCAAGTACCAAGTTACCGTTATTGATGCTCCAGGTCACAGAGATTCATCAAGAACATGATTACTGG TACTTCTCAAGCTGACTGTGCTATCTTGATTATTGCTGGTGGTGTGCGTGAATTCGAAGCCGGTATCTCAAGGATGGT CAAACGAGAAACAGCTTTGTTGGCTTACCTTGGGTGTTAGACAATTGATTGTTGCTGTCAACAAGATGGACTCCG TCAAATGGGACGAATCCAGATTCCAAGAAATTGCAAGGAAACCTCCAACCTTATCAAGAAGGTTGGTTACAACCCAA AGACTGTTCCATTCTGCCAATCTCTGTTTGAACGGTGACAACATGATTGAAGCTACCACCAACGCTCCATGGTACAA GGGTTGGGAAAAGGAAACCAAGCCGGTGTGCTCAAGGGTAAGACTTTGTTGGAAGCCATTGACGCCATTGAACAAC CATCTAGACCAATGACAAGCCATTGAGATTGCCATTGCAAGATGTTTACAAGATCGTGGTATTGGTACTGTGCCAGT CGGTAGAGTTGAAACCGGTGTCATCAAGCCAGGTATGTTGTTACTTTCCGCCAGGTTGTTACCACATGAAGTCAA GTCCGTTGAAATGCATCGAACAATTGGAACAAGGTGTTCCAGGTGACAACGTTGGTTTCAACGTCGAAGAAGCTTTCC GTTAAGGAAATCAGAAGAGGTAACGCTCTGTGGTGACGCTAAGAACGATCCACCAAGGTTGCGCTTCTTTCAACGCT ACCGTCATTGTTTTGAACCATCCAGGTCAAATCTCTGCTGGTTACTCTCCAGTTTTGGATTGTCACTGCTCACATTGCT TGTAGATTGACGAATTGTTGAAAAGAACGACAGAAGATCTGGTAAGAAGTTGGAAGACCATCAAAGTCTTGAA GTCCGGTGACGCTGCTTGGTCAAGTTCGTTCCATTAAGCCAATGTGTGTTGAGCTTTCAGTGAATACCCACCATTA			

Table S2.6 *PGK1* gene sequence, including 1 kb regions at upstream of start codon and downstream of stop codon. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters.

Gene Name	<i>PGK1</i>		
GenBank Accession number	NM_001178725		
DNA sequence	Length	1251	bp
<p>ATGTCTTTATCTTCAAAGTTGCTGTCCAAGATTTGGACTTGAAGGACAAGCGTGTCTTCATCAGAGTTGACTTCAACGT CCCATTGGACGGTAAGAAGATCACTTAACCAAAGAATTGTTGCTGCTTTGCCAACCTCAAGTACGTTTTGGAACAC CACCCAAGATACGTTGTCTTGGCTTCTCACTTGGGTAGACCAAACGGTGAAGAAACGAAAAATACTCTTTGGCTCCAG TTGCTAAGGAATTGCAATCATTGTTGGGTAAGGATGTACCTTCTGAACGACTGTGTCGGTCCAGAAGTTGAAGCCGC TGCAAGGCTTCTGCCAGGTTCCGTTATTTTGTGGAAAACCTGCGTTACCACATCGAAGAAGAAGTTCCAGAAAAG GTCGATGGTCAAAGGTTCAAGGCTTCCAAGGAAGATGTTCAAAGTTTCAGACACGAATTGAGCTCTTTGGCTGATGTT TACATCAACGATGCCTTCGGTACCCTCACAGAGCTCACTCTTCTATGGTCGGTTTCGACTTGCCACAACGTGCTGCCG GTTTTCTGTTGGAAAAGGAATTGAAGTACTTCGGTAAGGCTTTGGAGAACCAACCAGACCATTCTGGCCATCTTAGG TGGTGCCAAAGTTGCTGACAAGATTCAATTGATTGACAACTTGTGGACAAGGTCGACTCTATCATCATTGGTGGTGGT ATGGCTTTCACCTTCAAAGAAGTTTTGGAAAACACTGAAATCGGTGACTCCATCTTCGACAAGGCTGGTGTGAAATCG TTCCAAAGTTGATGAAAAGGCCAAGGCCAAGGGTGTGGAAGTCGTCTGCCAGTCGACTTCATCATTGCTGATGCTTT CTCTGCTGATGCCAACCAAGACTGTCACTGACAAGGAAGGATTTCCAGCTGGCTGGCAAGGGTTGGACAATGGTCC AGAATCTAGAAAGTTGTTTCTGCTACTGTTGCAAAGGCTAAGACCATTGTCTGGAACGGTCCACCAGGTGTTTTCGAA TTCGAAAAGTTCGCTGCTGGTACTAAGGCTTTGTTAGACGAAGTTGTCAAGAGCTCTGCTGCTGGTAACACCGTCA TTGGTGGTGGTGACACTGCCACTGTCGTAAGAAGTACGGTGTCACTGACAAGATCTCCCATGTCTCTACTGGTGGTG GTGCTCTTTGGAATTATTGGAAGGTAAGGAATTGCCAGGTGTTGCTTTCTATCCGAAAAGAAATAA</p>			
Genomic DNA +/- 1 kb flanking sequences			
<p>AGACCGCAATTTTTCGAAGAAGTACCTTCAAAGAATGGGGTCTTATCTTGTGTTTGAAGTACCAGTACGAGGATAATA ATAGAAATGATAATATACTATAGTAGAGATAACGTCGATGACTTCCCATACTGTAATTGCTTTTAGTTGTGATTTTTAG TGTGCAAGTTTCTGTAAATCGATTAATTTTTTTCTTCTCTTTTATTAACCTTAATTTTTATTTTAGATTCTGACTTC AACTCAAGACGCACAGATATTATAACATCTGCATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGGAAAGAGTGAG GAACTATCGCATACCTGCATTTAAAGATGCCGATTTGGGCGCAATCCTTTATTTGGCTTACCCTCATACTATTATCA GGGCGAGAAAAGGAAGTGTTCCTCCTTCTGAAATGATGTTACCCTATAAAGCAGTGGCTCTTATCGAGAAAG AAATTCGCTCGTGTGATTTGTTGCAAAAAGAACAAAACGAAAAACCCAGACAGCTGCATCTCTGCTTCTCT ATTGATTGCAGCTTCCAATTTGCTCACACAACAAGTCTAGCGACGGCTCACAGGTTTTGTAACAAGCAATCGAAGGT TCTGGAATGGCGGGAAGGGTTTAGTACCACATGCTATGATGCCACTGTGATCTCCAGAGCAAAGTTCGTTGATCG TACTGTTACTCTCTCTTTCAAACAGAATTGTCCGAATCGTGTGACAACAACAGCCTGTTCTCACACTCTTTTCTTCT AACCAAGGGGGTGGTTAGTTAGTAGAACCTCGTAAACTTACATTTACATATATAAATTGCATAAAATGGTCAA TGCAAGAAATACATATTTGGTCTTTTCTAATTCGTAGTTTTTCAAGTCTTAGATGCTTTCTTTTCTCTTTTACAGATC ATCAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACAATGCTTTATCTTCAAAGTTGCTGTCCAAGATTTGG ACTTGAAGGACAAGCGTGTCTTCATCAGAGTTGACTTCAACGTCCATTGGACGGTAAGAAGATCACTTCTAACCAAAG AATTGTTGCTGCTTTGCCAACCATCAAGTACGTTTTGGAACACCACCAAGATACGTTGTCTTGGCTTCTCACTTGGGTA GACCAAACGGTGAAGAAACGAAAAATACTTTGGCTCCAGTTGCTAAGGAATTGCAATCATTGTTGGGTAAGGATG TCACCTTCTGAACGACTGTGTCGGTCCAGAAGTTGAAGCCGCTGTCAAGGCTTCTGCCAGGTTCCGTTATTTTGTG GAAAACCTGCGTTACCACATCGAAGAAGAAGGTTCCAGAAAGGTCGATGGTCAAAGGTTCAAGGCTTCCAAGGAAGA TGTTCAAAGTTCAGACACGAATTGAGCTCTTTGGCTGATGTTTACATCAACGATGCCTTCGGTACCCTCACAGAGCT CACTCTTCTATGGTCGGTTTCGACTTGCCACAACGTGCTGCCGTTTTCTGTTGGAAAAGGAATTGAAGTACTTCGGTA AGGCTTTGGAGAACCAACGACACTTCTTGCCATCTAGGTGGTATGGCTTTACCTTCAAGAAGGTTTTGAAAAACT CAACTTTGGACAAGGTCGACTCTATCATCATTGGTGGTATGGCTTTACCTTCAAGAAGGTTTTGAAAAACT GAAATCGGTGACTCCATCTTCGACAAGGCTGGTGTGAAATCGTTCAAAGTTGATGGAAGAAGGTTTGAAGGCAAGGG TGTCGAAGTCTTGGCAGTCGACTTCATCATTGCTGATGCTTTCTGCTGATGCCAACCAAGACTGTCACTGACA AGGAAGGATTCCAGCTGGCTGGCAAGGGTTGGACAATGGTCCAGAATCTAGAAAGTTGTTGCTGCTACTGTTGCAA AGGCTAAGACCATTGTCTGGAACGGTCCACCAGGTGTTTTGCAATCGAAAAGTTGCTGCTGGTACTAAGGCTTTGTT AGACGAAGTTGCAAGAGCTCTGCTGCTGGTAACCCGTCATCATTGGTGGTGGTGGTACTGCTGCTAAGAA GTACGGTGTCACTGACAAGATCTCCATGTCTCTACTGGTGGTGGTGGTCTTTTGAATTATTGGAAGGTAAGGAATTG CCAGGTGTTGCTTTCTATCCGAAAAGAAATAAATTGAATTGAATTGAAATCGATAGATCAATTTTTTTCTTTCTCTTC CCCATCTTACGCTAAAATAATAGTTTATTTTATTTTTGAATTTTTTATTTATATACGTATATATAGACTATTTATTTAT</p>			



```

CTTTAATGATTATTAAGATTTTTATTAATAAAAAAAAAATTCGCTCCTCTTTAATGCCTTTATGCAGTTTTTTTTCCCATTCG
ATATTCTATGTTTCGGGTTTCAGCGTATTTAAGTTTAATAACTCGAAAATTCTGCGTTCGTTAAAGCTTTTCGAGAAGGAT
ATTATTTTCGAAATAAACCGTGTGTGAAGCTTGAAGCCTTTTTGCGCTGCCAATATTCTTATCCATCTATTGTAATCTTT
AGATCCAGTATAGTGTATTCTTCTGCTCCAAGCTCATCCACTTGCAACAAAAAAGTCTAATCTTCTGCAATAATTTCC
ATCCTTGGCATTAGAGACATATATTGGTCAATCGGTTTTAATTTGTTTCTTCTTCTTCTTCTTCAATTAACAATGTG
CCTTCATCATTTTTTCATTTCTTTCAGTTCCTTAATTCTCTGGAGACAATTTGACTCCCATGAAAAATTTCTTGATTATCTC
AGAAGAATACCACCTTTTCTTCTCCATACCAACAATCTTCTGCTGTACGAAATCTCTCCACTATTCTTTTAAGAAATAG
CTTTTCCAAGTTTGGCGTTAGCTGTAATAAACAATGGATATAGCCATCCTTGTACAACCTAATACAAAGCGTTTCCATGA
TTTTTGCCAACCTCGGTGCTGTCATTACAAAACGGCTTGAAAAATAAAGATCAATGTCACCACACTTGGAAATAGCCCT
ATTATAACTTCCCTGTAATTCGACTTGGCATTAGGATCAATGCCACGCAGCGCTTTTGAACCTTCTTAAATGTGTGA
AACATTCATTCCGAGACATCTGCATAACCAATCGTCGTAATATGACCAACCAATAAAA
    
```

Table S2.7 mCherry gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	mCherry		
GenBank Accession number	-		
DNA sequence	Length	711	bp
ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTC CGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTG AAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCTCAGTTCATGTACGGCTCCAAGGCTACG TGAAGCACCCCGCGACATCCCCGACTACTTGAAGCTGTCTTCCCCGAGGGCTCAAGTGGGAGCGCGTGATGAACT TCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCTTGACGAGCGGCGAGTTCATCTACAAGGTGAAGCTG CGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGAT GTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCT GAGGTCAAGACCACCTACAAGGCCAAGAAGCCGTGCAGCTGCCGGCGCCTACAACGTCAACATCAAGTTGGACATC ACCTCCACAACGAGGACTACACCATCGTGAACAGTACGAACGCGCCGAGGGCCGCACTCCACCGCGGCATGGA CGAGCTGTACAAGTAA			
Protein sequence	Length	236	aa
MVSKGEEDNMAIIEKFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKH PADIPDYLKLSFPEGFKWERVMNFDGGVVTVTQDSSLQDGEFIYKVKLRGTFNFPDGPVMQKKTMGWEASSERMYPED GALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*			



Chapter 3

Identification and characterization of an α -galactosylceramide synthase







3.1 Identification of an α -galactosylceramide synthase from *Bacteroides fragilis*

The identification of a glycosyltransferase that can catalyze the synthesis of this molecule using UDP-galactose and ceramide is pivotal for this project. The following chapter covers the identification of an enzyme that can perform the reaction mentioned above.

The first step was to identify an organism that can synthesize α -GalCer. Then, sequences were analyzed by using bioinformatics tools and some potential candidates were found. The candidates' activity was analyzed in *E. coli* using ceramide-NBD assay. Finally, the enzyme that showed α -GalCer synthase activity was characterized to study its kinetic and structure properties.

NOTE: There was a turning event in this part of the thesis. At the time this investigation started, no α -GalCer synthase was identified. Thus, the literature study and *in silico* analysis were done previous to the publication of BF9343_3149 (UniProt A0A380YRQ3) as α -GalCer synthase by OKINO *et al.*³⁸. However, it has been decided to describe our work as it was performed before the mentioned publication in 2020.

3.1.1 Literature study

Although agelasphin was first identified in 1993 as an antitumor agent in the marine sponge *Agelas mauritanus*^{26,27}, more recently it was isolated from *Bacteroides fragilis*, a common component of the human colon microbiota generally commensal but that can become pathogenic and caused infections if displaced into the bloodstream or surrounding tissues following trauma, surgery, or disease³⁷. α -GalCer and other bacterial GSLs are exogenous ligands of CD1d and play important roles in the host's immune system. The bacterial origin of α -GalCer supports the hypothesis that the identified congeners in the sponge *Agelas* were actually produced by symbiotic bacteria²⁵.

Recently, Okino *et al.*³⁸ identified a glycosyltransferase from *Bacteroides fragilis* that can catalyze the reaction of α -GalCer synthesis. This protein showed homology to ceramide UDP-glucuronosyltransferase (Cer-GlcAT) from *Z. mobilis*, another transferase that uses ceramide as acceptor. They studied different substrates for the *B. fragilis* GT, showing that ceramide, but apparently not diacylglycerol, was the acceptor, and UDP-galactose or UDP-glucose (with preference for the first) were the donor substrates, but not UDP-GalNAc, UDP-GlcA, UDP-GlcNAc or GDP-Man. The α configuration of the glycosidic linkage confirmed the enzyme as a retaining glycosyltransferase synthesizing α -GalCer. Other congeners have also been identified in other gut bacteria such as *Bacteroides vulgatus* and *Prevotella capri*¹⁰⁸. The *Bacteroides* α -GalCer is thought to be a critical signaling molecule in gut physiology. Studies with mice identified a different congener of α -GalCer with a β -hydroxylated palmitic acid and C18-sphinganine (α -GalCerMLI) in the mice colon that was not



detected in germ-free mice¹⁰⁸. Because its production was dependent on diet and inflammation, it was suggested that its effector function through iNKT cells is important in gut energy and immunity. Although the synthetic pathway for this unique α -GalCer is unknown, the identified *Bacteroides* GT may be involved in its production.

α -GalCer has gained attention as vaccine adjuvant for immunotherapy of tumors by induction of potent natural killer cell-dependent anti-tumor cytotoxic responses^{109–112}. Since chemical synthesis of α -GalCer is tedious with low yields^{113,114}, biotechnological production by biocatalysis and cell factory approaches are of current interest, thus requiring better knowledge of its biosynthetic enzymes.

3.1.2 *In silico* methods for α -GalCer synthase identification

After identification of *B. fragilis* as a natural producer of α -GalCer, it was hypothesized that the glycosyltransferase that catalyzes its synthesis could be identified by using *in silico* approaches. By aligning a pool of *B. fragilis* with known non-processive retaining glycosyltransferases that use diacylglycerol as acceptor, several GT candidates were identified.

Although diacylglycerol and ceramide are not the same molecule, they have a similar chemical structure and usually, GT that use DAG, are promiscuous and can also use ceramide. This could indicate a similar structure and, therefore, a similar sequence.

3.1.2.1 Multiple sequence alignment of *B. fragilis* glycosyltransferases

To identify the α -GalCer synthase in *B. fragilis*, GTs from this microorganism that could also show galactosylceramide transferase activity were searched using the CAZy database, which compiles identified and annotated GTs¹¹⁵. A total of 83 putative GT genes from *B. fragilis* NCTC9343 strain are annotated in this database, however, only retaining GTs (families 3, 4, 5, 8 and 35) were considered in the analysis since it is a key feature in the α -GalCer synthase. GT3, 5 and 35 members were discarded as they contain enzymes involved in glycogen and starch metabolism. Therefore, they would not contain a GT using lipid acceptors. Families 4 and 8 does contain enzymes that use lipids as acceptor but only GT4 enzymes have been reported to use ceramide or diacylglycerol as acceptors. The 22 annotated GT4 sequences from *B. fragilis* genome were aligned with 2 biochemically characterized GT4 glycolipid synthases: the non-processive monoglucosyldiacylglycerol synthases from *Acholeplasma laidlawii*^{116,117} and *Streptococcus pneumoniae*¹¹⁸.

Results from the multiple sequence alignment (Figure S3.1 in this chapter's Annexes) are shown in a Neighbor-Joining tree in Figure 3.1. Among the *B. fragilis* GT4 sequences, only two were clustered with the query sequences: BF9343_3149 (UniProt A0A380YRQ3), and BF9343_1306 (UniProt Q5LFK6).

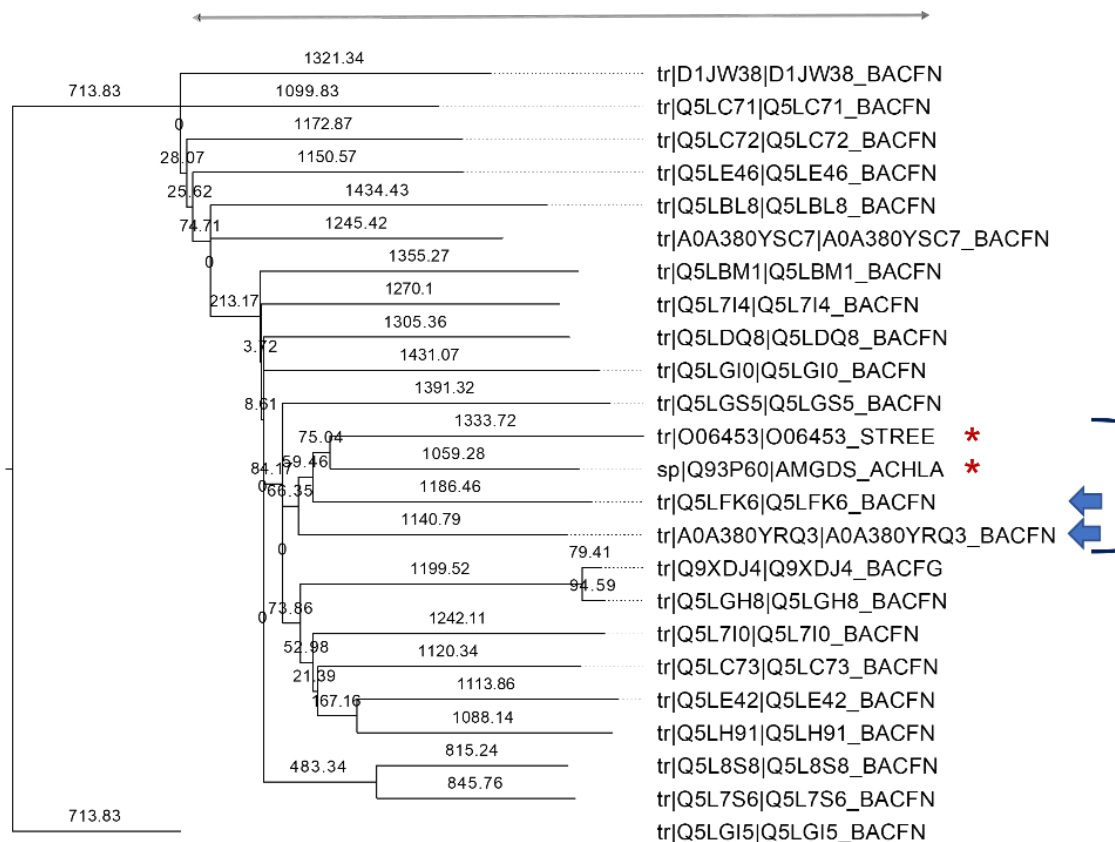


Figure 3.1 Neighbor-Joining phylogenetic tree of annotated GT4 enzymes. Sequences from *Bacteroides fragilis* and query sequences of monoglucosyldiacylglycerol synthases from *Acholeplasma laidlawii* (UniProt Q93P60) and *Streptococcus pneumoniae* (UniProt O06453) (marked with *). Closest *B. fragilis* sequences are indicated with arrows.

3.1.2.2 BLAST search for candidate sequences

The following step was to search for homologous sequences of the candidates with identified activity so they could be discarded if the activity was another than glycolipid synthase.

For this analysis other four GT genes from *B. fragilis* among the non-classified GT sequences in CAZY were also considered: BF9343_3589 (UniProt Q5L962), BF9343_0008 (UniProt Q5LJ89), BF9343_0585 (UniProt Q5LHL6) and BF9343_0009 (UniProt A0A380YSP3). These sequences were not included in the previous alignment since they are in the non-classified family of GT sequences and, therefore, they do not align with other GT sequences.

BLAST searches with these 6 candidate sequences were performed against Swiss-Prot, refseq and TrEMBL databases. Candidates with high homology ($\geq 50\%$) to other enzymes with known activity were discarded. Thus, BF9343_0008 and BF9343_0009 showed high homology with D-inositol 3-phosphate glycosyltransferase (57.72% and 74.89%, respectively) and thus, were rejected. By contrast,

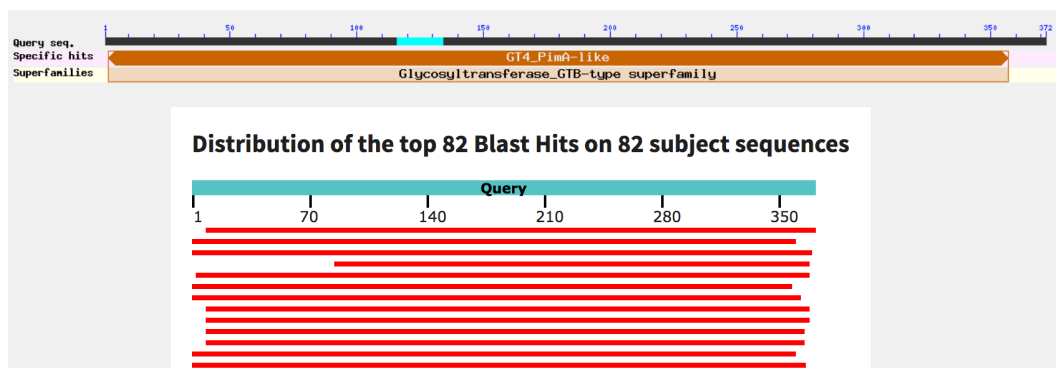


BF9343_3589 and BF9343_0585 did not show any significant hits and BF9343_1306 and BF9343_3149 showed homology to some uncharacterized putative glycosyltransferases. Moreover, BF9343_3149 showed 23.98% homology with α -monoglucosyldiacylglycerol synthase from *Acholeplasma laidlawii*, therefore, BF9343_3149 and BF9343_1306 were considered as candidates. In conclusion, the GT4 sequences BF9343_3149 and BF9343_1306, and the non-classified GT sequences BF9343_3589 and BF9343_0585 were considered as potential candidates for ceramide glycosyltransferase activity. BLAST results are shown in Figure 3.2, hits are listed in the table and the graphic shows the query cover and score of each hit (in the same order). Arrows indicate the mentioned homologies with enzyme with known activities.

Chapter 3. Identification and characterization of an α -galactosylceramide synthase

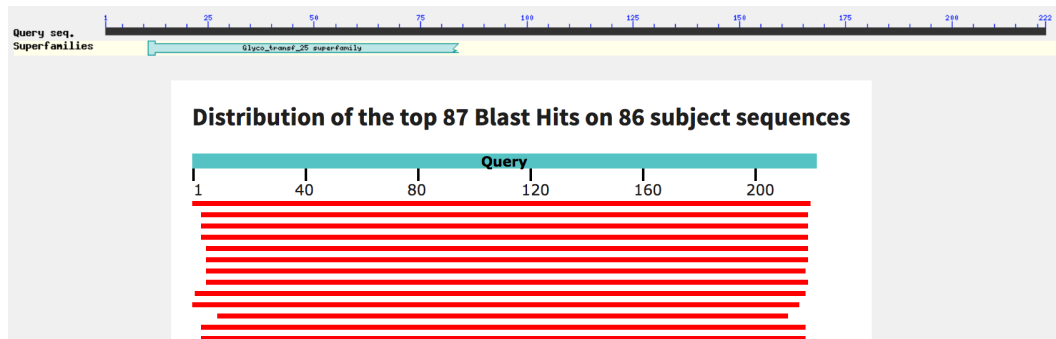
BF9343_0008

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
hypothetical protein BSHG_1569 [Bacteroides sp. 3_2_5]	Bacteroides sp...	724	724	100%	0.0	99.46%	372	EES88650.1
glycosyltransferase [Enterococcus durans]	Enterococcus d...	707	707	97%	0.0	99.45%	364	MZH34337.1
glycosyl transferase family 1 [Tannerella forsythia]	Tannerella forsy...	554	554	98%	0.0	75.62%	621	KKY61921.1
glycosyltransferase [Prevotella sp.]	Prevotella sp...	507	507	96%	8e-174	68.51%	591	MBQ44483676.1
glycosyl transferase [Prevotella sp. HMSC073D09]	Prevotella sp. H...	473	473	98%	2e-163	62.50%	367	OFQ19641.1
glycosyltransferase [Prevotella sp.]	Prevotella sp...	471	471	99%	4e-163	61.35%	368	RKW57467.1
glycosyl transferase [Tannerella forsythia]	Tannerella forsy...	437	437	76%	3e-150	77.03%	312	PDP44537.1
D-inositol 3-phosphate glycosyltransferase [Tannerella forsythia]	Tannerella forsy...	427	427	98%	4e-145	57.72%	386	SCQ20826.1



BF9343_0009

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
glycosyl transferase [Enterococcus durans]	Enterococcus d...	465	465	100%	2e-165	100.00%	222	MZH34338.1
hypothetical protein BSHG_1568 [Bacteroides sp. 3_2_5]	Bacteroides sp...	464	464	100%	6e-165	99.55%	222	EES88649.1
glycosyl transferase [Tannerella forsythia]	Tannerella forsy...	350	350	98%	7e-120	73.64%	224	PDP44536.1
hypothetical protein BFO_1699 [Tannerella forsythia 92A2]	Tannerella forsy...	348	348	98%	3e-119	73.18%	224	AEW19757.1
glycosyl transferase [Paludibacter sp. 47-17]	Paludibacter sp...	345	345	99%	6e-118	72.85%	224	OJX91985.1
glycosyltransferase [Bacteroides pyogenes]	Bacteroides pyo...	357	357	98%	3e-117	75.80%	623	TYK32344.1
D-inositol-3-phosphate glycosyltransferase [Bacteroides pyogenes]	Bacteroides pyo...	354	354	98%	6e-116	74.89%	623	MBR8719830.1
glycosyl transferase family 1 [Tannerella forsythia]	Tannerella forsy...	352	352	97%	5e-115	74.19%	621	KKY61921.1
D-inositol 3-phosphate glycosyltransferase [Tannerella forsythia]	Tannerella forsy...	350	350	97%	1e-114	74.19%	612	SCQ21685.1

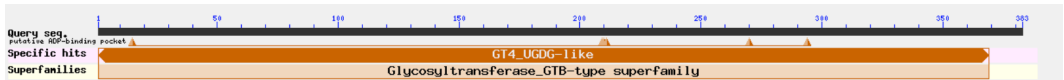


Chapter 3. Identification and characterization of an α -galactosylceramide synthase

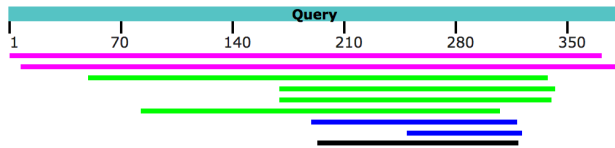


BF9343_3149

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Alpha-monoglucosyldiacylglycerol synthase; Short=Alpha-MGS; Short=MGlCDAG synthase; AltName: Acholeplasma la...	<i>Acholeplasma la...</i>	87.4	87.4	96%	9e-18	23.98%	398	Q93P60.1
RecName: Full=Uncharacterized glycosyltransferase MJ1607 [Methanocaldococcus jannaschii DSM 2661]	<i>Methanocaldoco...</i>	82.0	82.0	97%	6e-16	23.99%	390	Q59002.1
RecName: Full=Putative teichuronic acid biosynthesis glycosyltransferase TaaC [Bacillus subtilis subsp. subtilis st...	<i>Bacillus subtilis s...</i>	69.7	69.7	74%	8e-12	25.00%	389	Q32272.1

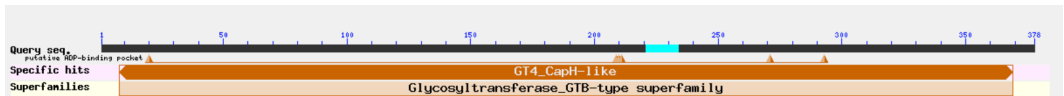


Distribution of the top 9 Blast Hits on 9 subject sequences

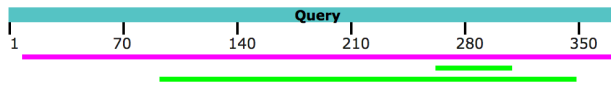


BF9343_1306

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Putative glycosyltransferase EpsF [Bacillus subtilis subsp. subtilis str. 168]	<i>Bacillus subtilis s...</i>	164	164	96%	9e-46	29.73%	384	P71055.1
RecName: Full=N,N'-diacetylbaicilosaminyl-diphospho-undecaprenol alpha-1,3-N-acetylgalactosaminyltransferase...	<i>Campylobacter j...</i>	54.7	54.7	12%	5e-07	53.19%	376	Q0P9C9.1
RecName: Full=N-acetyl-alpha-D-glucosaminyl L-malate synthase; Short=GlcNAc-Mal synthase; AltName: Full=L...	<i>Bacillus anthracis</i>	53.9	53.9	67%	8e-07	25.95%	381	Q81ST7.1

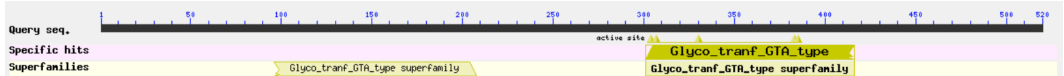


Distribution of the top 3 Blast Hits on 3 subject sequences

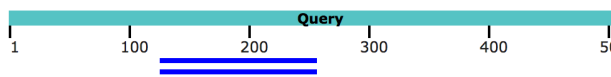


BF9343_0585

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Validoxylamine A glucosyltransferase [Streptomyces hygroscopicus subsp. jinggangensis 5008]	<i>Streptomyces hygro...</i>	44.7	44.7	25%	0.001	24.50%	422	H2K893.1
RecName: Full=Validoxylamine A glucosyltransferase [Streptomyces hygroscopicus subsp. limoneus]	<i>Streptomyces hygro...</i>	43.9	43.9	25%	0.002	24.50%	453	Q15JF5.1



Distribution of the top 2 Blast Hits on 2 subject sequences



Chapter 3. Identification and characterization of an α -galactosylceramide synthase

BF9343_3589

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Alpha-(1,3)-fucosyltransferase B; AltName: Full=Galactoside 3-L-fucosyltransferase Drosophila...	Drosophila mela...	50.8	50.8	40%	7e-06	28.67%	444	Q9VLC1.2
RecName: Full=Alpha-(1,3)-fucosyltransferase 10; AltName: Full=Fucosyltransferase X; Short=Fuc-TX; Short=Fu...	Gallus gallus	47.4	47.4	25%	9e-05	30.95%	475	Q8AWB5.2
RecName: Full=Alpha-(1,3)-fucosyltransferase 11; AltName: Full=Fucosyltransferase X; Short=Fuc-TX; Short=Fu...	Oryzias latipes	47.0	47.0	26%	1e-04	33.71%	497	Q5F2N2.1
RecName: Full=Alpha-(1,3)-fucosyltransferase 10; AltName: Full=Fucosyltransferase X; Short=Fuc-TX; Short=Fu...	Xenopus laevis	43.5	43.5	24%	0.002	30.95%	469	Q6NTZ6.2

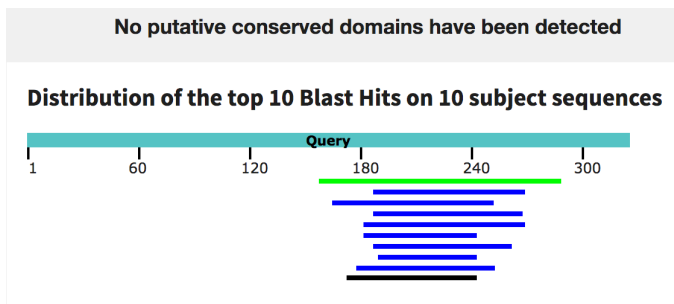


Figure 3.2 BLAST search of candidate sequences. Hits are listed in the table, and the representation of the query cover and scores are shown in the graphic below. Arrows indicate known activities.



3.1.3 Analysis of candidate sequences expressed with pET28a vector in *E. coli*

3.1.3.1 Cloning and expression in *E. coli*

Cloning

For expression in *E. coli*, the four GT candidates were cloned into pET28a. This vector is a high expression vector for *E. coli* with a C-terminal His-tag for purification. This would allow to purify the enzyme for further biochemical characterization.

Sequences were ordered from GeneArt (ThermoFisher) in synthetic vectors. Cloning was carried out using restriction enzymes NcoI and XhoI. When ordered, the genes contained both sites at 5' and 3' ends, respectively. Besides, the sequences also contained NheI and Bpu1102I sites, since the vector allows to use these sites to clone the sequences with an N-terminal His-tag and was considered in case C-terminal His-tag caused problems. Figure 3.3 shows the added sequences at both ends of the sequences.

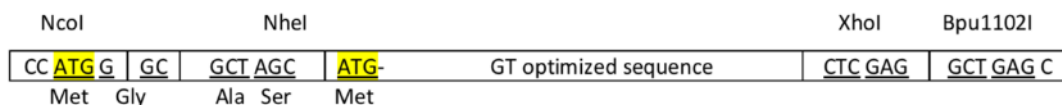


Figure 3.3 Scheme for ordered GT sequences with additional bp that include restriction sites. Additional sequences are indicated in letters. Two start codons (ATG) were present. Start codon in the final plasmid would use the first codon, adding an extra 4 residues to the final protein. Yellow codons indicate possible start codons available for cloning with NcoI or NheI.

In addition, sequences were codon-optimized for *S. cerevisiae* expression that would be used later in this project, with some modifications to improve expression in *E. coli*.

Figure 3.4 shows the scheme of the cloning strategy followed for plasmid construction. The four final designed vectors are shown in Figure 3.5, which shows the restriction sites, gene sequences and plasmid elements on each construct.

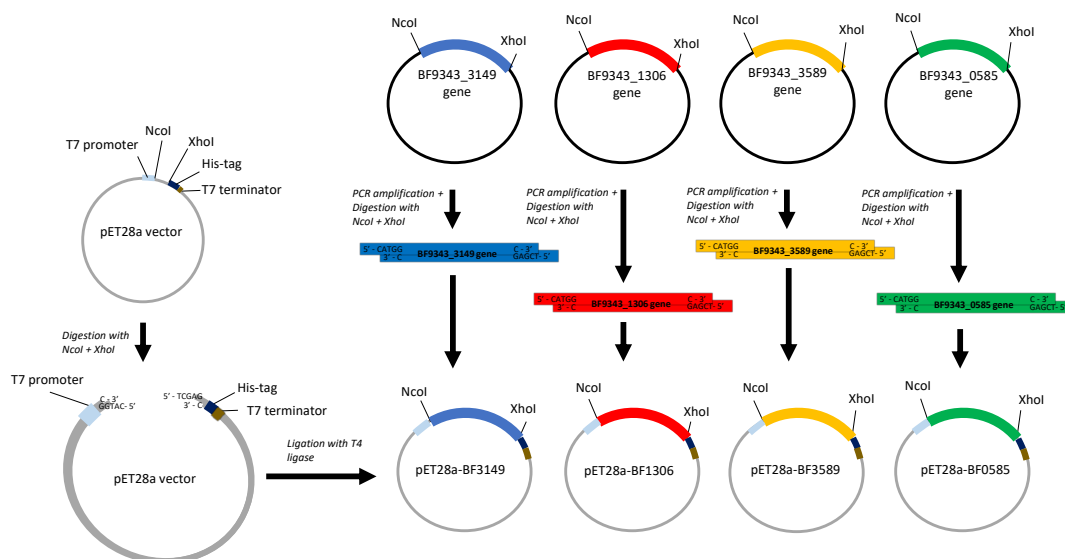


Figure 3.4 Strategy followed for GT cloning into pET28a vector. GT sequences were ordered as synthetic genes. GT genes were amplified by PCR and the digested with NcoI and XhoI restriction enzymes. pET28a vector was also digested with NcoI and XhoI. After purification, fragments are ligated with T4 ligase, yielding 4 different plasmids: pET28a-BF3149, pET28a-BF1306, pET28a-BF3589 and pET28a-BF0585.

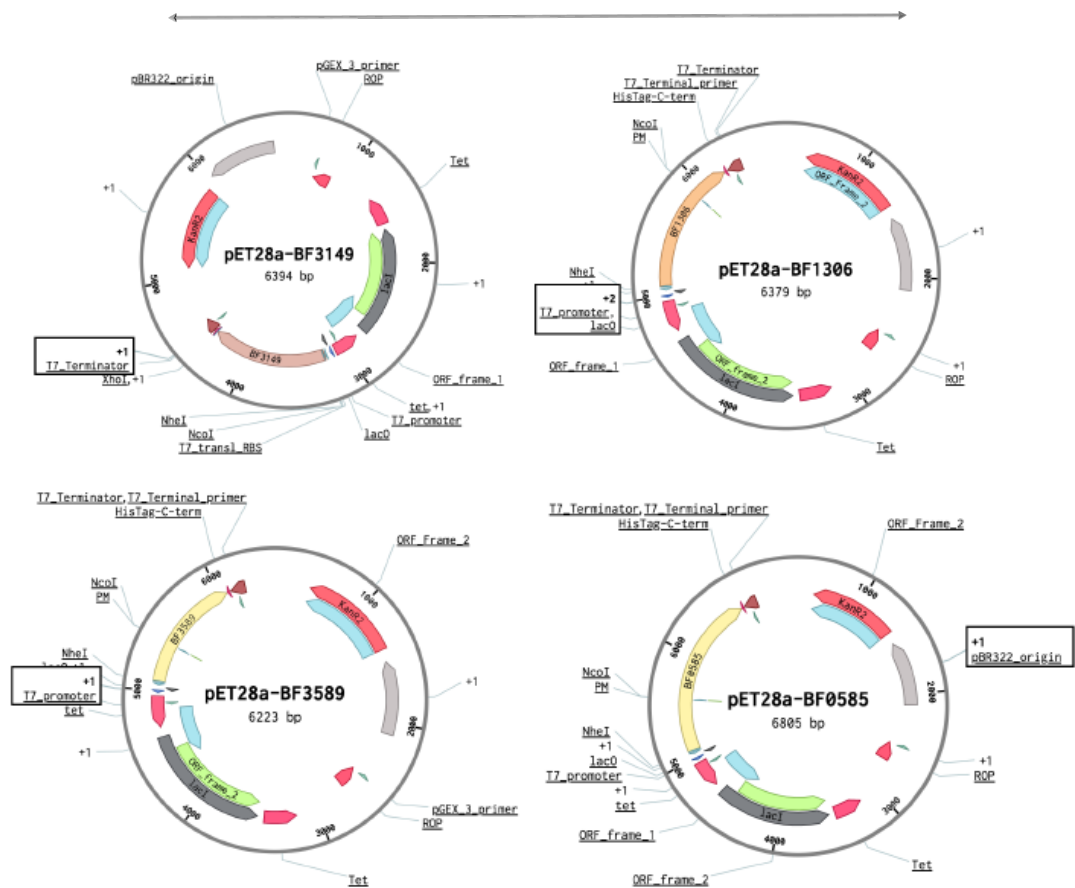


Figure 3.5 Vector maps for the final assembled vectors for *B. fragilis* candidate GT expression. Vector elements are marked in colored arrows inside the circle of the plasmid that indicate sequence direction. In the center, the name of the vector and its size. Legend: Grey, pBR322 origin. Dark grey, lacI. Small blue and light blue lines, T7 promoter and RBS, respectively. Brown/orange/yellow, inserted GT sequences (Top-left: BF9343_3149, Top-right: BF9343_1306, Bottom-left: BF9343_3589, Bottom-right: BF9343_0585). Red (short): His-tag. Strong brown: T7 terminator. Red: Kanamycin R.

For each gene, sequences were amplified with PCR using primers that covered the added sequences with the restriction sites and part of the gene (see Table 3.1 and Figure 3.6), and then purified. Amplified sequences and pET28a vector were then digested with NcoI and XhoI enzymes following a preparative digestion protocol and then purified again to remove the remaining short bp that were cut. Finally, each gene sequence was ligated to the vector at ratio 3:1 (insert:vector) using T4 DNA ligase, incubating at room temperature for 5 min and finally transforming *E. coli* DH5 α with the product of ligation. A negative control of transformed *E. coli* with digested pET28a was also performed. Colonies were tested by Colony PCR (with primers T7 promoter and T7 terminator, Figure 3.7) and positives were sent for sequencing.



Table 3.1 List of primers and templates used to amplify each fragment for pET28a vectors for each *B. fragilis* candidate GT. Table indicates the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Fragment	Template	Primers	Length of fragment (bp)
Vector backbone	pET28a	pET28a_fw	5369
		pET28a_rv	
BF1306 sequence	pMAT-Q5LFK6	Q5LFK6-fw	1148
		Q5LFK6-rv	
BF3149 sequence	pMAT-Q5LAE3	Q5LAE3_fw	1163
		Q5LAE3_rv	
BF3589 sequence	pMAT-Q5L962	Q5L962_fw	992
		Q5L962_rv	
BF0585 sequence	pMARQ-Q5LHL6	Q5LHL6_fw	1574
		Q5LHL6_rv	

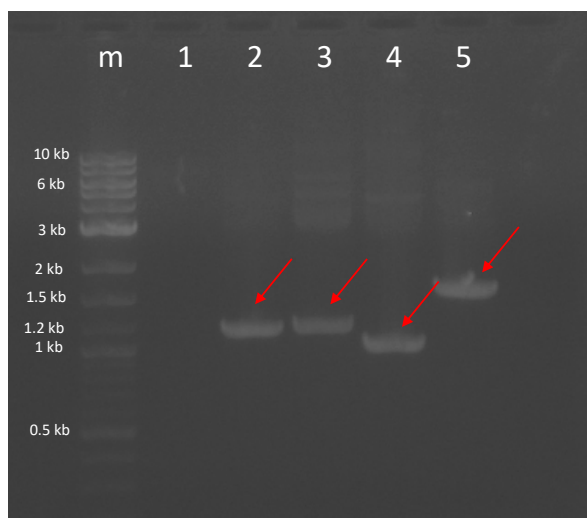


Figure 3.6 Agarose gel electrophoresis results of PCR for pET28a-GT vector synthesis. Lane labeled on top as *m* contains DNA marker (1 kb plus DNA ladder from NEB) with some indications of fragments sizes. Arrows show fragments that have been amplified after PCR. Results of amplification of pET28a vector (lane 1), BF1306 (lane 2), BF1349 (lane 3), BF3589 (lane 4) and BF0585 (lane 5) genes.

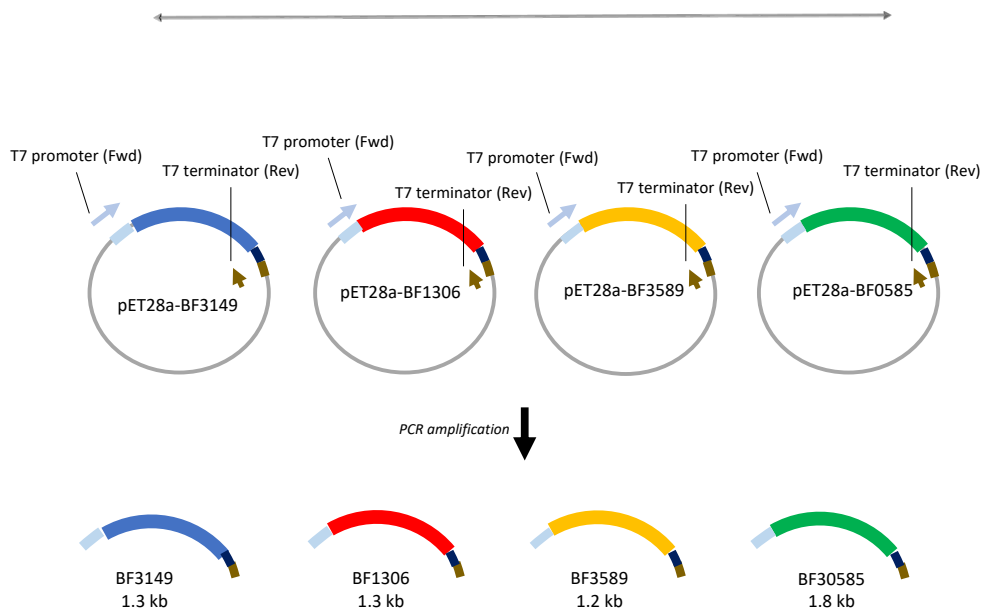


Figure 3.7 Colony PCR analysis of pET28a-GT vectors. T7 promoter and T7 terminator primers were used for amplification. Each GT gene yielded different fragment sizes: 1.3 kb for BF3149, 1.3 for BF1306, 1.2 for BF3589 and 1.8 for BF0585.

Expression

After successful cloning, all four plasmids with each candidate gene together with pET28a, which will be used as negative control, were used to transform *E. coli* BL21(DE3)Star strain for protein expression. Although several conditions for protein expression were tested (only for BF3149 gene), no important differences were observed among the different conditions, therefore, it was decided to use standard IPTG induction at 1 mM final concentration at 25°C for 4h. Expression was induced when culture reached an OD of 1 and growth was set at 37°C.

Preparative cultures of 300 mL were set up for protein analysis, cultures were induced with IPTG at 25°C and cell pellets were resuspended in phosphate buffer and lysed by sonication in a Soniprep 150 sonifier at 4°C (10 min, 10 s ON/ 20 s OFF, 50% amplitude). Lysed samples were centrifuged at 25,000xg for 60 min and soluble and insoluble fractions were separated and stored. SDS-PAGE gel electrophoresis analysis was carried out.

Results (Figure 3.8) showed that cell extracts and pellets from BF9343_3149 and BF9343_0585 presented the expected new proteins with molecular masses of 45.0 and 63.8 kDa respectively. By contrast, expression of BF9343_3589 and BF9343_1306 proteins with theoretical masses of 40.4 and 45.5 kDa were not clear compared to the control (empty pET28a).

Nevertheless, all four cell extracts were assayed with UDP-Gal and Cer-NBD (C6-ceramide with fluorescent label) for galactosyltransferase activity.

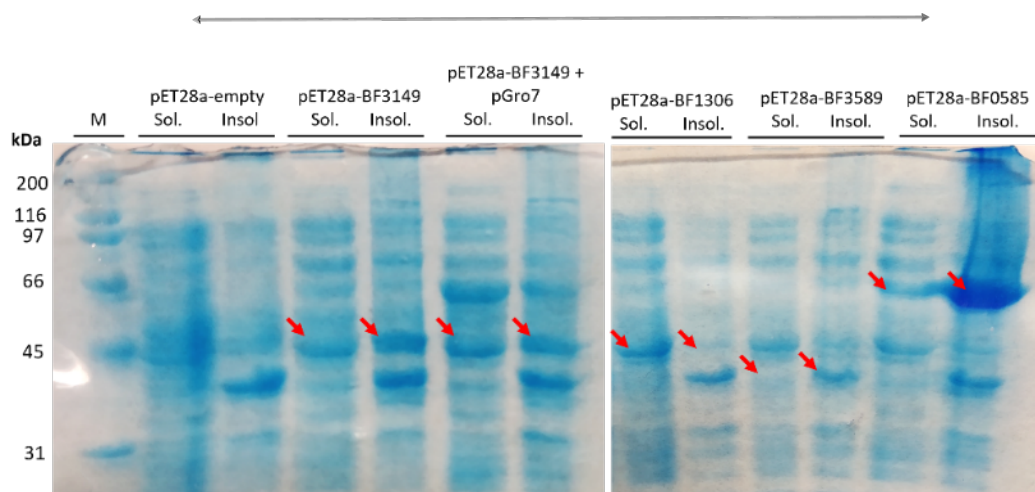


Figure 3.8 SDS-PAGE gel electrophoresis results of expression of *B. fragilis* GT candidates in *E. coli*. Labels indicate the soluble and insoluble fraction for each culture; M: protein molecular mass marker. Arrows show the expressed proteins. Expected MW for BF3149, BF1306, BF3589 and BF0585 proteins were 45.0, 45.5, 40.4 and 63.8 kDa, respectively. Expression with pGro7 vector, expressing GroES and GroEL chaperones, will be discussed later in the text.

3.1.3.2 Ceramide-NBD glycosyltransferase assay with candidates

To prove the glycosyltransferase activity of the candidates, an assay was carried out using cell-free extracts of the cultures expressing each of these candidates, plus pET28a as control. In this assay, C6-ceramide with the fluorescence label NBD, Cer-NBD, bound at the fatty acid position, is used as acceptor and UDP-galactose as donor. Reaction is monitored by HPLC with fluorescence detection with since both product (galactosylceramide-NBD) and substrate (Cer-NBD) have the fluorescent label that allows for their detection (see Material and methods chapter).

For the reaction, the Cer-NBD acceptor is solubilized with an equimolar amount of BSA, the cell extract is added, and reactions initiated by addition of UDP-Gal at pH 7.5, 37°C, and monitored by HPLC. Only the cell extract expressing BF9343_3149 showed activity synthesizing the corresponding galactosylceramide (GalCer) although the conversion rate was low (Figure 3.9).

To confirm the activity of BF9343_3149, it was decided to recreate the conditions of the reaction and monitor the reaction at longer times (Figure 3.10). HPLC results showed the expected peak corresponding to GalCer and that after 120 min, conversion rate was not yet complete.

Results confirm that only BF9343_3149 (from now on also named as Bf3149) but not BF9343_1306, BF9343_3589 and BF9343_0585, is a glycosylceramide synthase.

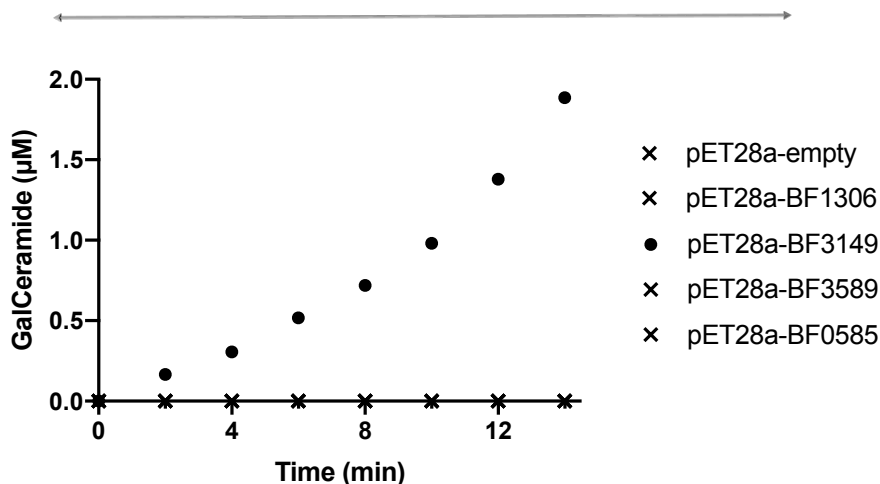


Figure 3.9 Time-course reaction monitoring of galactosyltransferase activity of a cell-free extract expressing all GT candidates. The reaction was carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and a cell-free extract (as enzyme) from *E. coli*/pET28a-GT in phosphate buffer at pH 7.5 and 37 $^{\circ}$ C.

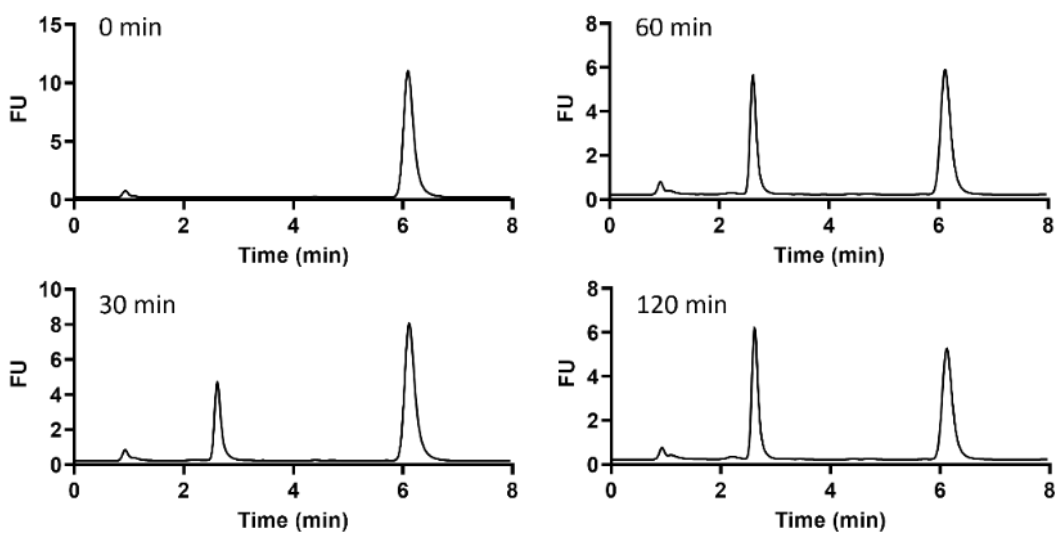


Figure 3.10 Time-course reaction monitoring of galactosyltransferase activity of a cell-free extract expressing BF9343_3149. The reaction was carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and, as enzyme source, a cell-free extract from *E. coli*/pET28a-BF3149 (\approx 4 mg/mL total protein) in phosphate buffer at pH 7.5 and 37 $^{\circ}$ C.



3.1.4 Expression analysis of Bf3149 in pET22b-Strep-SUMO vector

BF3149 sequence was analyzed to identify possible structures that could difficult recombinant protein expression: TMHMM 2.0^{119,120} (used to predict transmembrane domains using hidden Markov models) did not predict any transmembrane domains in BF3149 sequence; and SignalP 5.0¹²¹ (predicts signal peptides presence in protein sequences) did not predict any signal peptide in BF3149 sequence.

The full length protein was expressed as a C-terminal His-tagged protein in a pET28a vector. Expression in *E. coli* was explored at different conditions since yield with standard IPTG induction at 25 °C was low and the protein mainly remained in the insoluble fraction after cell lysis. Induction at different temperatures and co-expression with pGro7 vector¹²² encoding for GroEL and GroES chaperones were assayed. For cultures co-expressing pGro7, a previous 30 min incubation at 37°C with 1% L-arabinose when OD reached 0.5 (therefore, after 30 min it was at OD of 1 and IPTG was added). Plasmid expressing BF9343_3149 with chaperones from pGro7 also showed the expected size of 45 kDa, plus the band corresponding to the GroEL chaperone of 60 kDa (previous Figure 3.8). BF9343_3149 expression yields did not improve when co-expressed with chaperones, therefore, from that point onwards the preparative protein expression was performed with IPTG induction at 25°C with no chaperone co-expression.

Solubilization of the cell extract to isolate solubilized protein for purification was further explored. Cells were lysed by sonication in different buffers. Phosphate buffer alone and containing CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate), Tris buffer with DTT (dithiothreitol) and HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer containing CHAPS and glycerol were assayed. Only cell extract with buffers without detergent presented activity, indicating that the detergent could interfere with the active enzyme.

Despite the presence of active enzyme in the soluble cell extract with phosphate buffer, the purification by metal affinity chromatography did not work. Since activity was only found in the flowthrough and not in the elution step with imidazole, it was thought that the His-tag was not exposed or was hydrolyzed. Thus, it was decided to use another vector with a different affinity tag.

3.1.4.1 Cloning and expression

BF3149 gene was subcloned into a pET22b expression vector fused with Strep-tag at the N-terminus, and also the SUMO protein between Strep and BF3149 gene trying to enhance protein solubility as reported for other proteins^{123,124}. Figure 3.11 shows plasmid design of pET22b-Strep-SUMO-BF3149 and Figure 3.12 shows the CPEC assembly strategy followed for the construction of this vector.

Cloning was carried out using pET22b-Strep-SUMO-mKate2 as vector and substituting mKate2 gene for BF3149 (obtained from pET28a-BF3149) sequence by assembly cloning. Vector and sequence were

amplified using primers from Table 3.2 and, after PCR products showed the expected fragment size of 5.4 and 1.2 kb, respectively (Figure 3.13), they were cloned into the final vector by CPEC assembly.

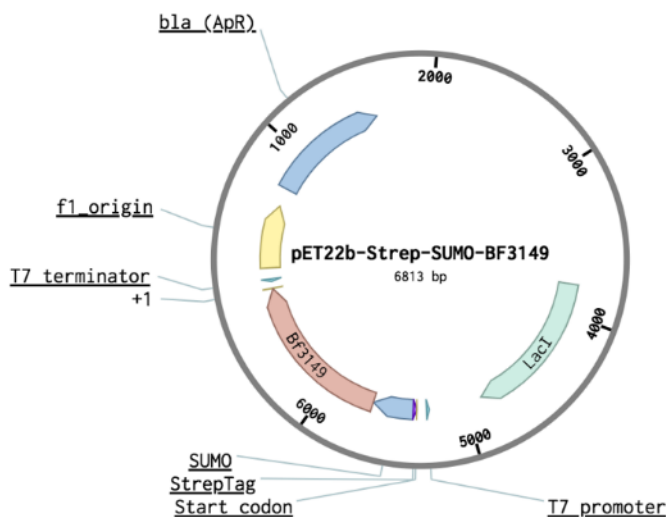


Figure 3.11 Vector map of pET22b-Strep-SUMO-BF3149. Vector elements are marked in colored arrows inside the circle of the plasmid that indicate sequence direction. In the center, the name of the vector and its size. Legend: Blue (top), Ampicillin R gene. Green, lacI. Blue, strong blue (short), T7 promoter, StrepTag, respectively. Blue, SUMO protein. Brown, BF3149 gene. Blue (short) T7 terminator. Green, f1 origin. pBR322 origin not shown.

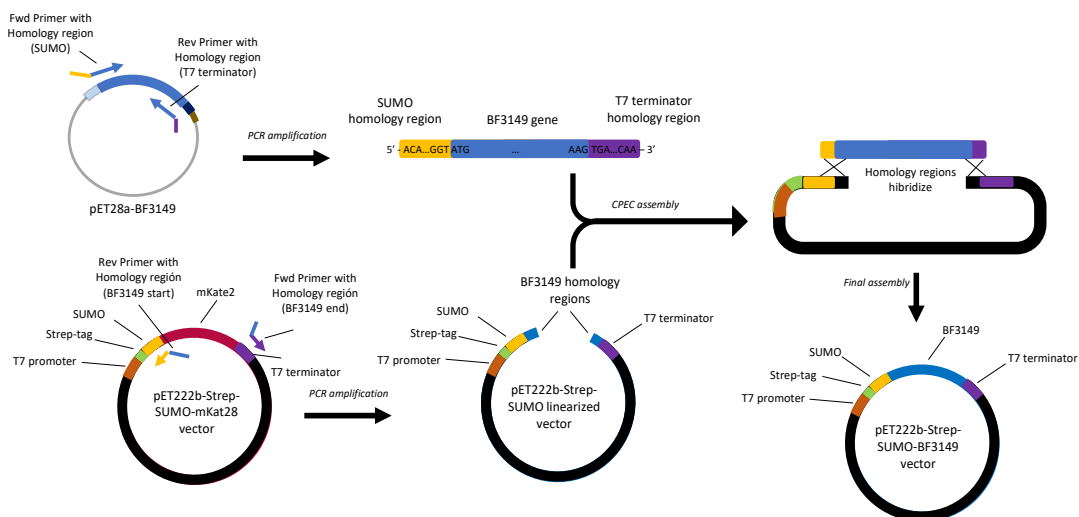


Figure 3.12 CPEC strategy followed for pET22b-Strep-SUMO-BF3149 construction. BF3149 gene was amplified using primers with 20 bp homology regions to SUMO and T7 terminator from pET28a-BF3149. pET222b-Strep-SUMO-mKate28 vector was linearized, removing mKate2 sequence, with homology regions to BF3149. Fragments were assembled by hybridizing homology regions.



Table 3.2 List of primers and templates used to amplify each fragment for pET22b-Strep-SUMO-BF3149. Table indicates the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences are listed in Table S3.1 from this chapter's Annexes.

Fragment	Template	Primers	Length of fragment (bp)
pET22b-Strep-SUMO backbone	pET22b-Strep-SUMO-mKate2	bb_N_BF3149_fw bb_N_BF3149_rv	5410
BF3149	pET28a-BF3149	N_BF3149_fw N_BF3149_rv	1189

After assembly, *E. coli* DH5 α was transformed with CPEC product and Colony PCR (with primers T7 promoter and T7 terminator) was used to analyze the presence of the vector. Positive colonies were sent for sequencing. Finally, a colony showing 100% homology with the expected sequence of the designed construct, was used to transform *E. coli* BL21(DE3)Star for expression analysis.

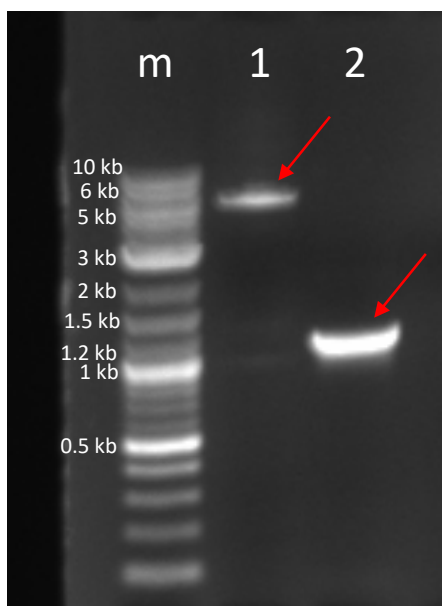


Figure 3.13 Agarose gel electrophoresis results of PCR for pET22b-Strep-SUMO-BF3149 construction. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows show fragments that have been amplified after PCR. Lane 1: Results of amplification of pET22b-Strep-SUMO vector backbone. Lane 2: BF3149 gene.

3.1.5 Comparison of activity between vectors with BF3149 and purified fraction

After successful cloning and purification of pET22b-Strep-SUMO-BF3149, activity was compared with pET28a-BF3149 vector. Cultures and cell disruption were carried out following the mentioned protocol for each of the constructs. Cell-free extracts were obtained, and extracts were set to a similar protein concentration before carrying out the ceramide-NBD assay.

Results in Figure 3.14 showed that the cell-free extract in phosphate buffer from pET22b-SUMO-Bf3149 showed 3.5 times higher activity than the previous one (pET28a-BF3149), proving the effectiveness of the SUMO fusion protein. After these results, it was proceeded to purify the enzyme with Strep-tag affinity chromatography.

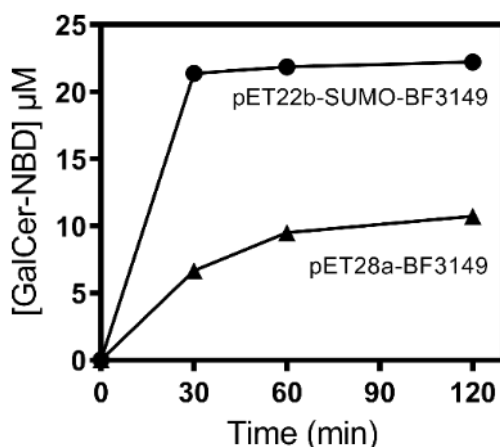


Figure 3.14 Ceramide-NBD activity assay of α -GalCer_{GT} (BF9343_3149) cell-free extracts from different expression systems. Reactions were carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and cell-free extracts of pET28a-BF3149 and pET22b-SUMO-BF3149 cultures (total protein concentration in the assay 9.9 ± 0.4 mg/mL) in 50 mM phosphate buffer, 150 mM NaCl at pH 7.5 and 37 $^{\circ}$ C.

3.1.5.1 Purification with Strep-tag

BL21(DE3)Star cells transformed with pET22b-Strep-SUMO-BF9343_3149 were grown in 600 mL LB containing ampicillin at a final concentration of 100 μ g/mL, inoculated at initial OD of 0.05 and incubated at 37 $^{\circ}$ C until OD = 1. At this point, protein expression was induced by adding IPTG to a final concentration of 0.1 mM and cultures were incubated at 30 $^{\circ}$ C for 4h. Cells were harvested and stored at -20 $^{\circ}$ C. Purification was carried out with the frozen pellets as described in the Material and methods chapter. Final eluted fraction was dialyzed with Amicon Ultra 15 10 kDa (Millipore) to remove d-desthiobiotin. Finally, it was concentrated, and volume adjusted to 1 mL with lysis buffer. Protein was quantified using the Bradford assay.

SDS-PAGE gel electrophoresis (Figure 3.15) was carried out using samples of soluble fraction, insoluble fraction, flow-through (FT), purified fraction (before and after concentration). Concentrated purified fraction showed the expected size of 56 kDa for Strep-SUMO-BF3149 fusion protein.

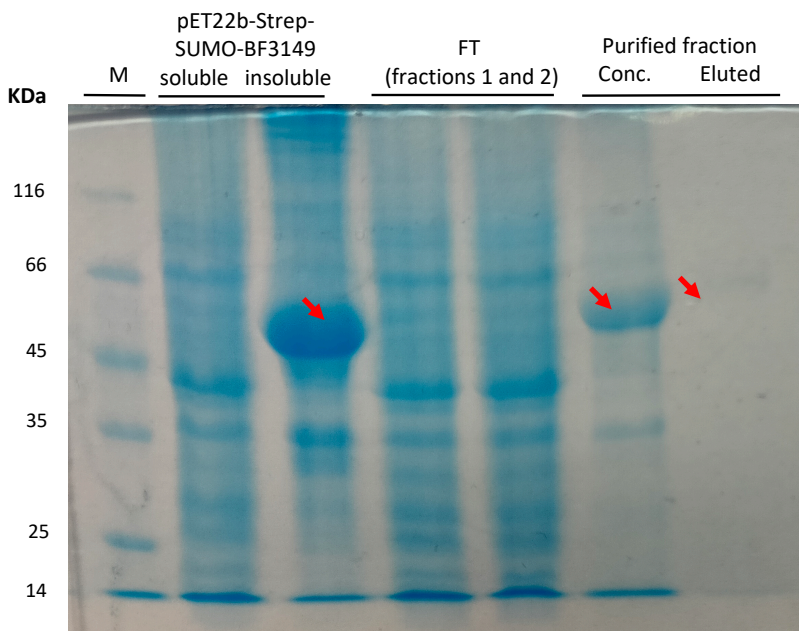


Figure 3.15 SDS-PAGE gel electrophoresis results of purification of construct Strep-SUMO-BF3149. Labels indicate the soluble and insoluble fraction for each culture; M: protein molecular mass marker. Arrows show the expressed proteins. Expected MW for Strep-SUMO-BF3149 was 56 kDa.

Therefore, purification using Strep-trap affinity chromatography was successful. The eluted protein after dialysis was about 0.39 mg/mL with an overall yield of 0.6 mg per liter of culture.



3.2 Characterization of Bf3149

Following successful purification of Bf3149 fused with SUMO protein, the next step was to characterize its biochemical features. There are two analysis approaches: *in silico* sequence and structure analysis and kinetic properties characterization. Each approach will give more information on different feature of Bf3149 protein.

All experiments in the characterization were carried out using galactosylceramide transferase activity assay with slight variations of the protocol for the determination of each parameter (described in Materials and methods chapter). Purified enzyme was used in all experiments, unless indicated otherwise. Strain culture, protein expression and purification were carried out following previous protocols (described in Materials and methods chapter).

At this point of the project, Okino *et al.*³⁸ already reported the identification of the α -GalCer synthase from *B. fragilis* and the characterization of its product analysis by HILIC-phase LC-ESI-MS/MS and comparison with C18- α -GalCer and C18- β -GalCer standards, which showed that the product of Bf3149 had a retention time that matched with C18- α -GalCer standard. After this initial identification of the enzyme activity by Okino *et al.*³⁸, here we report the biochemical characterization of the enzyme.

3.2.1 Specific activity

Specific activity was calculated by determining initial velocities of different reactions at different concentrations of purified enzyme. Reactions were monitored by HPLC and GalCer concentration was quantified from the chromatograms using the relative areas of substrate and product.

Reaction conditions were determined at 1.25 mM UDP-Gal donor and 25 μ M Cer-NBD acceptor solubilized as a complex with BSA in equimolar concentration (Cer-NBD:BSA 1:1) in phosphate buffer, pH 7.5, 37°C. Time course monitoring by HPLC showed that the enzyme synthesizes α -GalCer as unique product (Figure 3.16), and the reaction is complete after 15 min (Figure 3.17A).

The specific activity was calculated to be $3.27 \cdot 10^{-2} \text{ s}^{-1}$ from the linear dependence of initial velocity with enzyme concentration (Figure 3.17B).

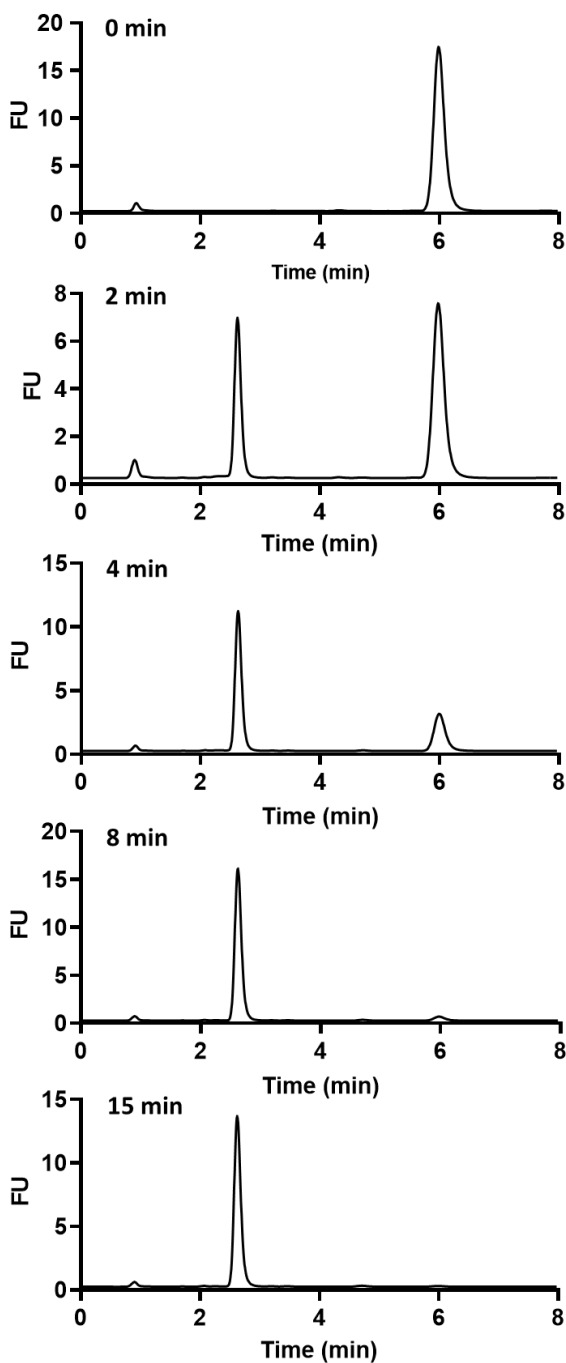


Figure 3.16 Time-course reaction monitoring of galactosyltransferase activity of purified α -GalCer synthase. The reaction was carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and, 2.75 μ M enzyme. in phosphate buffer at pH 7.5 and 37 $^{\circ}$ C.

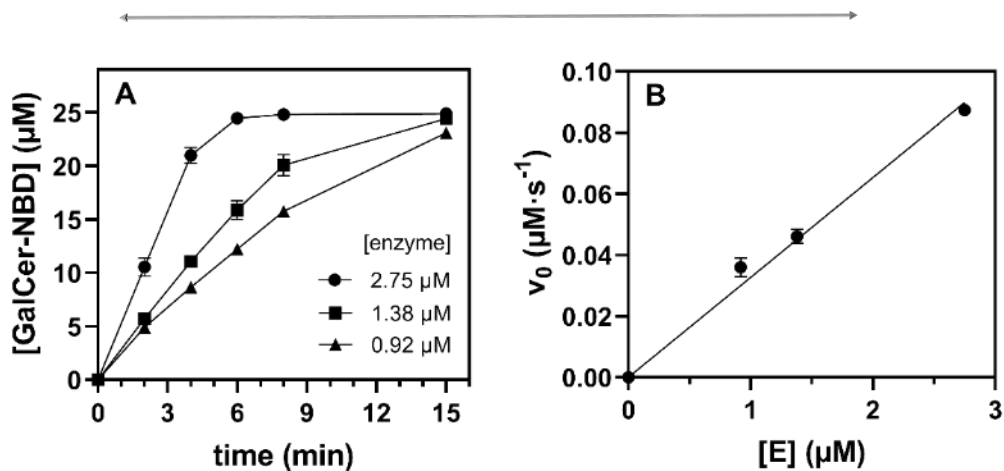


Figure 3.17 A) Time-course of BF3149 α -GalCer GT reaction. B) Linear dependence of initial rates with enzyme concentration. Reactions were carried out with 25 μM of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 1 to 3 μM enzyme in 50 mM phosphate buffer, 150 mM NaCl at pH 7.5 and 37 $^{\circ}\text{C}$.

3.2.2 Enzyme stability

Unexpectedly, the stability of the purified enzyme was a key issue. To evaluate the stability, assays were carried out at time 0 h after purification, and every 30 min until 2h, then at 4h, 6h (5h for Tris buffer) and 24h. Cell-free extract was also tested for stability. Reaction conditions were the same, but different buffers for protein solubilization were tested.

Enzymatic activity was only completely maintained for 2 h (points at 30, 60 and 90 min not included), about 39% at 6 h and approximately lost at 24 h. This behavior was similar in Tris buffer and also in the presence of 20% glycerol. On the other hand, in the cell extract, the enzyme was stable until 6 h and kept 42% of activity after 24 h (Figure 3.18).

Results indicate that the purified protein has a poor stability, therefore, all activity data presented is for freshly expressed and purified protein (less than 2 h storage after purification).

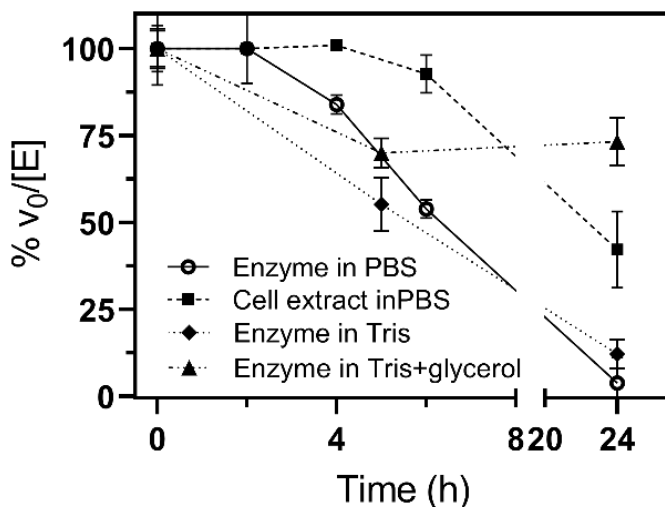


Figure 3.18 Stability of α -GalCer synthase from *B. fragilis*. The specific activity ($v_0/[E]$) of freshly prepared enzyme stored at 4°C in phosphate, pH 7.5 was measured at different times. Assay conditions: 25 μ M Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 1.39 μ M purified enzyme or soluble cell ex-tracts (0.77 μ M active enzyme), in 50 mM phosphate or Tris buffer, 150 mM NaCl at pH 7.5 and 37 °C.

3.2.3 Metal binding

Since in glycosyltransferases cation requirement is common, metal binding was also studied. Buffer used for protein purification and activity assays did not contain added metal cations and was changed from phosphate to Tris buffer since phosphate causes precipitation of cations. For each condition, a solution of cation was added to the final specified concentration. EDTA was also tested since it captures metal cations from the enzyme, causing a reduction of its activity if they are bound to the enzyme.

Results (Figure 3.19) showed that treatment with 200 mM EDTA did not reduce the activity relative to untreated and freshly purified enzyme indicating that the enzyme either does not bind divalent cations or that the metal cation is strongly bound to the active site. Mg^{2+} and Mn^{2+} at 0.25 mM slightly increased enzyme activity, but the effect was reduced at higher concentration, with some inhibition at 1 mM Mg^{2+} , which reduced activity to 70%. Whereas Ca^{2+} at 0.25 mM does not significantly affect the enzyme activity, Zn^{2+} inactivated the enzyme.

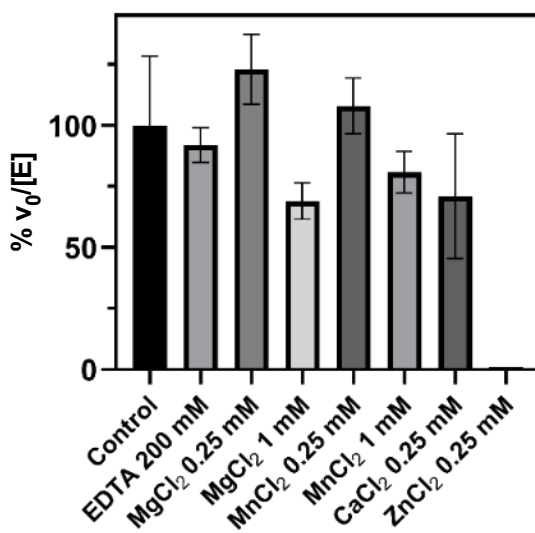


Figure 3.19 Effect of metal cations on α -GalCer synthase activity. Specific activities ($v_0/[E]$) were compared and plotted as % activity relative to the activity of freshly purified enzyme (control). Reaction conditions: 25 μ M Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 35-70 nM enzyme in 20 mM Tris buffer, 200 mM NaCl at pH 7.5 and 37 $^{\circ}$ C.

GT enzymes with a GT-A fold are metallodependent enzymes and exhibit the conserved DXD motif that binds a metal cation, which facilitates leaving group departure by coordinating the phosphate groups of the sugar nucleotide donor. In contrast, GT-B enzymes, such as α -GalCer synthase, do not have the DXD signature but instead use positively charged side chains and/or hydroxyls and helix dipoles to facilitate leaving group departure during catalysis⁴². Metal cations are not essential for enzymatic activity but in some cases reaction rates are accelerated by certain metal cations¹²⁵. Results show that in α -GalCer synthase there is no evidence of metal cation requirement, since EDTA did not reduce enzyme activity, and addition of Mg^{2+} or Mn^{2+} did not significantly affect the activity, but Zn^{2+} did inactivate the enzyme. However, related GT4 glycolipid synthases such as GlcAT from *Z. mobilis* and α -monoglucosyldiacylglycerol synthase from *A. laidlawii* did show metallodependent activity despite the lack of DXD motif^{38,117}.

Further studies should be considered, since it is possible that the metal cation is strongly bound to the enzyme. To test this hypothesis, an experiment of protein unfolding and refolding with EDTA to remove

possible cations should be performed. However, at this moment, this experiment is not an option since protein refolding would take a longer time, which would cause that the enzyme would already be inactivated due to its instability.

3.2.4 Optimal pH and temperature

Optimal pH and temperature for the enzyme was also determined. For each property, a set of different conditions at different pH and temperatures were set. For pH, citrate-phosphate buffer was used instead of phosphate buffer, so the pH range was widened. Two set of experiments with different pH but with some overlapping pH points were carried out. The curve was determined by combining results of both set of experiments, by normalizing the results calculating activity (initial velocity per enzyme concentration, $V_0/[E]$). For temperature, the same was done with two sets of experiments, with single and common temperature points. Curve was plotted by adjusting the points the same way as for pH.

The pH profile (Figure 3.20A) of specific activity using phosphate-citrate buffers follow a bell-shaped curve with the optimum catalytic efficiency at a pH around 7.3. Between 6 and 8.5, the enzyme retained more than 70% activity, while it is essentially inactive at pH < 5 and > 9.5.

As for temperature (Figure 3.20B), α -GalCer synthase remained catalytically active from 25°C to 40 °C with an optimal temperature around 30-35 °C, and it was nearly inactivated at 50 °C, and lost all activity at 55°C. This indicates a poor thermostability, which agrees with the observed poor stability of the enzyme even at storage at 4°C.

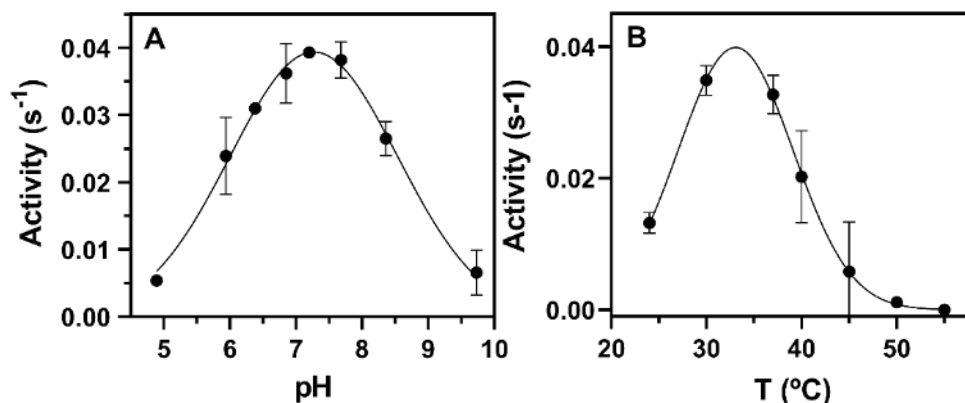


Figure 3.20 Optimal pH and temperature profiles of α -GalCer synthase activity. Reactions were carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and different enzyme concentrations. A) pH profile between 5 and 9.5 with 50 mM/50mM citrate/phosphate buffers adjusted at each pH, 37 °C and 830 nM enzyme. B) Temperature profile between 25 and 55 °C in phosphate buffer at 7.5 and 1.1-3.2 μ M enzyme.

3.2.5 Enzyme kinetics and substrate affinity

3.2.5.1 Michaelis-Menten curves for UDP-galactose and ceramide-NBD

Kinetics were determined by varying UDP-Gal at saturating Cer-NBD concentration (25 μM) solubilized with BSA, and varying Cer-NBD concentration at saturating UDP-Gal concentration (1.25 mM) at pH 7.5 and 37°C. As with pH and temperature experiments, different sets of conditions were tested, using common concentrations between sets to later normalize the results by calculating activity.

The results were plotted and generated a curve that obeyed Michaelis-Menten kinetics with some substrate inhibition (Figure 3.21). K_M for UDP-Gal at saturating ceramide-NBD concentration (25 μM), was calculated (using analysis tool of Prism 8 software) to be 108 μM , with a catalytic efficiency of 0.22 $\text{mM}^{-1}\cdot\text{s}^{-1}$ and a slight donor substrate inhibition can be observed above 2 mM.

For NBD-ceramide acceptor at saturating donor conditions (1.25 mM), kinetics shows substrate inhibition with maximum activity at 25 μM of Cer-NBD, K_M of 4.6 μM and K_I of 161 μM .

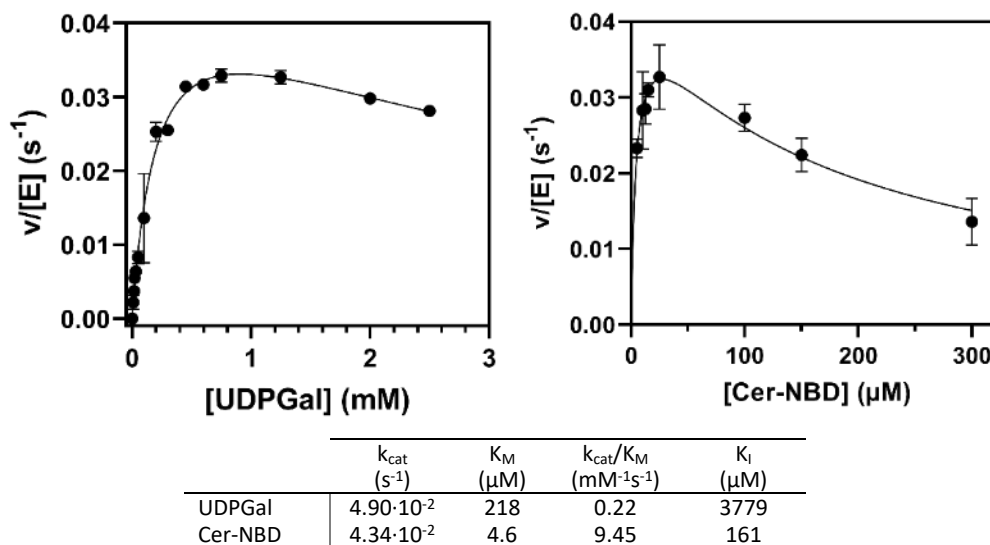


Figure 3.21 Kinetics of α -GalCer synthase. A) Michaelis-Menten curve for UDP-Gal. Reactions conditions: 10-2,500 μM UDP-Gal, 25 μM Cer-NBD in the presence of equimolar BSA, 0.2-0.5 μM enzyme, at pH 7.5 and 37°C. B) Michaelis-Menten curve for Cer-NBD. Reactions conditions: 1.25 mM UDP-Gal, 5-300 μM Cer-NBD in the presence of equimolar BSA, 0.35-0.75 μM enzyme, at pH 7.5 and 37°C.

3.2.5.2 Specific activity comparison: UDP-Gal vs. UDP-Glc

Okino *et al.*³⁸ already tested different substrates for α -GalCer synthase from *B. fragilis*. For acceptors, it appeared that only ceramide was accepted, but not diacylglycerol. For donors, UDP-Gal, UDP-Glc, UDP-GalNAc, UDP-GlcA, UDP-GlcNAc and GDP-Man were tested, but only UDP-Gal and UDP-Glc were accepted by the enzyme, with a preference for UDP-Gal.

In order to quantify this preference, specific activity was calculated at the same concentration of donor using both substrates. The experiment was carried out as with previous specific activity determination conditions, but at different enzyme concentrations.

Results showed that UDP-Gal presents 14.5-fold higher specific activity than UDP-Glc using NBD-ceramide as acceptor at 1.25 mM/25 μ M donor/acceptor ratio (Figure 3.22). Assuming saturating conditions for both donors and acceptor, UDP-Gal is a better donor than UDP-Glc in terms of k_{cat} .

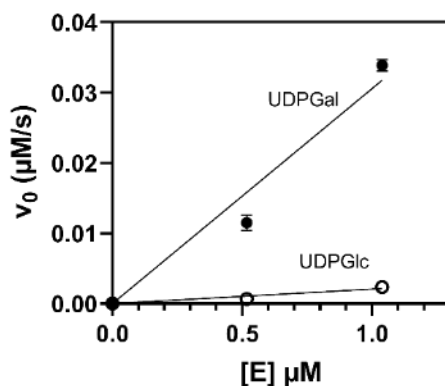


Figure 3.22 α -GalCer synthase activity on UDP-Gal and UDP-Glc donor substrates. Reactions conditions: 1.25 mM UDP-Gal or UDP-Glc, 25 μ M Cer-NBD in the presence of equimolar BSA, different enzyme concentrations (0.52 to 1.04 μ M) at pH 7.5 and 37°C. Specific activities: UDP-Gal, 0.031 ± 0.003 s⁻¹, UDP-Glc, 0.0021 ± 0.0003 s⁻¹.

3.2.6 Ceramide in mixed micelles and activator lipids

Although Okino *et al.* reported that ceramide with C12 chain length (C12-Cer-NBD) was the preferred acceptor, whereas 50% activity was observed with C6-Cer-NBD and no activity was detected with diacylglycerol (DG-NBD), the conditions on how the lipidic acceptor substrate was presented were not detailed. Other glycolipid synthases are membrane associated proteins, (i.e. monoglucosyldiacylglycerol synthases from *A. laidlawii*¹¹⁷ and *Streptococcus pneumoniae*¹¹⁸, processive MG517 from *Mycoplasma genitalium*⁸⁴ or PimA from *M. smegmatis*¹²⁶), which require an activation by anionic lipids, and are active with the lipid acceptor presented in mixed micelles.

As described above, Bf3149 α -Gal synthase kinetic characterization was performed with Cer-NBD solubilized with BSA. However, it was decided to evaluate the effect of presenting the acceptor Cer-

NBD in DOPC and/or DOPG mixed micelles (DOPC, dioleoylphosphatidylcholine as neutral lipid and DOPG, dioleoylphosphatidylglycerol as anionic lipid).

Micelles were prepared by adding the appropriate amount of DOPC and/or DOPG, and Cer-NBD, solubilized with chloroform, and mixed in a tube for each condition. Chloroform was evaporated under N_2 stream. Then, the lipid film was solubilized in phosphate buffer by vortexing extensively until no lipid film was observed. Finally, the solution with lipid mixture were treated for 30 min in an ultrasound bath to form micelles. Micelle formation was analyzed using ZetaSizer (Malvern Panalytical).

For each condition, Cer-NBD micelles were mixed with the enzyme, with the appropriate amount of buffer, and incubated on ice for 30 minutes. In addition, in the analysis there were also included: a control with Cer-NBD solubilized with BSA, and a reaction with Cer-NBD solubilized with BSA but also containing DOPC 1mM micelles without Cer-NBD. After incubation of enzyme with micelles, UDP-Gal was added to start the reaction. After 60 minutes, a fresh aliquot of enzyme was added to evaluate the effect of fresh enzyme.

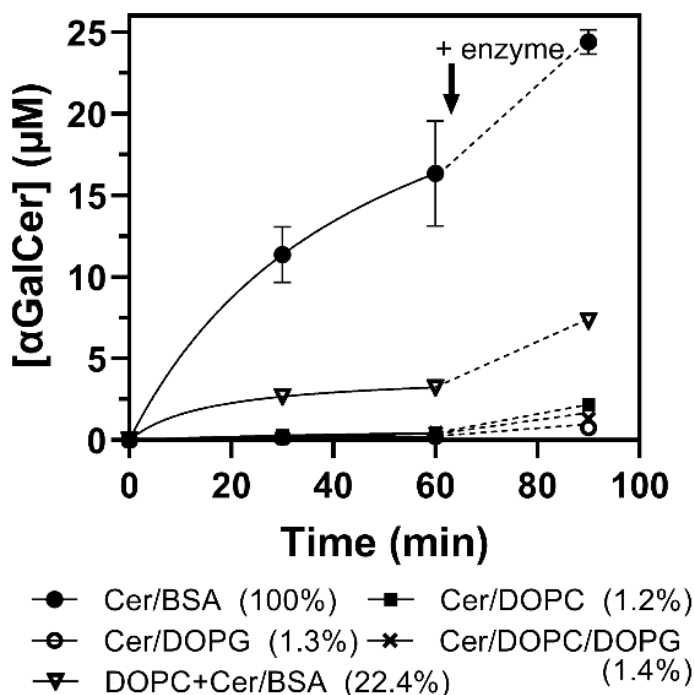


Figure 3.23 Reaction monitoring of α -GalCer GT activity using ceramide in mixed micelles as acceptor. Reactions were carried out with 25 μ M of Cer-NBD, 1.25 mM of UDP-Gal, 0.25 μ M en-enzyme. Ceramide was solubilized with BSA or in mixed micelles with DOPC (1 mM) or DOPC/DOPG 0.6/0.4 (total 1 mM) or DOPG (1 mM). Also, activity was measured with Cer-NBD in BSA in the presence of DOPC (1 mM). Legend (in parenthesis): % activity at 30min. At 60 min reaction, a fresh aliquot of enzyme was added (indicated with an arrow).

Results (Figure 3.23) show that enzyme activity decreases considerably in ceramide/DOPC or DOPG micelles as compared with Cer-NBD solubilized with BSA. Neither the presence of neutral nor anionic



lipids in vesicles activate the enzyme, but rather inhibit since addition Cer-NBD solubilized with BSA to the enzyme preincubated with DOPC vesicles only results in 22% of the original activity. Therefore, these findings indicate that Bf3149 GT is fully active with the ceramide acceptor solubilized with BSA, but activity is largely reduced when using ceramide in lipid vesicles and no activation by anionic lipids was observed.³⁸

Addition of fresh enzyme did not show any conclusive results. Despite the addition of fresh enzyme (not previously incubated with DOPC/DOPG micelles) seems to show higher initial velocity on time 60-90 min than 0-30 min, it is not clear if the higher velocity is due to higher enzyme concentration or adding enzyme free of detergent.

These results are in accord with previous results with buffer containing CHAPS, a detergent used to improve protein solubility, which proved to inhibit Bf3149 α -GalCer synthase activity. It seems that the enzyme is not membrane-associated since the presence of detergents in the lysis buffer when extracting the protein did not improve protein solubilization. Altogether these results suggest that the enzyme takes the lipidic acceptor presented by a lipid-protein complex rather than from membranes, however, it requires further studies to elucidate the mechanism.

3.2.7 Structure analysis

A structural analysis of Bf3149 was performed to evaluate the topology of the active site and identify key residues interacting with UDP-galactose. The analysis is carried out using the three-dimensional AlphaFold model.

Bf3149 amino acid sequence was aligned with GT4 enzymes with solved crystal structure (Table S3.2 and Figure S3.2 this chapter's Annexes). In addition, the AlphaFold model of the α -GalCer GT was structurally aligned with the 3D structures of solved GT4 enzymes (Figure 3.24). Bf3149 GT does not differ from the typical GT-B fold of GT4 enzymes, comprising two separate $\beta/\alpha/\beta$ Rosmann-fold domains that form an interdomain substrate-binding crevice.

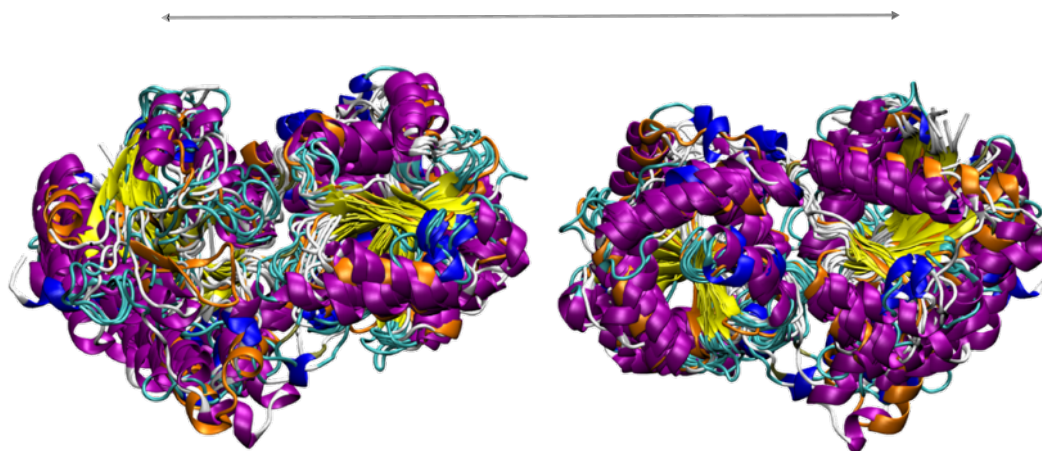


Figure 3.24 Structural alignment of 10 selected GT4 enzymes (2JIM, 5D00,3OKA, 2IV7, 2GEJ, 4XSU, 5N80, 6D9T, 2IUU, 5ZFK from Table S3.2 in this chapter's Annexes) and the α -GalCer GT (orange). Figures generated with VMD¹⁵¹.

To locate the UDP-Gal donor binding site in the enzyme model, docking experiments were performed, highlighting first shell amino acid residues (Figure 3.25). The closest protein with solved X-ray structure is the phosphatidylinositol mannosyltransferase (PimA) from *Mycobacterium smegmatis* (Uniprot: AOQWG6) with 42.4% similarity and 24.4% identity (Table S4.2 in this chapter's Annexes). Structural alignment of the modeled α -GalCer GT with UDP-Gal ligand with the 3D structure of a complex PimA with bound GDP-Man substrate¹²⁶ allowed to compare the similar binding site topology and first shell amino acid residues in the binding site (Figure 3.26). Residues of α -GalCer GT that interact with the UDP nucleotide moiety comprise Val209, Arg269, Val294 and Glu297 (uracil), Lys216, Phe284 and Leu293 (ribose), whereas Thr19, His117, and Asp289 interact with the galactose unit. The enzyme conserves the EX₇E motif (Asp289 and Glu297) typical of GT4¹²⁷, but with an Asp for the first Glu. Acceptor docking experiments have not been attempted because the 3D enzyme model corresponds to the free enzyme, and it is expected a domain closure upon acceptor binding as often observed in GT enzymes with GT-B fold¹²⁸.

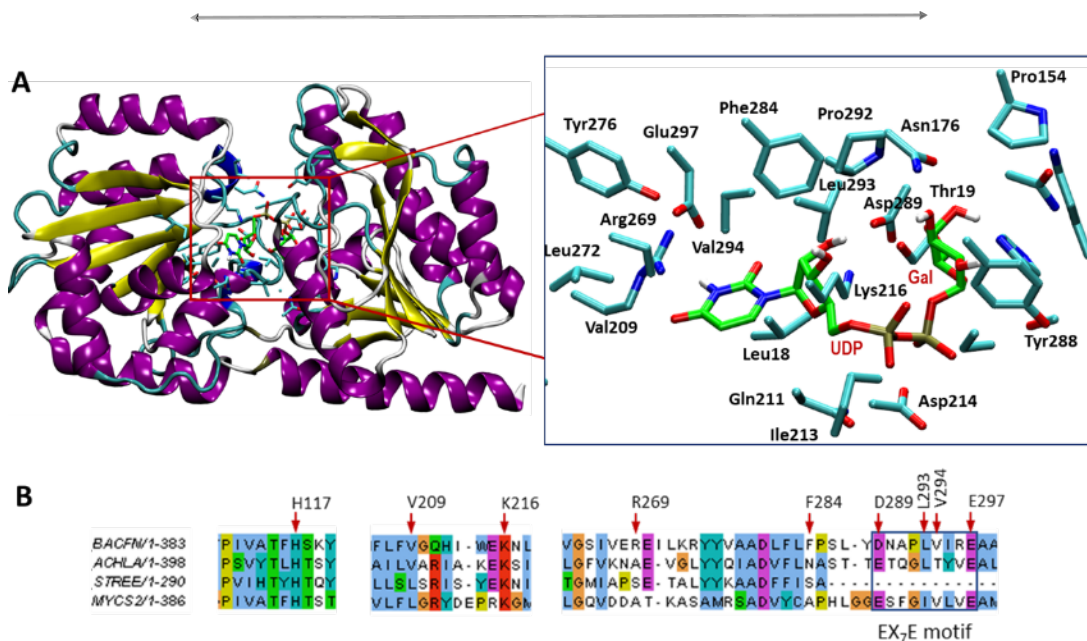


Figure 3.25 UDP-galactose donor location in modeled α -GalCer GT with identified interacting residues. A) Structural model of α -GalCer GT with bound UDP-Gal, generated by docking UDP-Gal into the AlphaFold model of the free enzyme (retrieved from Uniprot A0A380YRQ3). Right, magnification of the donor binding site, showing the amino acid side chains of first shell residues interacting with UDP-Gal. B) Sequence alignment of regions surrounding donor binding residues, GT4 glycolipid synthases from *B. fragilis* (α -GalCer GT, top), *A. laidlawii*, *S. pneumoniae*, and *M. smegmatis*.

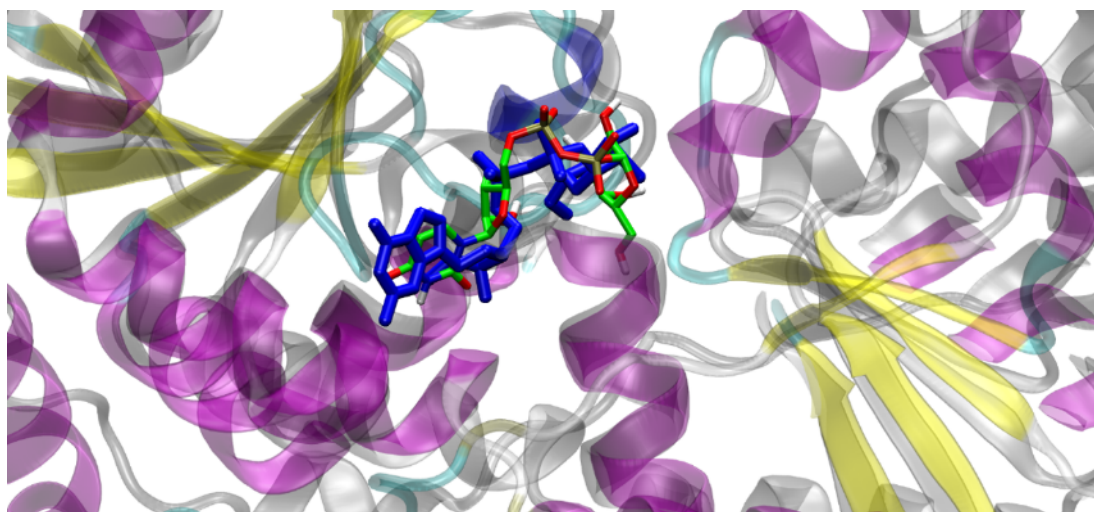


Figure 3.26 Structural superposition of the modeled α -GalCer GT with bound UDP-Gal (colored by secondary structure) and PimA from *Mycolicibacterium smegmatis* with bound GDP-Man (PDB 2GEJ) (silver; GDP-Man in blue). Figure generated with VMD¹⁵¹.



3.3 Summary

This chapter has covered the identification and characterization of a α -GalCer synthase from *B. fragilis*. Identification was carried out by *in silico* approach. From 83 putative glycosyltransferases annotated in the CAZY database, four candidates were finally selected from GT4 family, which contains GT that use ceramide or diacylglycerol as acceptors. Only one of those candidates, BF9343_3149, which has been previously identified by Okino *et al.*³⁸, proved to have α -galactosylceramide synthase activity.

Further characterization showed that α -GalCer synthase Bf3149 is a non-processive GT, maintains the α linkage of UDP-Gal in the α -GalCer product, which means it matches the retaining mechanism of action of GT4 family³⁸.

Although the protein expressed mainly as an insoluble protein, fusion with SUMO protein improved its solubility. Moreover, using Strep tag proved to have a higher efficacy to obtain pure enzyme, which yielded about 0.6 mg/L of culture. However, the purified protein was found to be quite unstable, losing its activity after 24h.

Enzyme shows maximal activity at pH 7.3 and 30-35°C and fully inactive at 55°C indicating poor thermostability in agreement with the observed instability upon storage even at 4°C.

Although GT enzymes with a GT-A fold are metallodependent enzymes with the conserved DXD motif that binds a metal cation, GT-B enzymes do not have the DXD signature but may still show metallodependent activity, such as close glycolipid synthases as GlcAT from *Z. mobilis* and α -monoglucosyldiacylglycerol synthase from *A. laidlawii*. Results, however, indicate that none of the metal cations tested improved enzyme activity, although Zn^{2+} showed to inactivate the enzyme. Furthermore, EDTA did not reduce the activity. This evidence indicates that α -GalCer synthase from *B. fragilis* does not have a metal cation requirement.

Substrate affinity analysis proved that the enzyme prefers UDP-Gal over UDP-Glc as donor substrate (showing 14.5-fold higher specific activity for the first). This supports results from Okino *et al.*³⁸ publication, in which those were the only substrate accepted by α -GalCer synthase, tested together with UDP-GalNAc, UDP-GlcA, UDP-GlcNAc and GDP-Man.

As for lipid acceptor, it was also reported that Bf3149 synthase used ceramide, and C12 chain length was preferred over C6 ceramide, but with diacylglycerol (DAG) no activity was detected. Our results prove that presence of detergents inhibit enzyme activity (CHAPS, DOPC and DOPG have been tested), and the enzyme takes the lipidic acceptor presented by a lipid-protein complex rather than from membranes (ceramide-BSA was used), but it requires further studies to elucidate the mechanism. This could indicate that the enzyme is not membrane-associated. Activation by anionic lipids was not observed, contrary to other glycolipid synthases that require this activation.

Structural analysis of the modelled Bf3149 α -GalCer synthase showed the enzyme has the common GT-B fold structure and the typical EX₇E motif essential for activity of GT4 enzymes¹²⁷, but with an Asp for the first Glu.



Finally, the results of the biochemical characterization of α -GalCer synthase from *B. fragilis* prove that there are more studies that should be done, since some features of the interactions of this enzyme are still to be elucidated. These future results could help improving features such as protein stability over time, thermostability, and even use of other lipid acceptors such as diacylglycerol, which could be complexed with BSA to test its use by the enzyme.



3.4 Annexes Chapter 3

Table S3.1 List of primers used in Chapter 3 and their sequences. Table indicates name of the primer as mentioned in the chapter, sequence, and length (Lngth). Underlined nucleotides indicate the added homology region for assembly.

Name	Sequence	Lngth
pET28a_fw	CTCGAGCACCACCACCACCACCTG	26
pET28a_rv	GGTATATCTCCTTCTTAAAGTTAAAC	26
Q5LFK6-fw	<u>GTTTAACTTTAAGAAGGAGATATACCATGCGTGATGGTAAACCTATTGAA</u> ATC	53
Q5LFK6-rv	<u>GTGGTGGTGGTGGTGGTGCTCGAGCTTGTGCAACAAGTTCAAGTAGATA</u> CG	51
Q5LAE3_fw	<u>GTTTAACTTTAAGAAGGAGATATACCATGATTGGTTTGTTAATGATTG</u>	49
Q5LAE3_rv	<u>GTGGTGGTGGTGGTGGTGCTCGAGCTTGTTGCCGTTACGCTTGATCAATC</u>	50
Q5L962_fw	<u>GTTTAACTTTAAGAAGGAGATATACCATGAAGAAGGTTTTATTCCAATC</u>	50
Q5L962_rv	<u>GTGGTGGTGGTGGTGGTGCTCGAGGTAGAACAGACGGTAGATACGGTT</u> ACG	51
Q5LHL6_fw	<u>GTTTAACTTTAAGAAGGAGATATACCATGAATGTTAACAACCTACTACTG</u>	51
Q5LHL6_rv	<u>GTGGTGGTGGTGGTGGTGCTCGAGGTACTCTTCGATCTTGCCAACCTGGT</u> AC	52
bb_N_BF3149_fw	<u>TCAAGCGTAACGGCAACAAGTGAGATCCGGCTGCTAACAAAG</u>	42
bb_N_BF3149_rv	<u>TCATTAACAACAACCAATCATACCACCAATCTGTTCTCTG</u>	39
N_BF3149_fw	<u>ACAGAGAACAGATTGGTGGTATGATTGGTTTGTTAATGATTG</u>	43
N_BF3149_rv	<u>TGTTAGCAGCCGGATCTCACTTGTGCCGTTACGCTTG</u>	39
T7 promoter	TAATACGACTCACTATAGGG	20
T7 terminator	GCTAGTTATTGCTCAGCGG	19

Chapter 3. Identification and characterization of an α -galactosylceramide synthase

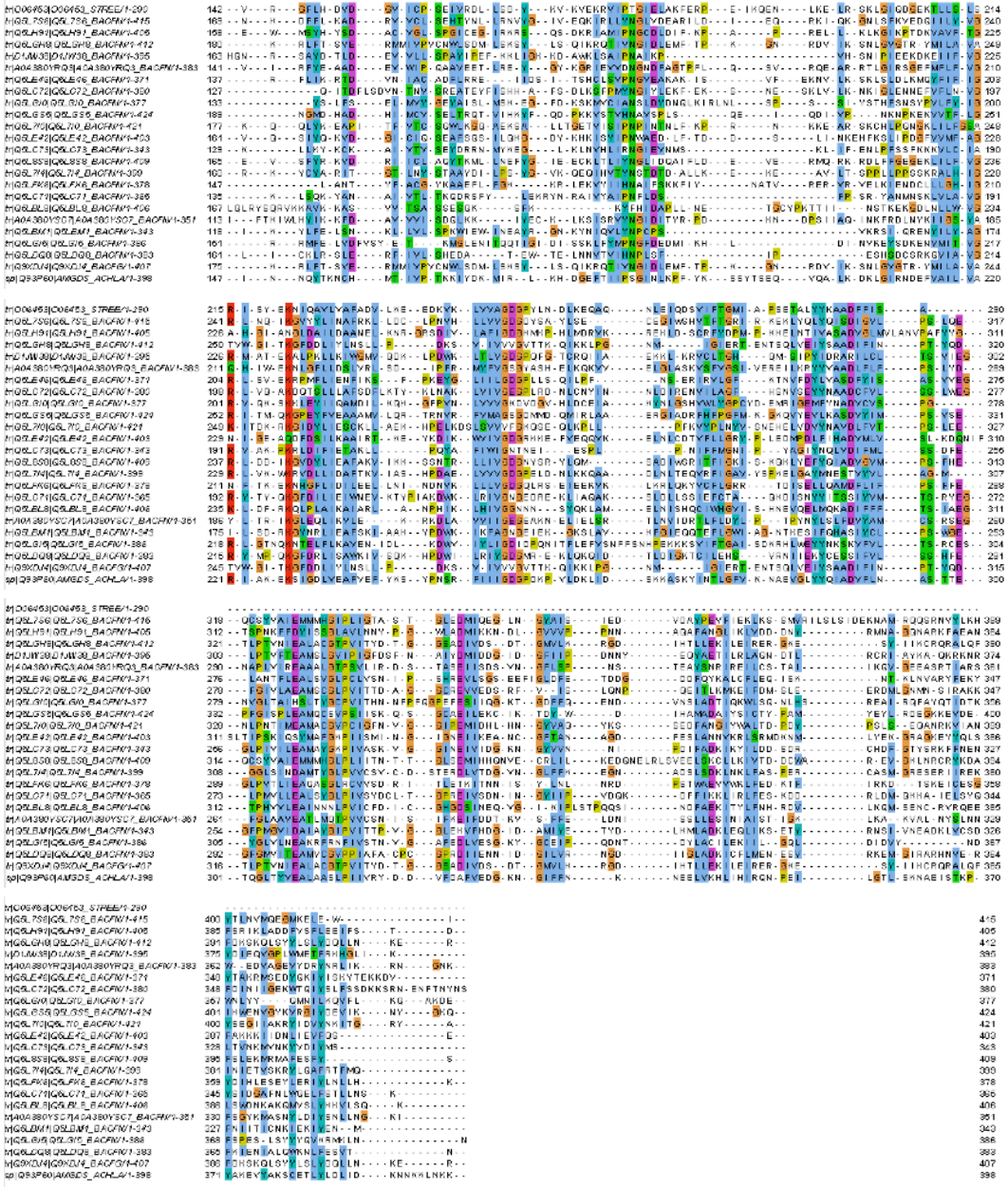


Figure S3.1 Multiple sequence alignment of annotated GT4 enzymes from *Bacteroides fragilis* and query sequences of monoglucosyl diacylglycerol synthase from *Streptococcus pneumoniae* [UniProt O06453] and *Acholeplasma laidlawii* [UniProt Q93P60] (first and last sequences in the alignment, respectively). Alignment with T-Coffee using JalView¹²⁹.

Chapter 3. Identification and characterization of an α -galactosylceramide synthase



Table S3.2 GT4 enzymes with solved crystal structure (October 2022). Proteins included in the sequence alignment (Figure S3.2) and structural alignment (Figure 3.24) are indicated in columns “Sequence alignment” and “Structural alignment”, respectively.

Name	Organism	Uniprot	PDB	Ligand	Resolution	Seq. Align. ¹	Struct. Align. ²
UDP-GlcNAc L-malate α -N-acetylglucosaminyltransferase	<i>Bacillus anthracis</i> str. Ames	Q81ST7	2JIM	-	3.10	YES	YES
UDP-GlcNAc L-malate α -N-acetylglucosaminyltransferase	<i>Bacillus anthracis</i> str. Sterne	Q81ST7	3BMO	UDP and L-malate	3.31	-	REMOVED
UDP-GlcNAc L-malate α -N-acetylglucosaminyltransferase	<i>Bacillus subtilis</i> subsp. <i>subtilis</i>	P42982	5D00	N-acetylglucosaminyl-malate and UMP	2.15	YES	YES
UDP-GalNAc: α -1,4-N-acetylgalactosyltransferase	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	O52905	6EJI	-	2.30	YES	YES
UDP-GlcNAc: 1L-myo-inositol-1-P α -N-acetylglucosaminyltransferase	<i>Corynebacterium glutamicum</i> ATCC 13032	Q8NTA6	3C48	-	2.10	YES	YES
			3C4V	UDP and 1L-INS-1-P.	2.60	-	YES
α -mannosyltransferase	<i>Corynebacterium glutamicum</i>	Q8NNK8	3OKA	-	2.20	YES	YES
UDP-GalNAc: GalNAc-PP-Und α -1,3-N-acetylgalactosaminyltransferase (WbnH)	<i>Escherichia coli</i> O86:H2	PODMP6	4XYW	-	2.20	YES	YES
UDP-Glc: L-glycero-D-manno-heptose II α -1,3-glucosyltransferase I (WaaG, RfaG)	<i>Escherichia coli</i> str. K-12 substr. MG1655	P25740	2IV7	-	1.6	YES	YES
FTT1235c (LpcC)	<i>Francisella tularensis</i> subsp. <i>tularensis</i> SCHUS4	Q5NFJ9	5I45	-	1.35	-	REMOVED
sucrose phosphate synthase (SpsA)	<i>Halothermothrix orenii</i>	B2CCB8	2R60	-	1.80	YES	YES
			2R68	P-(0-6)- β -D-Fruf-(2-1)- α -D-Glcp	2.40	-	YES

Chapter 3. Identification and characterization of an α -galactosylceramide synthase

cholesterol α -glucosyltransferase (HP0421)	<i>Helicobacter pylori</i> 26695	O25175	3QH P	-	1.50	YES	REMOVED
phosphatidylinositol mannosyltransferase	<i>Mycobacterium smegmatis</i> MC2 155	A0QWG6	2GEJ	GDP-Man	2.60	YES	YES
			4N9 W	-	1.94	-	YES
ADP-Glc-dependent α -maltose-1-phosphate synthase	<i>Mycobacterium smegmatis</i> MC2 155	A0R2E2	6TVP	-	1.90	YES	YES
sucrose synthase	<i>Nitrosomonas europaea</i> ATCC 19718	Q820M5	4RBN	-	3.05	REMOVED	REMOVED
UDP-Glc: [heterocyst envelope polysaccharide] mannoside glucosyltransferase	<i>Nostoc sp.</i> PCC 7120 = FACHB-418	Q8YQW3	4XSO	-	2.01	YES	YES
			4XSP	-	2.15	-	YES
			4XSR	UDP-Glc	2.39	-	YES
			4XSU	α -D-Glcp + UDP	2.48	-	YES
RAF_ORF0434	<i>Rickettsia africae</i> ESF-5	C3PN56	7MI0	-	2.90	NO	YES
LPS 1,6-galactosyltransferase (RfaB)	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> str. LT2	Q06994	5N80	UDP	1.92	YES	YES
			6Y6I	UDP	1.92	-	YES
UDP-GlcNAc: teichoic acid α -N-acetylglucosaminyltransferase	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> 21178	A0A0H2W WV6	4X6L	UDP	3.19	REMOVED	REMOVED
			4X7 M]	UDP-GlcNAc	2.40	-	-
			4X7R	α -GlcNAc-glycerol and UDP	2.15	-	-
UDP-GlcNAc: teichoic acid α -N-acetylglucosaminyltransferase	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	A0A0D6HU AO	4WA C	-	2.40	REMOVED	REMOVED
		-	4WA D	UDP-GlcNAc	2.80	-	REMOVED

Chapter 3. Identification and characterization of an α -galactosylceramide synthase

Serine-O- α -N-acetylglucosaminyltransferase	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	A0A0H2UR G7	7EC1	-	1.85	REMOVED	REMOVED
			7EC3	β -D-GlcpNAc	2.50	-	-
			7EC6	ASP-SER-ASP	1.90	-	-
			7VFL	β -D-GlcpNAc	2.45	-	-
			7VFM	UDP and SD peptide-binding form	2.28	-	-
UDP-GlcNAc: L-malate α -N-acetylglucosaminyltransferase	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i>	A0A068A5A 2	6D9T	UDP	2.00	YES	YES
			6N1X	α -D-GlcpNAc	2.35	-	YES
UDP-GlcNAc: [Ser/Thr] O-N-acetylglucosaminyltransferase	<i>Streptococcus gordonii</i> M99	Q9AET5	5E9T	-	2.92	REMOVED	REMOVED
			5E9U	β -D-GlcpNAc	3.84	-	-
UDP-GlcNAc: protein [Serine] α -N-acetylglucosaminyltransferase	<i>Streptococcus pneumoniae</i> TIGR4	A0A0H2UR G7	4PQG	β -D-GlcpNAc	2.00	REMOVED	REMOVED
eurekanate-attachment enzyme (AviGT4)	<i>Streptomyces viridochromogenes</i> TUE57	Q93KV2	2IUUY	-	2.1	YES	YES
UDP-Glc: tetrahydrobiopterin α -glucosyltransferase	<i>Synechococcus elongatus</i> PCC 7942 = FACHB-805	Q31LX1	5ZE7	-	1.99	YES	YES
			5ZFK	UDP-BH2 complex	1.75	-	YES
sucrose-phosphate synthase (TeSPS; TII1590)	<i>Thermosynechococcus vestitus</i> BP-1	Q8DIJ5	6LDQ	-	1.92	REMOVED	REMOVED
UDP-Glc: α -1,2-glucosyltransferase (TII1591)	<i>Thermosynechococcus vestitus</i> BP-1	Q8DIJ4	7FG9	-	2.66	YES	YES

¹ Sequence alignment. A total of 26 GT4 were first selected. 6 were removed because of poor alignment. YES are the 18 sequences included in the sequence alignment in Figure S3.2.

² Structural alignment. Protein sequences removed from the sequence alignment were also removed from the structural alignment. Additionally, structures labeled "REMOVED" in red were also discarded because of missing residues in the X-ray structure. Structural alignment in Figure 3..

Chapter 3. Identification and characterization of an α -galactosylceramide synthase

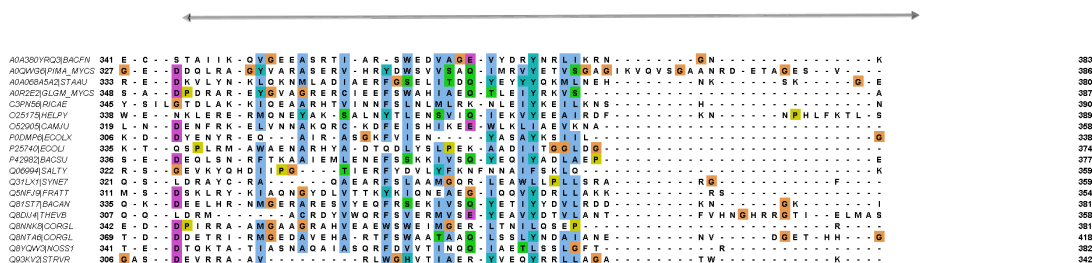


Figure S3.2 Sequence alignment of GT4 enzymes. Selected sequences from Table S3.2 (18 sequences + BF9343_149 (Uniprot A0A380YRQ3, *B.fragilis* α -GalCer_GT, on top). Alignment with T-Coffee using JalView¹²⁹.

Table S3.3 BF9343_149 gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	BF3149		
Locus Tag	BF9343_149		
DNA sequence	Length	1152	bp
<p>ATGATTGGTTTGTTTAATGATTGCTTCCACCTATCATGGACGGTGTTCCTTTGACTACTCAAATTACGCTTACTGGTTG CATCGTAAGGCTGGTAATGTTTGTGTTGTTACTC AAAATCTCCAGATGCTCGTGATGCTGAAGAATACCCAGTTTACC GTTACTCCTCTGTTCCAATCCAATGCGTAAGCCATACCGTTTGGGTTTCCCACGATTGATTGGCCATTCCACGAACTG ATCTCTCGTTTGTCTTTCGAATTGGTTCATGCTCATTGTCCATTTCTTCAGGTGCTTTGGCTATGCAAATGCCCGTGAA CAACATGTTCCAATCGTTGCTACTTCCACTTAAGTACCGTGCTGATTTCGAACGTGCTATTCCATCCCGTCTGTTGGTT AACTACTTGATCCGTAAGGTCATCCGTTTCTACGAAGCTGCTGATGAAGTTTGGATTCCACAAGCTGCAGTTGAAGAAA CTTTGCGTGAATACGGTTACAAGGTCGTATCGAAGTTGTTGATAACGGTAATGATTTCGCTGGTACTCCATTCTTGCA ATCCGTTTCGTAAGAAGCTCGTCGTACCTGGGATTCTGTTCTGGTGAATTCATGTTCTTGTTCGTCGGTCAACATATCT GGGAAAAGAATTTGGGTTTCTTGTGGACTCCTTGGTTCTGTTGCTGATATCCATTCCGTATGACTTCTGTTGTTCT GGTTATGCCTCTCATGAATTGAAGCAAAGGTTGTCGAATTTGGGTTTAGCCTCTAAGTTTCTTTTGTGGTTCCATCGT TGAGCGTGAATTTTGAAGCGTTATTACGTTGCTGCCGACCTGTTTTGTTCCCATCATTATGATAACGCCCATTTGG TTATTTCGTAAGCTGCCGTTTGGGTACACCATCTGTTTGTATCGTGATTCAACCGCTCCGAAATATCTCCGATTCTG TTAATGGTTTCTGTCTCAAATCCACTGAAGCTTCACTCAACCGTATTCGTGAAATCTGTGTTCCACGCCATTATTA AGCAAGTTGGTGAAGAAGCCTCTCGTACCATTGCTTCTTGGGAAGATGTTGCCGGTGAAGTTTACGATCGTTACAA CAGATTGATCAAGCGTAACGGCAACAAGTGA</p>			
Protein sequence	Length	384	aa
<p>MIGLFNDCFPPIMDGVSLTTQNYAYWLHRKAGNVVTPKSPDARDAEYYPVYRYSSVIPMRKPYRLGFPRIDWPFHERIS RLSFLVHAHCPFSSGALAMQIAREQHVPIVATFHHSKYRADFERAIPSRLLVNYLIRKVIREFYAADEVWIPQAAVEETLREYGY KGRIEVDVNDGDFAGTFLQSVRQEARRTLGIKSGEFMFLVGVQHIWEKNLGLFLDSLVRSLDIPFRMYFVGSYASHELKQ KVVELGLASKVSFVGSIVERILKRYVAADLFLFPLSDYDAPLVIREAAALGTPSVLIRDSTASEIISDSVNGFLSPNTEAYSNRI REILCSTAIKQVGEASRTIARSWEDVAGEVYDRYLRNLRKNGNK*</p>			

Table S3.4 BF9343_1306 gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	BF1306		
Locus Tag	BF9343_1306		
DNA sequence	Length	1137	bp
<p>ATGCGTGATGGTAAACCTATTGAAATCTTGCACATCGTTGGTACTCGTCCAGTTGGTGGTATTGGTGCTTTGTTGAAGA ACATTAACACTCCATCGACTTGAACAAGTTCCATTTCACTTACGTTTTCTCCGCCGATTCAACATTGGTGATTTGCGATA ACTACCTCCGTAAGTTGGTTCGATATCGTTGTTTCCATCTACCATCTGAAGTACCTGTTCTTGACTTTGAAGTTA TCTTCTGCTTCTACAAGCGTAAACGCAAAAAGACGATATCCATCTGTTTCATCTGCCAACACTGGTGTTCGGATTGG TTGTTTGCTAAGATCTACGGTATCCGATCCGTTCTTGTGATTCTCATTACCAAGTACTCTCCAAGAAGATCCGTTCC</p>			

<p>ATCCGTAAC TATTTCTTGAATCCCAACTATCTACTTGGCCAATACTTACTTTGCTTGTGGTTACAAGGCTGCCGAATTTGTTTGGTAAAAAGCGTTTGGAGAAGGTCTACATTATTCACAACGCCATCTTCTCCAAAAAGTTACTCTACAACGCTACCGTTCGTGAACGTGTTGTCGAATTGAAGATCGAAAACGACTGTTTGTGTTGGGTCATATTGGCAACTTACCCAAA GAAAAGAATCACGGTTTCTGATCGACATCTTGAAGAGTTGTTGAACATCAACGATAACGTCAAGTTGCTGTTGGTTG GTGATGGTCAATTGCGTCCGAAATTGAAGAAAAGGTCAAGTTGAAGCGTCTGCAGAAGTACGTTTGTGTTTTGGGTC GTCGTACCGACATCTGAATTATTGCAAGCAATGGACTTTTTGATCTCCCATCATTTTTGAGGGTTGCCAGTTACCT TGATTGAAGCTCAAGCTTCTGGTTGCGTTGCGTTGTTTCTGATCGTATCACCTTGGAACTAAGATTACCAACAACATC TCCTACTGAACTGAACCGTGATCTTACTTGGGCTGAAGTTGTTTGAAGTTGTTGGAAGATAAGACCTTCATCCG TAAGGACACCTCCAAAGAAATTTGCGAATCCGGTTACGATATCCACTTGAATCTGAATACTTGGAGCGTATCTACTTG AACTTGTGCAAGTGA</p>			
Protein sequence	Length	379	aa
<p>MRDGKPIELHIVGTRPVGIGALLKNINTSIDLNKFHFTYVFSADSNIGDFDNYVRKLGSDIVVFPYHLKYLFLYLVKVFICYKR NAKKYDIHVHSANTGVLDLLFAKIYIRIRILHSHSTKYSSKIRSNRYFLQPPTIYLANITYFACGYKAAEFVFGKKRLEKVYIHN AIFSCKFIYNATVRERVRVELKIENDCLLLGHIGNFTKEKNHGLIDILELLNINDNVKLLLVGDGQLRSEIEEKVKLRQLQKVC FLGRRTDISELLQAMDFLIFPFFELPVTLIEAQASGLRCVSDRITLETKITNNISYLNLRDPSTWAEVWVKLFEDKTFIRKD TSKEICESGYDIHLESEYLERIYLNLLHK*</p>			

Table S3.5 BF9343_3589 gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	BF3589		
Locus Tag	BF9343_3589		
DNA sequence	Length	981	bp
<p>ATGAAGAAGGTTTTTATTCCAATCAACCAAGATTCCAGTCAACGTCGAATTCCCAACCGTGTCCAATTTGGGGTA ACTACGAATTCATCATCTCTACCAAGAACCAGAACAAGAATACGATTACGTTGTTGTTTGGACGACATCGAATACTC GTTGGCTTTGATGTGTTGCAAGCAAAACATTTGCTTGTCACTGGTGAACCACCATACGTTAAGTTGTATCCACGTAAGT ACTTGAACCAAGTTCGGTCACTGCTATACCTGTCGAATCTTCTGTTTTGAAGCGTGATAACGCTGTTGTCTTATCCAGCTT TGCCTTGGATGTTGACTACAATTTCTACAACGACAAGCAGAAAGAGAGTTGTTGATCGATTACGACTTCTTGAAGAA TCGTCCAACCTTGACAGCGTAAGAACAAGATCTGTTTGTTCACCTCCAACAAGAAGATCTCCAAGGGTCATATTGAACGT ATCAAGTTGCGCTTGAAGCTGCAAGAAGAAATGCCAGATTTGATCGATATCTACGGTCTGTTTCCACCAACGTTGATT ACAAGTACGAAGTTATGGTCCAGTACAAATACGCCATCGTTATTGAAAACGCTCTTACCATATTACTGGACCGAAAA ATTGGCTGATCTTCTTGTGAGTTGCTACCAATCTATTTGGTGATCCACATATCGGCGATTTCTCAGTAAAGAAG AGATGGCCGTTATCGACATCGTAACCTTGTGATGAATCTAAGCAGACCATAAAAAGATCATCGATAACAACGCTACGA GAAGCAGTACGAAAACATATGTCATCCAGAGATAAGATCCTGGACAAGTACAATATGTTCTCTTGATCTCTCCACC TTGGATTCTATTCCAGCTAAGTTGGACAAAGAGAGAGTTGTTGTTGTTCCAATGCGTTTGTCTGTTTTCGACCGTATCCG TAACCGTATCTACCGTCTGTTCTACTGA</p>			
Protein sequence	Length	379	aa
<p>MKKVFIPINTKIPVERQFPNRPVIWGNYEFIISTKEPEQEYDYYVVLDDIEYSLRLMCCKQNICLFTGEPYVKLYPRKYLNQFG HYYTCQSSVLKRDNACLSYPALPWMLYYNFYNDKQKEELLIDYDFLKNRPTLQRKNKICLFTSNKKISKGHIERIKFALKLQEE MPDLIDIYSGFTNVDYKYEV MVQYKAI VIENC SYPPY WTEKLADTFLSGCYPIYFGDPHIGDFFSKEEMAVIDIRNFDESKQ TIKKIIDNNVYEKQYENICHARDKILDKYNMFLSISSTLDSIPAKLDEKLLSPMRLSVFDRIRNRIRLYF*</p>			

Table S3.6 BF9343_0585 gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	BF0585		
Locus Tag	BF9343_0585		
DNA sequence	Length	1563	bp
<p>ATGAATGTTAACAAACCTACTACTGAGAAGAAGCTGATCGATTTGAACAACGATATCATCCACAACCTTCGATGTCTCCA TCGTTATGTCTTTCTACAAGCGTTACACCGAGTTCCGTAAGTTTTGCCACATAATGCTCCATACTTGAACGTAACGGT ATCGAAGTTATCATCGTTTTGGATGATCCAGACGAAAAGTCTGAGTTGTTGATGTTGTTGCAAACTACCCATTATCA</p>			

<p>ACTGGAAGTTGATCATCAACGAACGTAACATGCCACGTAACCATGCTTCTGTTTGAATGTTGGTTGAAGCACGC TACCAAGAAGTACATCTTGCAAATTGATCCAGAAGTCGAATTCTTGACCGACATCATATGGCAAATGCGTGATGCCATT GAAAAGTACCCAATGCATTACATTTGGCCATGATGGCTTATGTCCATACGAACAAGAATTGACCGAGAACAATATCA AAGAGTTGGATTTCATTCTTGGGGCAACTTGATGGTTGAACGTAACCACTTGACAAAATTGACCGGTTACGACGAAAC TTTCATTACTTGGGGTGGTGAAGATAACAACATGCGTGCTGTTGGATATGTCCGGTATCAAGAAGTTCATTTTGCCA GAAGCTAAGACCATCCACCGTAAAAGAATTACGATCAAACGAGCGTTCCAAGCGTATCAACAAGCACTCTATTTCCG ATTGGCGTAAGATGAATTATCCATCCGAAGCTATTGCCAACAAGGATATTTGGGGTTCTGAATTCACAAGGTTATCTA CGATTGGCAAGACAATCAATACGCTAAGGATTTGTGTTACACCTACTTGAACAATTCATCGGTTTCGAAAATTAGACAT CCAGTGCTTTTCGTAAGCGTCAAAAAAATCGTTTTGTGCCAAGCCTACAACGAGGAAAAATTGATTGAAGGTTTCC TGACCAACATGGCCAATTACTTTGATGGTATCATCTTGTGGACGACGAATCTACTGATAGAATTGGGACTTAGCCAT CCATGATAAGATTATCTTGAAGGTCAAGAAGAAACGTTCCGGTTTTAACGATTTGGAGAACCGTAACATTCTGTTGGAT TTGTCTGCTTTCTCCAGTCTGAATGGTCTGCTTTATGGATATCGATGAGCGTTTCGACGAACGTTTCACTAATTCTCT GAGTTCGAGAACAACAAGAAATCCACGTTGTTTCATTCCGTGGTGTACTTCTGGGAACGATGAACAATCTTACAAGG GTGATATCCCAATTCCAACAAGGTATTTGACCGTTTACCCTATGTTCCGTCTATTGGTCATACCCATATTAACACCC ATAAGAAGTTGCACCTTCTGCTACCCCTACTTCAATAACTTGGCAGTCAACATCTTGTTCAGGACTATGGTTCC ATGAAGGAAAACGACCGTATCCGTAAGTACGAGCGTTACATCCAAGAAGATCAGCAAAAGGATATGTCATCCGGTTAC GATTACTGTTGAACCTCGAAAACCTGTACCAGTTGGACAAGATCGAAGAGTACTGA</p>			
Protein sequence	Length	521	aa
<p>MNVNKPTEKKLIDLNNDIIHNFVDSIVMSFYKRYTEFRKVLPHNAPYLQRNGIEVIIVLDDPDEKSELLMLLQNYPFINWKLII NERKHAPRNHASVLNVGLKHATKKYILQIDPEVEFLTDIIWQMRDAIEKYPMHYILAMMAYVPYEQELTENNICKELDFIPWG NLMVERNHLKLYKLGHYDEFITWGGEDNMRARLDMSGIKKFIPEAKTIHREKNYDPNERSKRINKHSISDWRKMNYPSEA IANKDIWGSEFNKVIYDWQDNQYAKDLCTYLQQFIGFEIRHPAAFRKRHKKIVLCOAYNEEKLIEGFLTNMANYFDGIILLD DESTDRTWDLAIHDKILKVKKKRSGFNLENRNILLDSAFFQSEWFCFMDIDERFDERFTNFSEFENNKEIHVVSFRGVYL WNDEQSYKGDIPNSNKGILTVYRMFRPIGHTHINTHKKLHFIATPYFTNTWQSNILFDYGSMEKENDRIRKERYIQEDQQK DMSSGYDYLLNSENLYQLDKIEEY*</p>			

Table S3.7 SUMO gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	BF0585		
Locus Tag	BF9343_0585		
DNA sequence	Length	294	bp
<p>ATGTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTA AAGGTGTCCGATGGATCTTCAGAGATCTTCTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTC GCTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCTGTACGACGGTATTAGAATCAAGCTGATCAGACCCCT GAAGATTGGACATGGAGGATAACGATATTATTGAGGCTCACAGAGAACAGATTGGTGGT</p>			
Protein sequence	Length	98	aa
<p>MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFKRQKEMDSLRFYDGIHQADQTPEDLDM EDNDIIEAHREQIGG</p>			





Chapter 4

Engineering a *S. cerevisiae* strain to produce α -galactosylceramide







4.1 Expression of BF9343_3149 in *S. cerevisiae*

4.1.1 Expression vectors for BF3149 in *S. cerevisiae* for different protein localization

The main objective of this part of the project is the insertion of the *Bacteroides fragilis* BF9343_3149 α -GalCer GT gene into a yeast strain to produce the α -galactosylceramide. However, the BF3149 expression plasmid was transformed in a non-modified strain to study phenotype and Bf3149 enzyme viability. If successful results were obtained, this construct would then be used to transform the ceramide-optimized strain.

In the Introduction chapter, it was explained that ceramide is synthesized in the endoplasmic reticulum (ER) and then transported to the Golgi apparatus (GA) where it is glycosylated. This knowledge needs to be taken into consideration when expressing a heterologous protein. If the heterologous enzyme is expressed and released into the cytoplasm, it may not find available substrates to catalyze its reaction. For that reason, three plasmids were designed to express the protein in different localizations.

The first plasmid was designed to express a cytoplasmic protein. Since UDP-galactose is present in the cytosol, Bf3149 may be able to be near a membrane with available ceramide, using the soluble UDP-galactose to perform the α -galactosylceramide synthase reaction.

The second plasmid was designed to redirect Bf3149 into the ER. Ceramide is synthesized in this organelle and then transported into the GA. The hypothesis of redirection of Bf3149 into the ER is that ceramide is still not glycosylated by endogenous glycosyltransferases of the GA and probably ceramide pool is still highly available. This strategy consists of fusing the signal peptide of α -mating factor of *S. cerevisiae* to the N-terminus of Bf3149 sequence, which will allow the translation of the protein into the ER^{130–132}. After that, Bf3149 would follow the secretion pathway. To keep Bf3149 in the ER, HDEL (Histidine-Aspartate-Glutamate-Leucine) sequence is also fused to the C-terminus of the protein, this target peptide prevents the protein to follow the secretion pathway by retrieving it to the ER, thus, maintaining the protein in this organelle^{130,133,134}.

Finally, the third plasmid was designed to redirect Bf3149 into GA, where is taking place the glycosylation of ceramide and proteins. It has been reported that UDP-galactose transporters are present in the *S. cerevisiae* GA¹³⁵, thus indicating UDP-Gal is available in this organelle. Furthermore, ceramide is transported into GA after being synthesized in the ER, that means that it will also be available for Bf3149. In GA, it also occurs the synthesis of inositol phosphorylceramide by the AUR1 inositol phosphoryl ceramide synthase. The strategy followed to redirect Bf3149 to this organelle was previously described^{136,137} and uses *KRE2*. Kre2 protein is a GA native protein that is anchored to the GA membrane by a transmembrane domain in its N-terminus sequence. The design was to fuse the first 100 residues of Kre2 protein to the N-terminus of Bf3149, forming a fusion protein Kre2-Bf3149. For that, *KRE2* gene sequence would be obtained and cloned into a vector with BF319 gene.



Nonetheless, Bf3149 actual location in yeast had to be checked after transformation with the new plasmids. Protein location was carried out using confocal fluorescence microscopy, which allows to observe fluorescent sources in the cell. As a control, the three same plasmids were constructed swapping Bf3149 sequence by mKate2 sequence. This would allow to check construct viability and use it as a control for localization analysis. Other fluorescent dyes would also be used to stain specific cellular organelles and, thus, see if protein and dyes would co-locate.

4.1.1.1 *Design of pSF-BF3149 vectors and fusion with mKate2 gene*

For yeast expression, BF3149 is desired to reach high levels of expression, therefore, a 2 μ plasmid is used. However, it was decided to use a constitutive promoter with mild strength, since the protein has been shown to be insoluble. The chosen plasmid was the commercial vector pSF-TPI1-URA3. This vector has a 2 μ origin, TPI1 promoter, which is a constitutive promoter of TPI1 with mild-to-strong expression levels¹³⁸, and CycE1 terminator. The cloning of BF3149 sequence into psF-TPI1 vector will be carried out using CPEC assembly (Figure 4.1), following the designed approach with the primers described in Table 4.1.

Three sets of vectors (each set containing three plasmids for cytosol, GA and ER relocation) will be designed and constructed: pSF-BF3149, with the enzyme gene (Figure 4.2), pSF-BF3149-mKate2, with a fusion construct of BF3149 and mKate2 (Figure 4.3), and pSF-mKate2, with mKate2 gene but no BF3149 (Figure S4.1 in this chapter's Annexes).

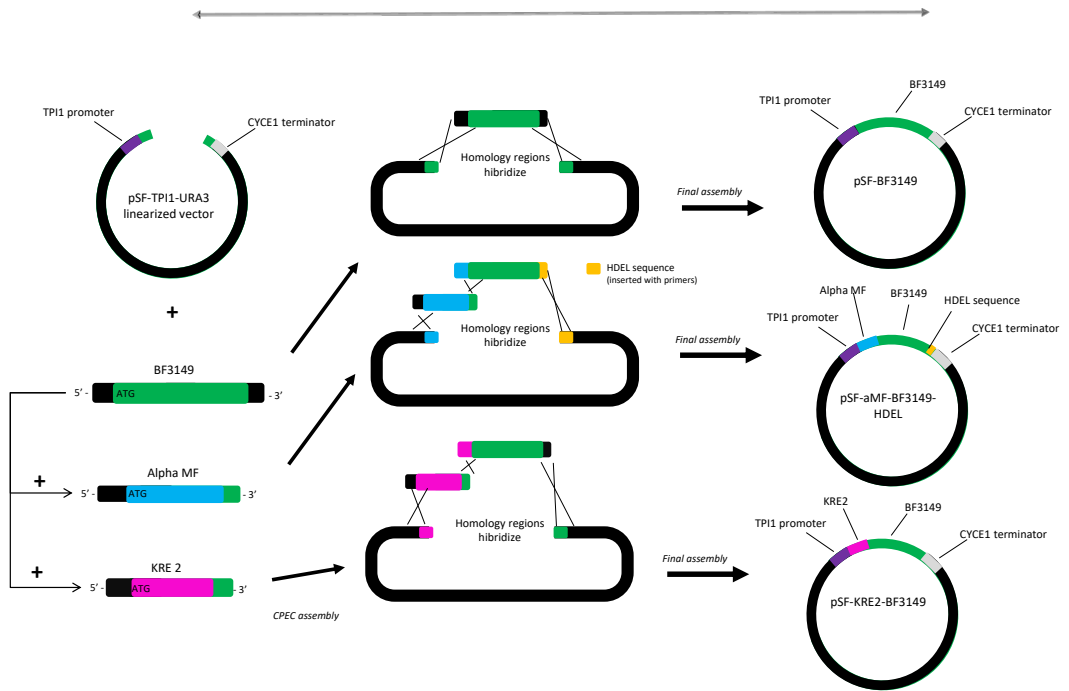


Figure 4.1 CPEC strategy followed for construction of pSF-BF3149 vectors. BF3149 gene was amplified using primers with 20 bp homology regions to adjacent regions (depending on the final plasmid, regions may be: Alpha MF or KRE2 at 5', or HDEL sequence or CYCE1 terminator at 3'). HDEL sequence is cloned by adding the 12 bp DNA sequence as overhanging regions at 5' of primer Fwd for pSF-TPI1 linearization, and 5' of primer Rev of BF3149. For pSF-BF3149, homology regions for BF3149 amplification are at TPI1 promoter and CYCE1 terminator. Although the assembly figure shows black regions, these indicate the backbone (including TPI1 and CYCE1).



Table 4.1 List of primer design used for pSF-TPI1-URA3 plasmid construction. Localization indicates the redirection approach of the plasmid. Fragment indicates the DNA fragment of the plasmid. Primer indicates the direction. 5' homologous region and 3' binding region indicate the sequences they will contain, in parenthesis the specific sequence they are targeted.

Localization	Fragment	Primer	5' homologous region	3' binding region
Cytosol	Vector backbone	Forward	BF3149 (end)	Vector (Stop codon + terminator CycE1)
		Reverse	BF3149 (start)	Vector (Promoter TPI1)
	BF3149 insert	Forward	Vector (Promoter TPI1)	BF3149 (start)
		Reverse	Vector (Stop codon + terminator CycE1)	BF3149 (end)
Golgi Apparatus	Vector backbone	Forward	BF3149 (end)	Vector (Stop codon + terminator CycE1)
		Reverse	KRE2 (start)	Vector (Promoter TPI1)
	KRE2 insert	Forward	Vector (Promoter TPI1)	KRE2 (start)
		Reverse	BF3149	KRE2 (end)
	BF3149 insert	Forward	KRE2 (end)	Bf3149 (start)
		Reverse	Vector (Stop codon + terminator CycE1)	BF3149 (end)
Endoplasmic reticulum	Vector backbone	Forward	BF3149 (end) + HDEL	Vector (Stop codon + terminator CycE1)
		Reverse	α -MF (start)	Vector (Promoter TPI1)
	α -MF insert	Forward	Vector (Promoter TPI1)	α -MF (start)
		Reverse	BF3149	α -MF (end)
	BF3149 insert	Forward	α -MF (end)	BF3149 (start)
		Reverse	HDEL + Vector (Stop codon + terminator CycE1)	BF3149 (end)

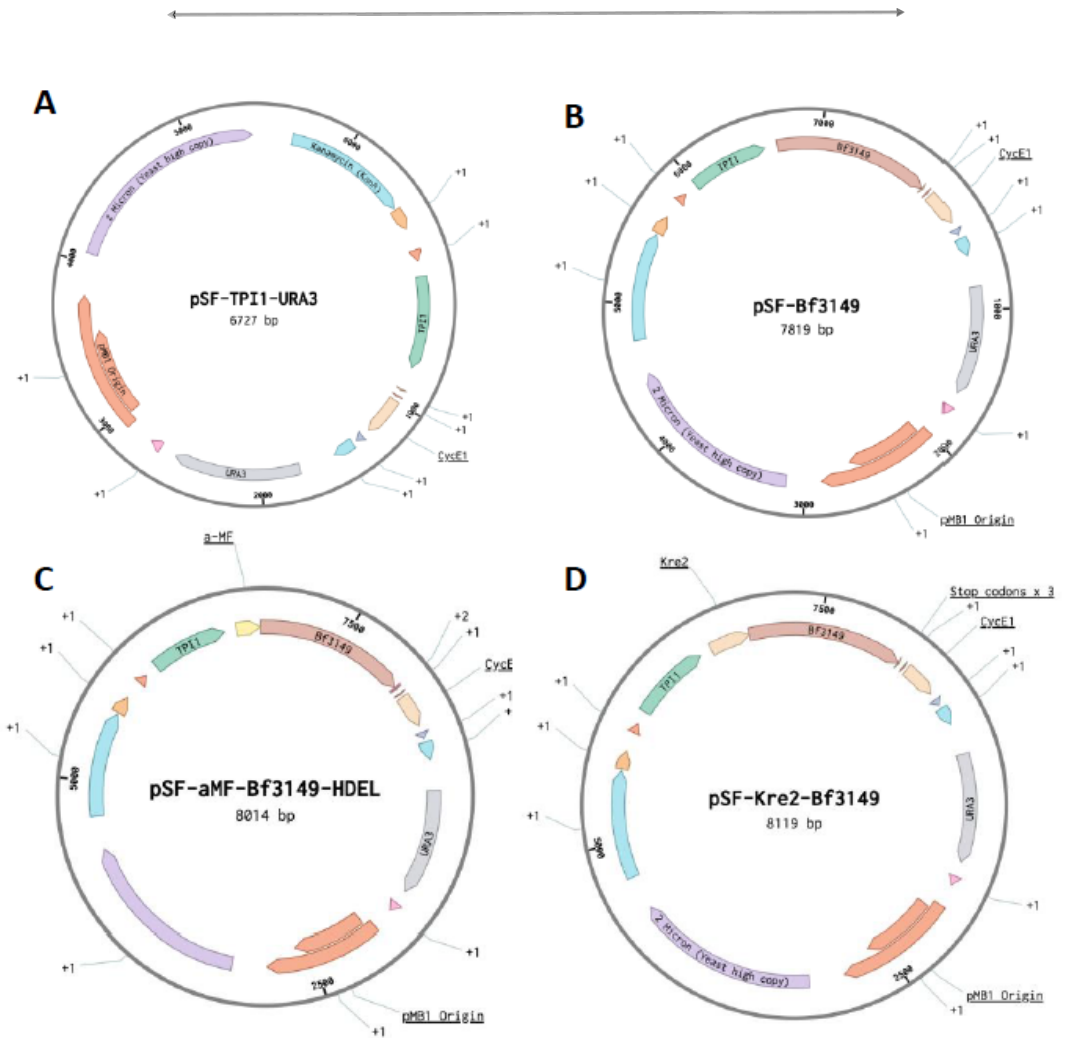


Figure 4.2 Vector map of pSF-TPI1-URA3 (A), pSF-Bf3149 (B), pSF- α MF-Bf3149-HDEL (C) and pSF-Kre2-Bf3149 (D). In the center, the name of the vector and its size. Legend: Light blue, Kanamycin R. Orange, RnG Bacterial Terminator. Red (short), 5' HS1 BetaGlobulin Insulator. Green, *TPI1* promoter. Brown (short), Stop codon x3. Light brown, *CycE1* terminator. Blue (short), T7 terminator. Light blue, RnG Bacterial terminator. Grey, URA3 selection marker. Pink, 3' HS1 BetaGlobulin Insulator. Red, pMB1 origin. Purple, 2 μ origin. Brown (B, C, D), Bf3149 gene (gene from *B. fragilis* BF9343_3149 α -GalCer synthase/GT). Yellow (C), α Mating Factor signal peptide. Light brown (D), *KRE2* sequence.

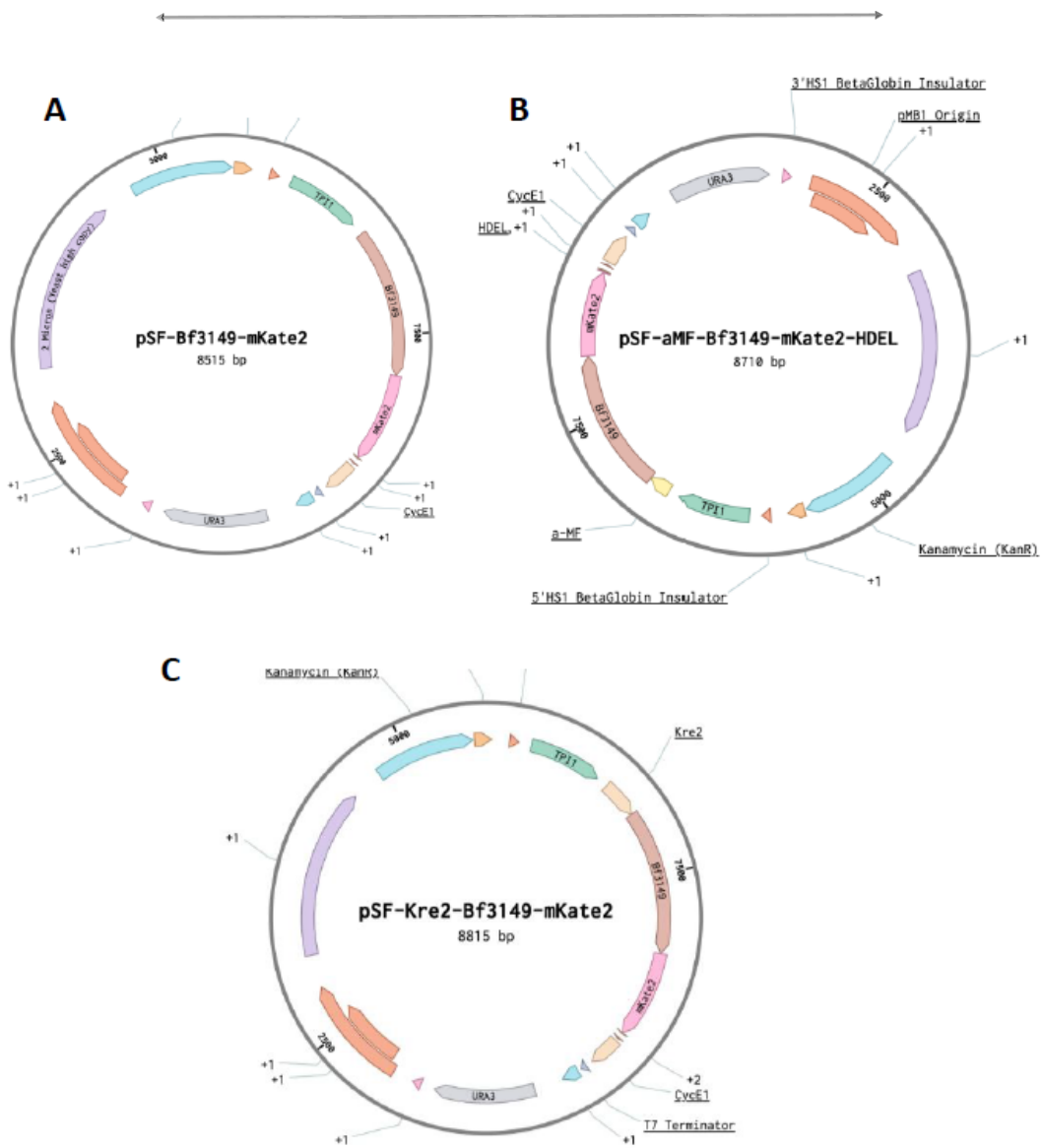


Figure 4.3 Vector map of **pSF-BF3149-mKate2 (A)**, **pSF- α MF-BF3149-mKate2-HDEL (B)** and **pSF-KRE2-BF3149-mKate2 (C)**. In the center, the name of the vector and its size. Legend: Light blue, Kanamycin R. Orange, RnG Bacterial Terminator. Red (short), 5' HS1 BetaGlobulin Insulator. Green, *TPI1* promoter. Brown (short), Stop codon x3. Light brown, CycE1 terminator. Blue (short), T7 terminator. Light blue, RnG Bacterial terminator. Grey, URA3 selection marker. Pink, 3' HS1 BetaGlobulin Insulator. Red, pMB1 origin. Purple, 2 μ origin. Brown (A, B, C), Bf3149 gene (gene from *B. fragilis* BF9343_3149 α -GalCer synthase/GT). Yellow (B), α Mating Factor signal peptide. Light brown (C), *KRE2* sequence. Pink (A, B, C), mKate2 gene.

4.1.1.2 Construction

pSF-BF3149, pSF- α MF-BF3149-HDEL and pSF-KRE2-BF3149 vectors:

Vector construction was carried out using CPEC assembly. Backbone was obtained from pSF-TPI1-URA3. BF3149 sequence was obtained from pET28a-BF3149. α Mating Factor signal peptide (α MF) was obtained by assembling two 100 bp primers by PCR (α MF_long_fw and α MF-long_rv, listed in Table S4.1 from Annexes of this chapter, when hybridizing both primers, the complete sequence was obtained), which after elongation formed the 183 bp of α MF signal peptide sequence. HDEL DNA sequence was added in the homologous regions of the primers, and KRE2 sequence (300 bp from the start codon) was obtained from yeast genomic DNA. pSF-TPI1-URA3 vector backbone and inserts had a 20 bp homologous sequence for assembly.

PCR amplifications were carried out with the primers listed in Table 4.2, with vector pSF-TPI1-URA3, pET28a-BF3148, genomic DNA and α MF DNA sequence to be used as templates.

Table 4.2 List of primers and templates used to amplify each fragment for pSF-BF3149 plasmids. Table indicates the name of the final vector to be constructed, the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Vector	Fragment	Template	Primers	Length of fragment (bp)
pSF-BF3149	Vector backbone	pSF-TPI1-URA3	cyt_bb_fw cyt_bb_rv	6710
	BF3149 insert	pET28a-BF3149	cyt_BfGT_fw cyt_BfGT_rv	1189
pSF-αMF-BF3149-HDEL	Vector backbone	pSF-TPI1-URA3	ER_bb_fw ER_bb_rv	6710
	α MF insert	α MF fragment	ER_aMF_fw ER_aMF_rv	223
	BF3149 insert	pET28a-BF3149	ER_BfGT_fw ER_BfGT_rv	1189
pSF-KRE2-BF3149	Vector backbone	pSF-TPI1-URA3	cyt_bb_fw GA_bb_rv	6710
	KRE2 insert	Genomic DNA	GA_Kre2_fw GA_Kre2_rv	340
	BF3149 insert	pET28a-BF3149	GA_BfGT_fw cyt_BfGT_rv	1189

Amplification results were checked by agarose gel electrophoresis. Figure 4.4 shows that BF3149 gene for pSF-BF3149 and pSF- α MF-BF3149-HDEL was amplified with the expected size of 1.2 kb. Vector backbone for the three vectors was amplified (6,7 kb) but showed a second shorter band around 1.5 kb, therefore, 6.7 kb band was removed using DNA Gel Extraction kit.

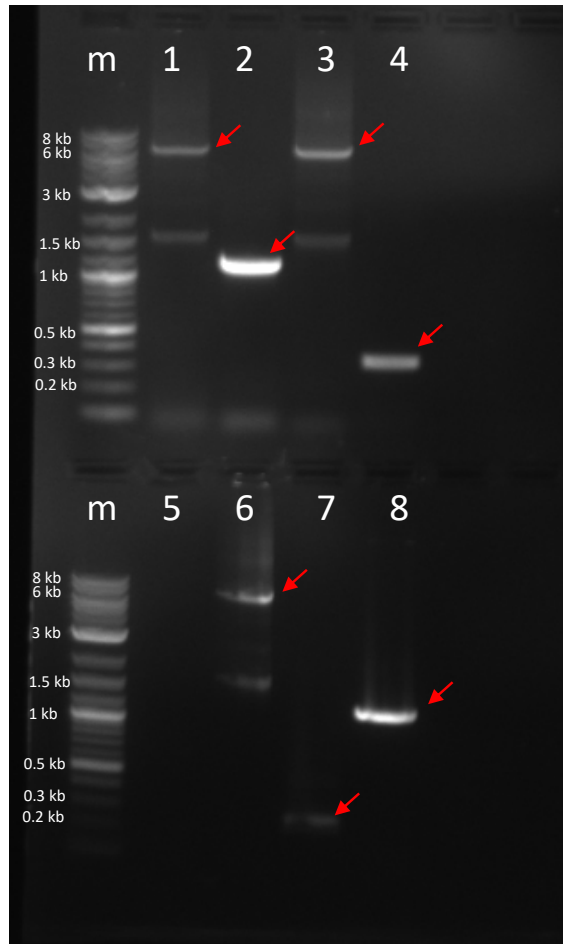


Figure 4.4 Agarose gel electrophoresis results of PCR amplification of pSF-BF3149 plasmids. *m* lane: DNA marker (1 kb plus DNA ladder from NEB), Lane 1: linearized pSF vector for pSF-BF3149. Lane 2: BF3149 gene for pSF-BF3149. Lane 3: linearized pSF vector for pSF-KRE2-BF3149. Lane 4: KRE2 fragment. Lane 5: BF3149 gene for pSF-KRE2-BF3149. Lane 6: linearized pSF vector for pSF- α MF-BF3149-HDEL. Lane 7: α -mating factor fragment. Lane 8: BF3149 gene for pSF- α MF-BF3149-HDEL.

Once all fragments were obtained, they were purified and quantified. Fragments were then used in CPEC assembly. For each vector reaction, an equimolar amount of each DNA fragment was used: backbone and BF3149 for pSF-BF3149, backbone, α MF and BF3149 sequences for pSF- α MF-BF3149-HDEL, and backbone, KRE2 and BF3149 sequences for pSF-KRE2-BF3149 (see Table 4.2). *E. coli* DH5 α cells were transformed with reaction products. All transformation yielded a high number of colonies.

Transformant colonies for each vector were analyzed by restriction digestion using BamHI and SacI restriction enzymes (Table 4.3 and Figure 4.5).

Table 4.3 Table of restriction enzymes and expected fragments for restriction analysis for plasmids pSF-TPI1-URA3 with only Bf3149 sequence.

Plasmid	Restriction enzymes	Length of fragments (bp)
pSF-TPI1-URA3	BamHI + SacI	6.6 + 0.1 kb
pSF-BF3149	BamHI + SacI	6.6 + 1.2 kb
pSF- α MF-BF3149-HDEL	BamHI + SacI	6.6 + 1.4 kb
pSF-KRE2-BF3149	BamHI + SacI	6.6 + 1.5 kb

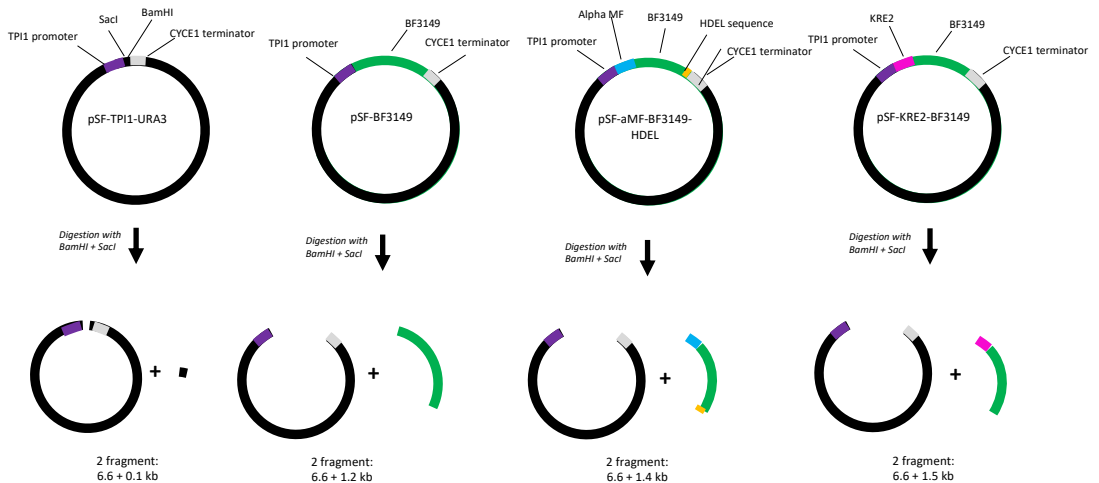


Figure 4.5 Digestion analysis of pSF-BF3149 vectors. Plasmids were extracted from different colonies transformed with pSF-BF3149, pSF- α MF-BF3149-HDEL and pSF-KRE2-BF3149) and digested with BamHI and SacI. pSF-TPI1-URA3 was also digested with the same restriction enzymes. All vectors were expected to show a large fragment of 6.6 kb, and a smaller fragment (not visible for pSF-TPI1-URA3 vector since it was 0.1 kb) of 1.2, 1.4 and 1.5 kb for pSF-BF3149, pSF- α MF-BF3149 and pSF-KRE2-BF3149, respectively.

Results (Figure 4.6.) matched the expected sizes of 6.6+0.1 kb for pSF-TPI1-URA3 (the 0.1 kb fragment could not be observed), 6.6+1.2 kb for pSF-BF3149 (only one colony matched the expected sizes), 6.6+1.4 kb for pSF- α MF-BF3149-HDEL (all colonies with the expected sizes), and 6.6+1.5kb for pSF-KRE2-BF3149 (all colonies with the expected sizes).

The plasmids from colony 4 with pSF-BF3149 plasmid, colony 1 with pSF-KRE2-BF3149 and colony 1 with pSF- α MF-BF3149-HDEL were sent for Sanger sequencing analysis, which confirmed the expected sequences.

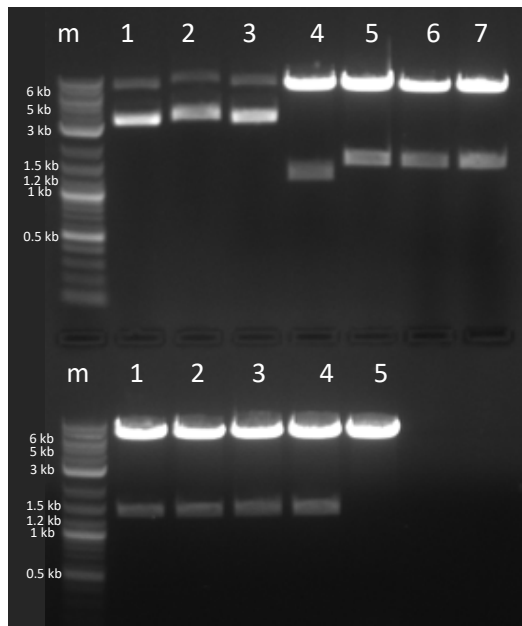


Figure 4.6 Agarose gel electrophoresis of restriction analysis of transformed colonies with pSF-Bf3149 plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Top: lanes 1-4: digestion of pSF-BF3149 plasmid from colonies 1 to 4. Lanes 5-7: digestion of pSF-KRE2-BF3149 plasmid from colonies 1 to 3. Bottom: Lane 1-4: digestion of pSF- α MF-BF3149-HDEL plasmid from colonies 1 to 4. Lane 5: digestion of pSF-TPI1-URA.

pSF-Bf3149-mKate2, pSF- α MF-Bf3149-mKate2-HDEL and pSF-KRE2-Bf3149-mKate2 vectors

Once pSF-Bf3149 vectors were synthesized, they were used as backbones to insert mKate2 sequence and obtain the plasmid with both genes. This construction would result as the fusion protein of α -GalCer synthase and mKate2 at N-terminus and C-terminus, respectively. All three vectors were linearized with PCR and mKate2 sequence was amplified using primers listed on Table 4.4 (Figure 4.7). Cloning was carried out by CPEC assembly (Figure 4.7)

pSF-BF3149-mKate2 and pSF-KRE2-BF3149-mKate2 used the same primers (cyt_bb_mK2_fw and cyt_bb_mK2_rv for backbone, and cyt_mK2_fw and cyt_mK2_rv for mKate2) for amplification of mKate2 and vector backbone since mKate2 is inserted at 3' of BF3149 sequence, which is the same for both plasmids. On the other hand, pSF- α MF-BF3149-mKate2-HDEL has the HDEL sequence at 3' of BF3149 sequence and primers homologous to these regions were different (ER_bb_mK2_fw and cyt_bb_mK2_rv for backbone, and cyt_mK2_fw and ER_mK2_rv for mKate2).

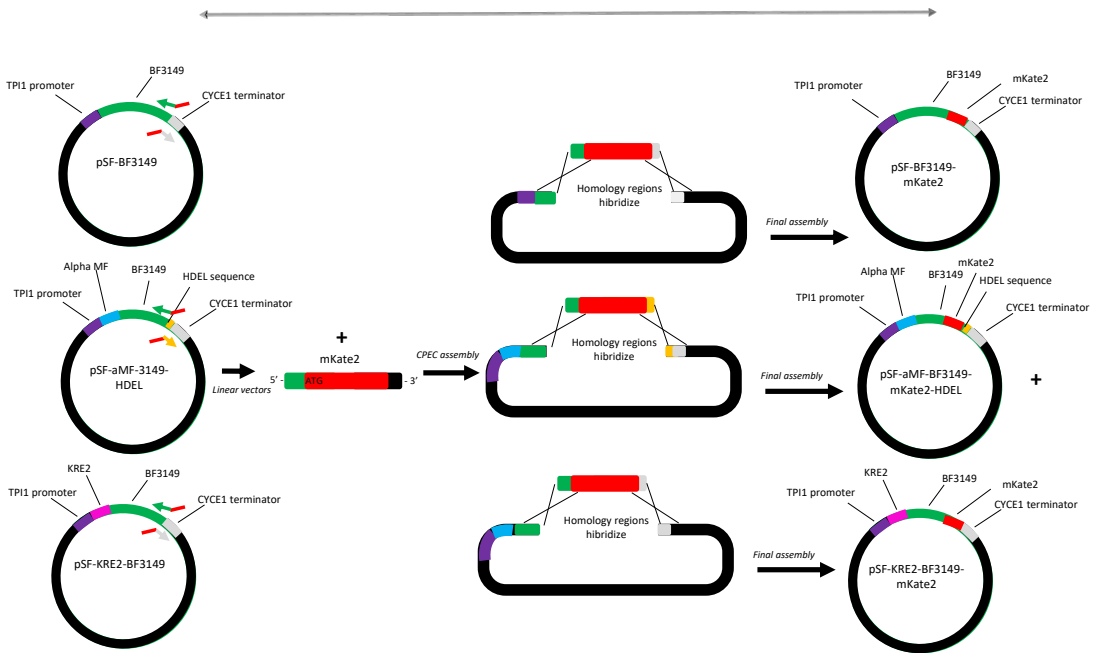


Figure 4.7 CPEC strategy followed for construction of pSF-BF3149-mKate2 vectors. The three pSF-BF3149 vectors were linearized with primers containing homology sequences to mKate2. mKate2 was also amplified with primers that added 20 bp homology regions to both ends. For pSF- α MF-BF3149-mKate2-HDEL plasmid, mKate2 sequence at 3' contained homology to HDEL sequence instead of CYCE1 terminator. Fragments were assembled by CPEC, hybridizing homology sequences to each adjacent region.

Table 4.4 List of primers and templates used to amplify each fragment for pSF-Bf3149-mKate2 plasmids. Table indicates the name of the final vector to be constructed, the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Fragments of backbone for pSF-BF3149-mKate2 and pSF-Kre2-BF3149-mKate2 are the same, so they are showed in the same row. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Vector	Fragment	Template	Primers	Length of fragment (bp)
pSF-BF3149-mKate2 / pSF-KRE2-BF3149-mKate2	Vector backbone	pSF-BF3149	cyt_bb_mK2_fw cyt_bb_mK2_rv	7859 / 8159
	mKate2 insert	pHES834	cyt_mK2_fw cyt_mK2_rv	736
pSF-αMF-BF3149-mKate2-HDEL	Vector backbone	pSF- α MF-BF3149-HDEL	ER_bb_mK2_fw cyt_bb_mK2_rv	8054
	mKate2 insert	pHES834	cyt_mK2_fw ER_mK2_rv	736



Figure 4.8 shows that vector backbones for pSF-Bf3149-mKate2, pSF- α MF-BF3149-mKate2-HDEL and pSF-KRE2-BF3149-mKate2 showed the expected sizes (7.9, 8 and 8.2 kb, respectively). DNA fragments of mKate2 were also amplified with the expected size (0.7 kb).

Once fragments were purified and quantified, they were used in CPEC assembly. CPEC products were then used to transform DH5 α *E. coli* cells. Transformant colonies for each vector were analyzed by colony PCR (with ColPCR_pSF_fw and ColPCR_pSF_rv primers, Figure 4.9) with the expected fragment sizes shown at Table 4.5. As control, pSF-TPI1-URA3 plasmid was also amplified using the same primers. Figure 4.10 shows that, although not all colonies amplified the expected fragment sizes, some colonies from pSF-BF319-mKate2, pSF- α MF-BF3149-mKate2-HDEL and pSF-KRE2-BF3149-mKate2 showed the expected fragment sizes of 2.2, 2.4 and 2.5 kb, respectively. Positive colonies were analyzed by sanger sequencing, which confirmed that the construction matched with the expected sequence.

Table 4.5 Table of primers used in Colony PCR for pSF-BfGT-mKate2 plasmids. Template name, primers used, and fragment length are shown. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Template	Primers	Length of fragment (bp)
Colony 1-4 of pSF-BF3149-mKate2	ColPCR_pSF_fw	2175
	ColPCR_pSF_rv	
Colony 1-4 of pSF-KRE2-BF3149-mKate2	ColPCR_pSF_fw	2475
	ColPCR_pSF_rv	
Colony 1-4 of pSF-αMF-BF3149-mKate2-HDEL	ColPCR_pSF_fw	2370
	ColPCR_pSF_rv	
pSF-TPI1-URA3 plasmid	ColPCR_pSF_fw	387
	ColPCR_pSF_rv	

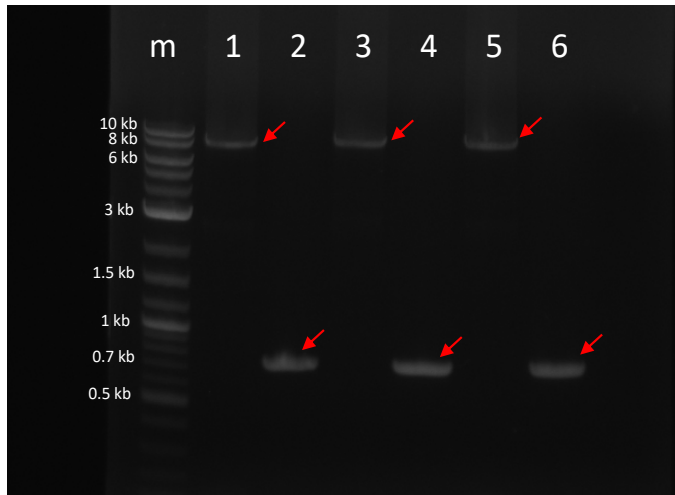


Figure 4.8 Agarose gel electrophoresis results of PCR amplification of pSF-BF3149-mKate2 plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows indicate fragments that matched the expected size. Lane 1: linearized pSF-BF3149. Lane 2: mKate2 gene. Lane 3: linearized pSF-KRE2-BF3149 vector. Lane 4: mKate2 gene. Lane 5: linearized pSF- α MF-BF3149-HDEL vector. Lane 6: mKate2 gene for cloning into pSF- α MF-BF3149-mKate2-HDEL vector.

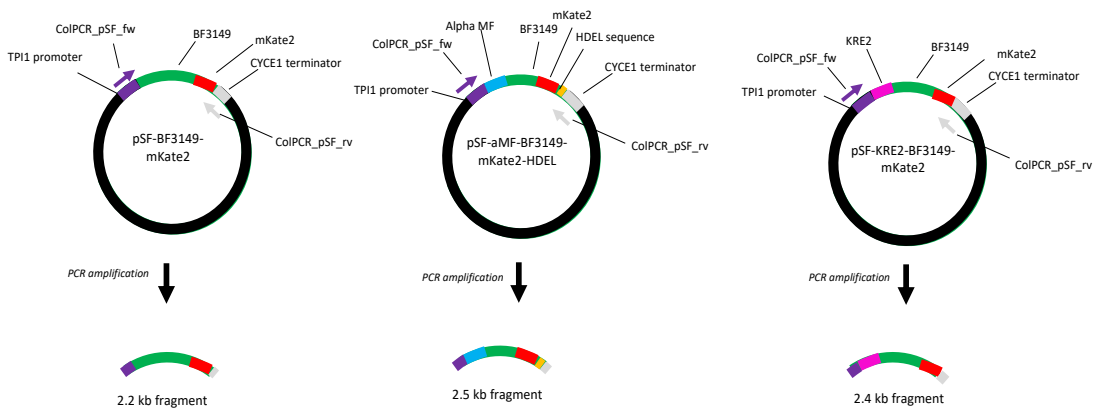


Figure 4.9 Colony PCR analysis of pSF-BF3149-mKate2 vectors. ColPCR_pSF_fw and ColPCR_pSF_rv primers were used for the amplification. These hybridized with TPI1 promoter and CYCE1 terminator, respectively. Fragments of 2.2, 2.5 and 2.4 kb were expected for pSF-BF3149-mKate2, pSF- α MF-BF3149-mKate2-HDEL and pSF-KRE2-BF3149-mKate2, respectively.

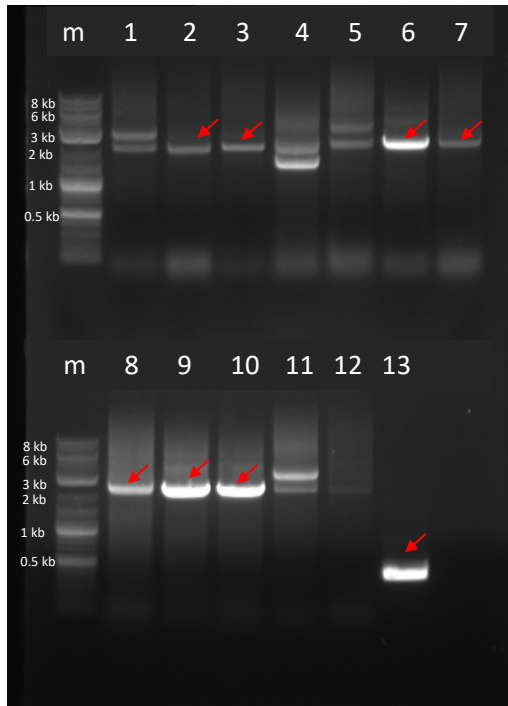


Figure 4.10 Agarose gel electrophoresis results of restriction analysis of transformed colonies with pSF-Bf3149-mKate2 plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows indicate fragments that matched the expected size. Lanes 1-4: Transformed colonies 1 to 4 of pSF-BF3149-mKate2. Lanes 5-8: Transformed colonies 1-4 of pSF-KRE2-BF3149-mKate2. Lanes 9-12: Transformed colonies 1-4 of pSF- α MF-BF3149-mKate2. Lane 13: Plasmid pSF-TPI1-URA3.

pSF-mKate2, pSF- α MF-mKate2-HDEL and pSF-KRE2-mKate2 vectors

Vectors containing mkate2 gene for expression in cytoplasm, GA and ER were also constructed. For these vectors, it was decided to amplify the backbone in two parts of similar length to increase PCR amplification efficiency (primers at 'half' the vector hybridized near the center of 2 μ origin). They contained the backbones from pSF-BF3149, pSF- α MF-BF3149-HDEL and pSF-KRE2-BF3149 vectors, removing BF3149 sequence, and they were obtained by PCR with primers listed on Table 4.6. It must be noted that pSF-mKate2 and pSF-mKate2 share one of the backbone halves (cyt backbone fw half) since it is the same sequence, thus, this fragment was only amplified once.

Table 4.6 List of primers and templates used to amplify each fragment for pSF-mKate2 plasmids. Table indicates the name of the final vector to be constructed, the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Fragment of cyt backbone fw half for pSF-mKate2 is the same as for pSF-Kre2-mKate2and, so it is not listed. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Vector	Fragment	Template	Primers	Length of fragment (bp)
pSF-mKate2	cyt backbone fw half	pSF-BF3149	cyt_bb_mK2_fw pSF_bb_half_rv	3557
	cyt backbone rv half	pSF-BF3149	pSF_bb_half_fw cyt_bb_mK2_rv	3213
	mKate2 insert	pHES834	mK2_cyt_fw cyt_mK2_rv	736
pSF-KRE2-mKate2	GA backbone rv half	pSF-Kre2-BF3149	pSF_bb_half_fw GA_mK2_bb_rv	3513
	mKate2 insert	pHES834	mK2_GA_fw cyt_mK2_rv	736
pSF- α MF-mKate2-HDEL	ER backbone fw half	pSF- α MF-BF3149-HDEL	ER_bb_mK2_fw pSF_bb_half_rv	3569
	ER backbone rv half	pSF- α MF-BF3149-HDEL	pSF_bb_half_fw ER_mK2_bb_rv	3396
	mKate2 insert	pHES834	mK2_ER_fw ER_mK2_rv	736

Amplified PCR fragments of cyt backbone fw and rv halves, mKate2 insert (for pSF-mKate2), GA backbone rv half, mKate2 insert (for pSF-KRE2-mKate2, and ER backbone fw and rv halves were checked by agarose gel electrophoresis. All fragments (Figure 4.11) matched the expected sizes (3.6,

3.2, 0.7 3.5, 0.7, 3.6 and 3.4 respectively). Fragment of mKate2 sequence for pSF- α MF-mKate2-HDEL is not shown. Then they were purified and quantified.



Figure 4.11 Agarose gel electrophoresis from PCR amplification of pSF-mKate2 plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows indicate fragments that matched the expected size. Lane 1: cyt backbone rv half fragment. Lane 2: ER backbone rv half fragment. Lane 3: GA backbone rv half fragment. Lane 4: cyt backbone fw half fragment. Lane 5: ER backbone fw fragment. Lane 6: mKate2 insert for cyt vector.

DNA fragments were assembled by CPEC with an equimolar amount of DNA. Reaction products were used to transform DH5 α *E. coli* cells. Four transformant colonies of each vector were cultured and plasmid was extracted. Then, plasmids were checked by restriction analysis (Table 4.7) with BamHI and NotI restriction enzymes. Figure 4.12 shows that two colonies of pSF-mKate2 matched the expected sizes of 6.6 and 0.8 kb. Further, three colonies matched the 6.6 and 1 kb of pSF- α MF-mKate2-HDEL vector, whereas only one colony of pSF-KRE2-mKate2 plasmid matched the expected sizes of 6.6 and 1.1 kb.

Plasmids from positive colonies were then sent for Sanger sequencing to analyze the sequence. Results confirmed that all constructs matched with the expected sequence.

Table 4.7 Restriction enzymes and expected fragments for plasmids pSF-mKate2. Name of plasmid, restriction enzyme used and the fragments resulting from digestion are shown.

Plasmid	Restriction enzymes	Length of fragments (bp)
pSF-mKate2	BamHI + NotI	6.6 + 0.8 kb
pSF-KRE2-mKate2	BamHI + NotI	6.6 + 1.1 kb
pSF- α MF-mKate2-HDEL	BamHI + NotI	6.6 + 1 kb

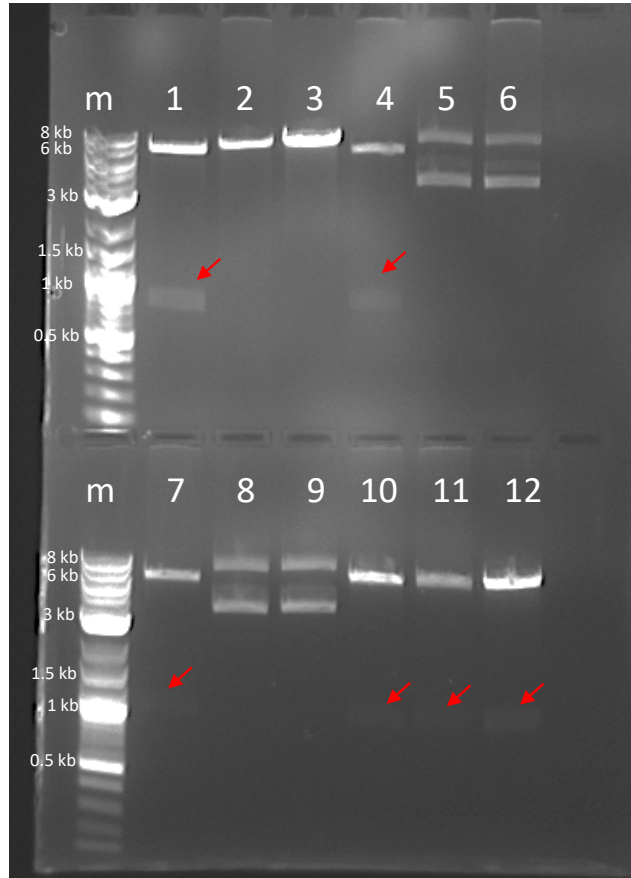


Figure 4.12 Agarose gel electrophoresis results of restriction analysis of transformed colonies with pSF-mKate2 plasmids. *m* lane contains DNA marker (1 kb plus DNA ladder from NEB). Arrows indicate fragments that matched the expected size. Lanes 1-4: Transformed colonies 1 to 4 of pSF-mKate2. Lanes 5-8: Transformed colonies 1-4 of pSF-Kre2-mKate2. Lanes 9-12: Transformed colonies 1-4 of pSF- α MF-mKate2-HDEL.

4.1.1.3 Yeast strains with vectors constructed from pSF-TPI1-URA3 plasmid

A total of three sets of vectors were constructed, making a total of 9 plasmids: (1) pSF-BF3149, (2) pSF- α MF-BF3149-HDEL and (3) pSF-KRE2-BF3149 for Bf3149 α -GalCer synthase expression, (4) pSF-BF3149-mKate2, (5) pSF- α MF-BF3149-mKate2-HDEL and (6) pSF-KRE2-BF3149-mKate2 for fusion protein of BF3149 and mKate2, and (7) pSF-mKate2, (8) pSF- α MF-mKate2-HDEL and (9) pSF-KRE2-mKate2 for expression of mKate2 fluorescent protein. Plasmids containing only BF3149, mKate2 or the fusion of both were used for cytosolic expression. Plasmids with KRE2 fused at N-terminus of BF3149, mKate2 or BF3149-mKate2 were used for Golgi apparatus relocation. Finally, plasmids with α mating factor signal peptide and HDEL sequences at N- and C-terminus of the construct (BF3149, mKate2 or



BF3149-mKate2), respectively, were used for relocation into the endoplasmic reticulum. *S. cerevisiae* RH6082 strain was transformed with the plasmids following the modified Gietz protocol¹⁰⁷.

Although fluorescent protein mKate2 yields reddish colonies due to its red fluorescence (mKate2 excitation/emission wavelengths are 588/633 nm), fusion constructs containing mKate2 may have its fluorescence reduced for the proximity of Bf3149. However, no red colonies were observed at first glance, even in the set of plasmids containing mKate2 (not fused to BF3149). Fluorescence was also measured with a fluorimeter, using cultures with cell lines expressing mKate2, and cell-free extracts of the same cultures after lysis (lyticase treatment followed by vortex with glass beads). However, no fluorescence was detected.

These results indicated that mKate2 was not expressed and it was inferred that the other constructs (with BF3149 and fusion BF3149-mKate2) were not expressed either. No enzymatic activity assays were carried out to check BF3149 expression, since it was assumed that if mKate2 (which is a relatively easier protein to express and synthesize) was not expressed, neither would be the enzyme. For that reason, it was decided to synthesize new expression plasmids using a different vector as backbone.



4.1.2 Expression vectors p2a33-BF3149

Since a vector with a fluorescent protein was already available, p2a33-mCherry, it was decided to use it to clone BF3149 sequence. This low copy number plasmid was very similar to the plasmids used in *SUR2* and *ISC1* expression vectors.

For the new set of expression vectors no plasmids for endoplasmic reticulum were constructed since they were constructed to cytoplasm and GA expression. Redirection in ER would be considered later depending on the results obtained with the previous vectors. Therefore, in this organelle, UDP-galactose would be more available, and ceramide would also be present since it is also converted to inositol phosphorylceramide.

4.1.2.1 Design and construction of p2a33-BF3149 vectors

The plasmid p2a33-mCherry contains the fluorescence protein mCherry sequence under the control of TEF1 promoter and ADH1 terminator. Fluorescence protein mCherry is similar to mKate2 in its excitation and emission wavelengths, so it would not interfere with other dyes that would be used in the localization experiment. The features of this plasmid are: origin of replication is CEN6/ARS4, (low-copy number plasmid), selectable marker URA3 (uracil auxotrophy), TEF1 promoter (strong constitutive expression) and ADH1 terminator (good mRNA stability).

The vectors were designed for BF3149 and BF3149-mCherry protein fusion to expression in cytoplasm and GA. Therefore, four different vectors were designed: p2a33-BF3149, p2a33-BF3149-mCherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry.

p2a33-mCherry was used as backbone, and BF3149 was inserted to replace mCherry sequence, or be inserted at 5' of mCherry gene, with a linker sequence between them. The selected linker was an 8-residue sequence consisting of Thr-Leu-Ala-Gln-Pro-Asp-Ala-Thr that confers some flexibility. This linker has been used previously with mg517 glycosyltransferase in the Laboratory of Biochemistry. The linker sequence was added in both homologous regions of vector and insert, since it was only 24 bp long, therefore, 22 bp sequence at 5' of linker were added at 5' of the vector forward half and the 22 bp sequence at 3' of linker were added at 3' of BF3149 insert sequence. Homology region between the 22 bp sequence was 20 bp long. PCR amplification was carried out using PrimeStar GXL DNA polymerase (TakaraBio) (see Material and methods chapter).

Backbone linearization was done by amplification in two halves of the vector and BF3149 was also amplified with homologous regions that contained the linker sequence, following the primer design in Table 4.8. Figure 4.13 shows the CPEC assembly strategy followed for the construction and Figure 4.14 shows the final designed plasmid maps.



Table 4.8 List of primer design used for p2a33-BF3149 plasmid construction. Localization indicates the organelles where the protein is targeted. Fragment indicates the DNA fragment of the plasmid. Primer indicates the direction. 5' homologous region and 3' binding region indicate the sequences they contain in parenthesis the specific sequence they are targeted.

Localization	Fragment	Primer	5' homologous region	3' binding region
Cytosol	cyt backbone fw half	Forward	BF3149 (end)	Vector (Stop codon + terminator ADH1)
		Reverse	-	Half vector (3' of URA3 terminator)
	cyt backbone rv half	Forward	-	Half vector (3' of URA3 terminator)
		Reverse	BF3149 (start)	Vector (Promoter TEF1)
	BF3149 insert	Forward	Vector (Promoter TEF1)	BF3149 (start)
		Reverse	Vector (Stop codon + terminator ADH1)	BF3149 (end)
Cytosol (with mCherry)	cyt backbone fw half	Forward	Linker	Vector (Stop codon + terminator ADH1)
		Reverse	-	Half vector (3' of URA3 terminator)
	cyt backbone rv half	Forward	-	Half vector (3' of URA3 terminator)
		Reverse	BF3149 (start)	Vector (Promoter TEF1)
	BF3149 insert	Forward	Vector (Promoter TEF1)	BF3149 (start)
		Reverse	Linker	BF3149 (end)
Golgi Apparatus	GA backbone fw half	Forward	Bf3149 (end)	Vector (Stop codon + terminator ADH1)
		Reverse	-	Half vector
	GA backbone rv half	Forward	-	Half vector
		Reverse	KRE2 (start)	Vector (Promoter TEF1)
	BF3149 insert	Forward	Vector (Promoter TEF1)	KRE2 (start)
		Reverse	Vector (Stop codon + terminator ADH1)	BF3149 (end)
Golgi Apparatus (with mCherry)	GA backbone fw half	Forward	Linker	Vector (Stop codon + terminator ADH1)
		Reverse	-	Half vector
	GA backbone rv half	Forward	-	Half vector
		Reverse	KRE2 (start)	Vector (Promoter TEF1)
	BF3149 insert	Forward	Vector (Promoter TEF1)	KRE2 (start)
		Reverse	Linker	BF3149 (end)

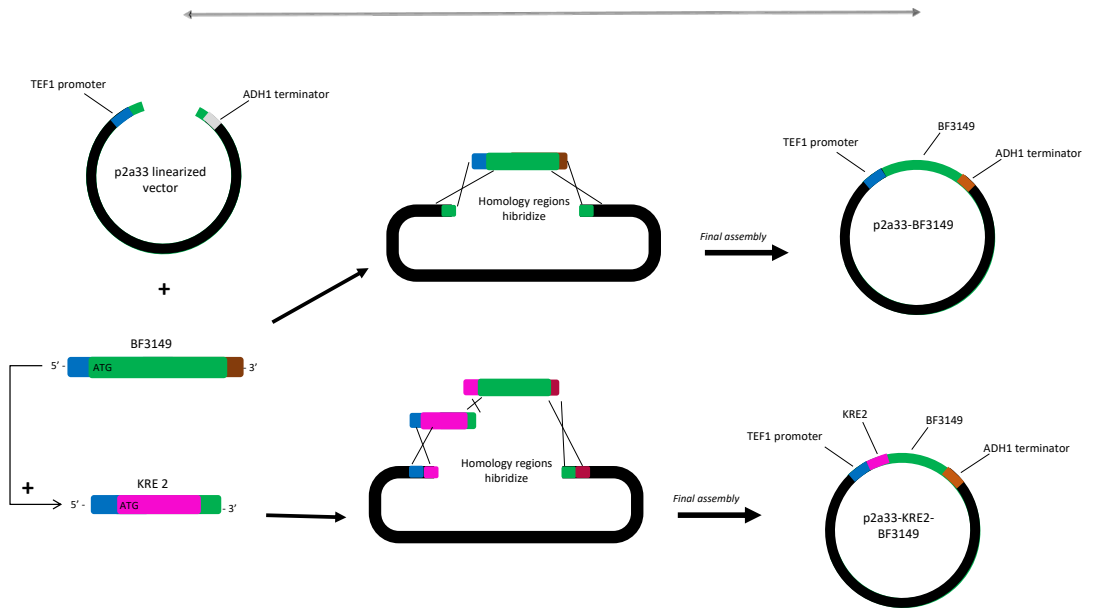


Figure 4.13 CPEC strategy followed for construction of p2a33-BF3149 vectors. Vector p2a33 was amplified from p2a33-mCherry, removing mCherry sequence, and using primers with 20 bp homology sequences to adjacent regions. BF3149 gene was amplified using primers with 20 bp homology sequences to adjacent regions (depending on the final plasmid, regions may be: TEF1 promoter or KRE2 at 5'. Fragments were assembled by hybridization of homologous sequences of adjacent regions.

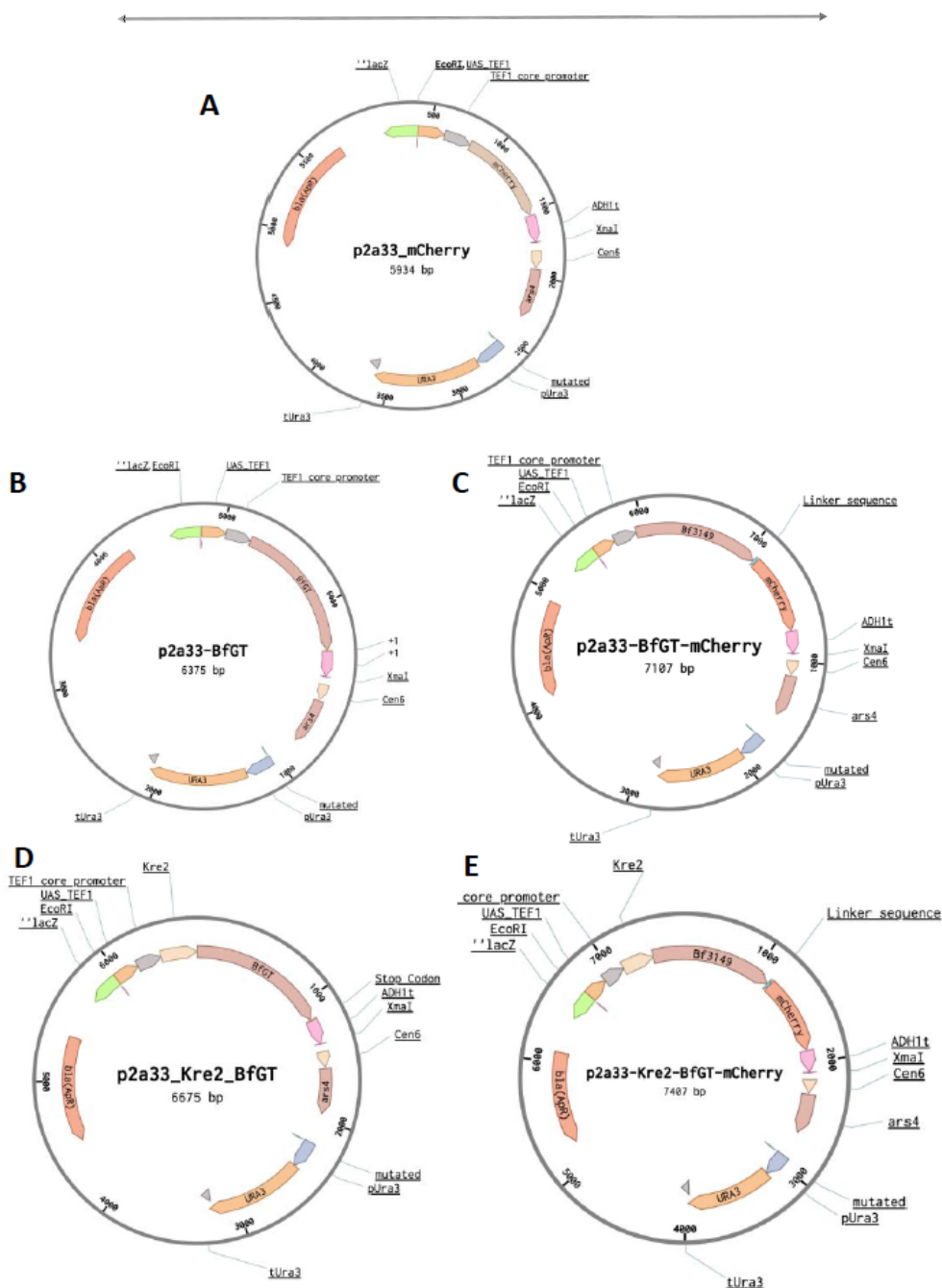


Figure 4.14 Vector map of p2a33-mCherry (A), p2a33-Bf3149 (B), p2a33-Bf3149-mCherry (C), p2a33-Kre2-Bf3149 (4) and p2a33-Bf3149--mCherry (E). Vector elements are marked in colored arrows inside the circle of the plasmid that indicate sequence direction. In the center, the name of the vector and its size. Legend: Green, lacZ. Orange, UAS (Upstream Activating Sequence) *TEF1*. Grey, *TEF1* core promoter. Red (A), mCherry gene. Pink, *ADH1* terminator. Light brown/Brown, CEN6/ARS4 sequences. Blue, *URA3* promoter. Orange, *URA3* selection marker. Red, Ampicillin R. Brown after *TEF1* promoter (B, D) or *KRE2* (C, E), Bf3149 gene. Light blue (C, E), linker sequence. Red (C, D), mCherry sequence. Light brown after *TEF1* promoter (C, E), *KRE2* sequence.



Since the same vector was used for all four plasmids, some linearized fragments of the backbone were the same for some of the plasmids: Fragment of p2a33 backbone forward half was the same for p2a33-BF3149 and p2a33-KRE2-BF3149. Fragment p2a33-mCherry backbone fw half was the same for p2a33-BF3149-mCherry and p2a33-KRE2-BF3149-mCherry. Fragment p2a33-KRE2 backbone rv half for p2a33-KRE2-BF3149 is the same for p2a33-KRE2-BF3149-mCherry.

PCR amplification using primers listed in Table 4.9 yielded the expected fragments (Figure 4.15) for p2a33 backbone fw half (2.4 kb), p2a33 backbone rv half (2.9 kb), BF3149 sequence for p2a33-BF349 (1.1 kb), BF3149 sequence for p2a33-BFB349-mCherry (1.1 kb), p2a33-mCherry backbone fw half (3.1 kb), and p2a33-mCherry backbone rv half (2.8 kb).

The remaining fragments for p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry (p2a33 backbone rv half, KRE2-BF3149 insert for p2a33-KRE2-BF3149, and KRE2-BF3149 insert for p2a33-KRE2-BF3149-mCherry) were amplified and showed the expected sizes (2.9, 1.5 and 1.5 kb, respectively), however, results are not shown.

DNA fragments were purified and quantified after successful amplification. Fragments were assembled with CPEC assembly, which was performed with an equimolar amount of each fragment for each of the four plasmids (following the relation between fragments from Table 4.9 and as indicated previously for non-listed fragments).

E. coli DH5 α cells were transformed with reaction product of each of the plasmids (p2a33-BF3149, p2a33-BF349-mCherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry). Transformant colonies of each plasmid were analyzed by Colony PCR with primers p2a33_ColPCR_fw and tADH_rv (see sequences in Table S4.1 from this chapter's Annexes). Positive colonies matching the expected amplification size (1.1, 1.9, 1.4 and 2.2 kb for (p2a33-BF3149, p2a33-BF349-mCherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry, respectively) were found (results not shown). One positive colony of each plasmid was sent for sequencing analysis, which confirmed that the inserted constructs matched the expected sequences.



Table 4.9 List of primers and templates used to amplify each fragment for p2a33-BF3149-mCherry plasmids. Table indicates the name of the final vector to be constructed, the name of the fragment for assembly, the template and primers used to obtain the fragment, and the length of this fragment. Primer sequences are listed in Table S4.1 from this chapter's Annexes.

Vector	Fragment	Template	Primers	Length of fragment (bp)
p2a33-BF1349	p2a33 backbone fw half	p2a33-mCherry	p2a33_fw p2a33_half_rv	2403
	p2a33 backbone rv half	p2a33-mCherry	p2a33_half_fw p2a33_rv	2863
	BF3149 insert	pSF-BF3149	p2a33_BfGT_fw p2a33_BfGT_rv	1149
p2a33-BF1349-mCherry	p2a33-mCherry backbone fw half	p2a33-mCherry	p2a33_mCh_fw p2a33_half_rv	3114
	p2a33-mCherry backbone rv half	p2a33-mCherry	p2a33_half_fw p2a33_mCh_rv	2863
	BF3149 insert	pSF-BF3149	BfGT_mCher_fw BfGT_mCher_rv	1149
p2a33-KRE2-BF1349	p2a33-KRE2 backbone rv half	p2a33-mCherry	p2a33_half_fw p2a33_Kre2_rv	2863
	KRE2-BF3149 insert	pSF-KRE2-BF3149	Kre2_BfGT_fw p2a33_BfGT_rv	1449
p2a33-KRE2-BF1349-mCherry	KRE2-BF3149 insert	pSF-KRE2-BF3149	Kre2_BfGT_fw BfGT_mCher_rv	1449

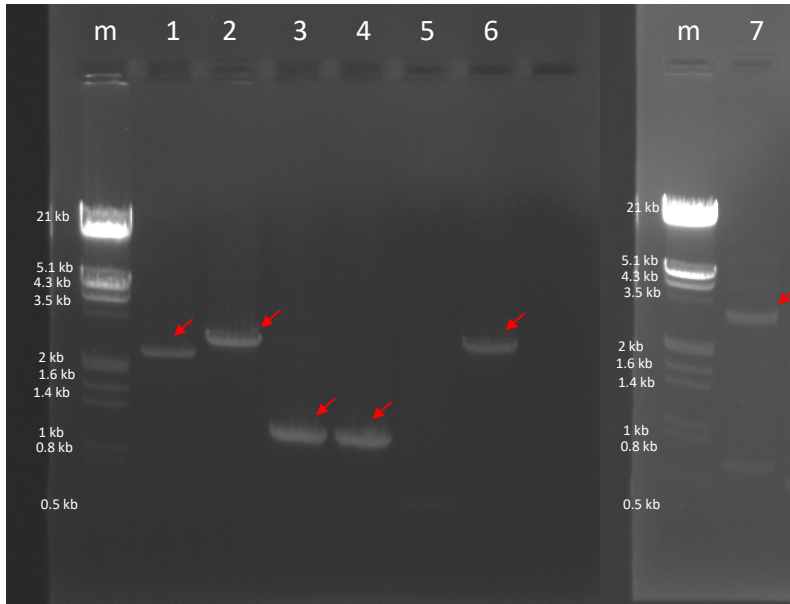


Figure 4.15 Agarose gel electrophoresis results of PCR amplification of p2a33-Bf3149 plasmids. *m* lane contains DNA marker (DNA marker III from SigmaAldrich). Arrows indicate fragments that matched the expected size. Lane 1: p2a33 backbone fw half fragment. Lane 2: p2a33 backbone rv half fragment. Lane 3: BF3149 insert for p2a33-BF3149 vector. Lane 4: BF3149 insert for p2a33-BF3149-mCherry vector. Lane 5: p2a33-mCherry backbone fw half. Lane 6: p2a33-mCherry backbone rv half. Lane 7: p2a33-mCherry backbone fw half (repetition).



4.1.3 Analysis of engineered yeast strains with p2a33 vectors

New synthesized plasmids (p2a33-BF3149, p2a33-BF3149-mCherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry) were used to transform *S. cerevisiae* strain RH6082 following the the modified Gietz protocol¹⁰⁷. p2a33-mCherry was also used as control and another strain was transformed with this plasmid, resulting in a total of five strains after transformation.

Resulting colonies from transformation showed that the strain with p2a33-mCherry showed red colonies (which indicated that construct was correctly expressed in the cytoplasm). The other strains containing fused mCherry did not show red colonies, however, this could be because of lower fluorescence due to the fusion.

To confirm the presence of the plasmids in all five strains, transformant colonies were checked by Colony PCR (primers p2a33_fw and tADH_rv) and it was confirmed that the plasmids were successfully transformed (results now shown). However, further analysis was required to confirm expression of BF3149 α -galactosylceramide synthase.

4.1.3.1 Ceramide-NBD glycosyltransferase assay with yeast cell-free extract

The first step was to check Bf3149 activity in the new host since the complications in its expression in *E. coli* could also be translated to yeast. Therefore, experiments using cell-free extract from *E. coli* expressing BF3149 were reproduced in yeast to perform ceramide-NBD- glycosyltransferase assay.

Strains analyzed in the experiment are shown in Table 4.10. *E. coli* strain expressing Strep-SUMO-BF3149 was added as control. Growth conditions were the following: A 3 mL inoculum of each strain was cultured in the appropriate selection media (Table 4.10) for 24h in agitation at 30°C. Then, the inoculum was added to 10 mL of fresh media to a final OD of 0.1. Cultures were incubated at 30°C in agitation until they reached OD 1. *E. coli* expression was carried out as described in Chapter 2.

Table 4.10 List of strains used for cell-free extract ceramide-NBD activity assay. Table includes strain name (strain + transformed plasmid) and selection media used for growth.

Strain	Selection media
RH6082 + p2a33-mCherry	SD - Ura
RH6082 + p2a33-BF3149	SD - Ura
RH6082 + p2a33-BF3149-mCherry	SD - Ura
RH6082 + p2a33-KRE2-BF3149	SD - Ura
RH6082 + p2a33-KRE2-BF3149-mCherry	SD - Ura
<i>E. coli</i> BL21(DE3)Star + pET22b-Strep-SUMO-BF3149	LB + Amp



Yeast strains did show low growth since after 8h only strain with p2a33-mCherry reached OD 1. Strains p2a33-BF3149, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry reached an OD around 0.7, however, strain p2a33-BF3149-mCherry did not significantly grow since its OD was 0.3 after 8h. Despite of this low growth, experiment was carried out with these cultures, treated following the protocol of Yeast *in vitro* galactosylceramide transferase activity assay described in the Material and methods chapter. Samples were taken at 30, 60 and 120 minutes by mixing 20 μ L of reaction with 80 μ L of methanol to stop reaction.

E. coli cells were treated as in previous galactosylceramide transferase activity assay using cell-free extract (see Material and methods chapter). Sampling was carried out at 4, 8, 12, 20 and 30 minutes, following the same procedure. This experiment would allow to see presence of ceramidases in yeast that could hydrolyze α -galactosylceramide. Then, after 60 min of *E. coli* reaction sample, the remaining volume of reaction mix was boiled for 5 minutes, and an equal volume of p2a33-mCherry yeast cell-free extract was added. Sampling was continued at 90 and 120 minutes.

Figure 4.16 shows *E. coli* control reaction results, which confirmed production of α -galactosylceramide, that is detectable at 4 min and increases to 17.7 μ M before yeast extract was added at 60 min (71% conversion). Following this addition, galactosylceramide levels did not significantly vary -a slight increase that might be explained because of some residual active enzyme. Thus, these results would confirm that there is no presence of a yeast ceramidase that is able to hydrolyze α -galactosylceramide. On the other hand, none of the yeast strains expressing Bf3149, as single or fusion protein, showed α -galactosylceramide production. No peaks were detected in HPLC chromatograms, neither in the cell-free extract nor the suspension solution that includes insoluble part of the cell. Although this could indicate that Bf3149 had no activity, it was also possible that the total amount of protein may be too low. OD did not reach the expected OD of 1 when cells were harvested for the experiment (p2a33-BF3149, p2a33-BF3149-Cherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry had an OD of 0.74, 0.29, 0.68 and 0.61, respectively).

After obtaining these results, it was decided to approach the assay as an *in vivo* ceramide-NBD activity assay.

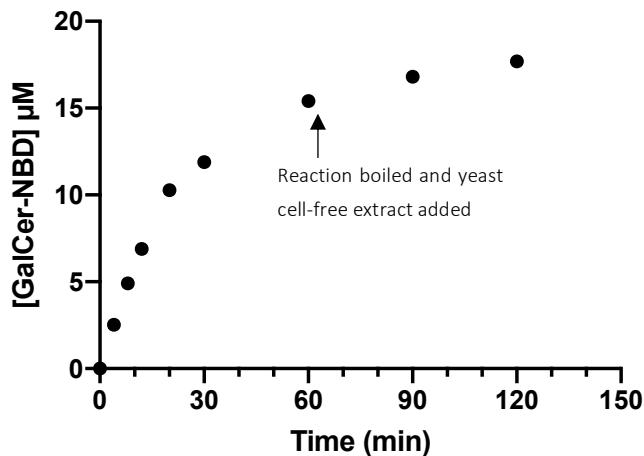


Figure 4.16 Reaction evolution of Bf3149 from *E. coli* cultures. Concentration of product (GalCer-NBD), in Y-axis, was followed for 120 min (X-axis). Arrow indicates the moment yeast extract was added to the reaction. GalCer concentration at 90 and 120 min points was adjusted after addition of yeast extract.

4.1.3.2 Ceramide-NBD glycosyltransferase *in vivo* assay

Ceramide-NBD presents the ability to enter the cell and follow the metabolism¹³⁹. Therefore, Bf3149 present in the cell might be able to use the fluorescent substrate to produce galactosylceramide. Enzyme and substrate could be closer which would improve reaction efficiency.

The same strains as in previous *in vitro* assay were tested again. Strain with p2a33-mCherry was also used as control. The protocol followed for the *in vivo* assay was adapted from Levine *et al.*¹³⁹.

Each strain was cultured until OD 1, and samples were treated following the protocol of Yeast *in vivo* galactosylceramide transferase activity assay described in the Material and methods chapter. This assay is based on the incorporation of Cer-NBD by the cells, following the metabolisms pathways, and finally being transformed into metabolites. Therefore, it was expected that yeast cells expressing Bf3149 enzyme could transform the labelled ceramide into galactosylceramide, which could be then monitored by HPLC. Cer-NBD was solubilized in BSA and mixed with cells during growth. Samples were taken at 15 and 30 min and treated as indicated in the mentioned protocol. Results were analyzed by HPLC with fluorescence detector.

Results (data not shown) from *in vivo* assay did not show any peaks corresponding to α -galactosylceramide elution in HPLC, and chromatograms from control strain with p2a33-mCherry were not different from the other strains. These results indicate that the metabolite was not produced by the strains expressing Bf3149 in the cytosol or Golgi apparatus, either as fusion or single protein.



4.1.3.3 Bf3149 location by fluorescence confocal microscopy

In order to check protein expression in the previous analyzed strains, confocal microscopy was used in strains with mCherry fusion constructs which would help know whether Bf3149 was expressed and, furthermore, where it was located in the cell. Bf3149 location would help understanding the problems with its expression in yeast. First, it would help checking the expression, and give information of where the protein is located, which could also explain the lack of activity observed in the assays. Location would be possible thanks to the red fluorescent protein mCherry. This protein is observable in a confocal fluorescence microscopy, and its location in a specific cell compartment could also indicate the presence of Bf3149.

Confocal microscopy can be used to localize the different organelles. There are three compartments that were of interest: cytoplasm (Cyt), Golgi apparatus (GA), endoplasmic reticulum (ER), nucleus (N) and cell wall (CW). Cytoplasm is indicated as a staining from cell wall to the nucleus, no dye was used to identify it in this experiment. Golgi apparatus can be distinguishable by a punctate pattern in the cell¹³⁹⁻¹⁴¹, ceramide-NBD was used to identify GA as it accumulates in this organelle^{139,141,142}. Endoplasmic reticulum usually appears as a structure that surrounds the nucleus^{130,143}, no dye was used to identify ER in this experiment. Nucleus appears as a round sphere within the cell outline, DAPI (blue fluorescence) is a common dye for the nucleus^{142,144}. Finally, cell wall is outside cell membrane and marks cell outline. No dye was used to identify cell wall, however, observing with optical microscopy allowed to see the outline of the cells.

The strains analyzed by confocal microscopy were strains with mCherry fusion proteins (p2a33-BF3149-mCherry and p2a33-KRE2-BF3149-mCherry) and strain with p2a33-mCherry. The experiment of localization with confocal microscopy was carried out following the protocol of Yeast *in vitro* galactosylceramide transferase activity assay in the Material and methods chapter.

Microscopy analysis was carried out with confocal microscope (Leica Laser Scanning Confocal Microscope TCS SP8), observing each strain under the microscope at 63x the appropriate conditions for fluorescence detection of mCherry (excitation/emission 587/610 nm), ceramide-NBD (ex/em. 466/536 nm) and DAPI (ex/em. 358/461 nm), as well as optical mode to see cell outline.

Specifics that need to be considered: First, ceramide-NBD is usually used as a specific dye for Golgi apparatus given to its tendency to accumulate in this organelle. At the microscope magnifications of the images, Golgi apparatus cannot be observed as a group of cisternae, but it appears as specific dots inside the cell, in some cases more than one can be found. This indicates the presence of different GA in the cell. In some of the results, however, the green staining appears all over the cells, which might indicate that either Cer-NBD did not enter the cell, or it was not accumulated in GA.

Although DAPI is a specific dye for the nucleus, most of the images (Figures 4.17, 4.18 and 4.19) do not show a clear position of this organelle. Probably the staining is poor and only some cells show a clear spot inside the cell (Figures 4.18 and 4.19). Although the protocol used was adapted from other



protocols for yeast, it is possible that DAPI could not enter the cell properly which led to a low nucleus staining.

Results

Results from confocal microscopy showed that mCherry was successfully expressed and localized in the cytosol in the yeast strain with p2a33-mCherry plasmid (Figure 4.17), since red color fills the cell shape. However, mCherry had a much lower fluorescence in the cells from strains expressing fusion proteins, p2a33-BF3149-mCherry and p2a33-KRE2-BF3149-mCherry (Figure 4.18 and Figure 4.19). Thus, indicating a lower protein level in these strains. This could indicate a lower level of expression for these constructs or an improper folding of the protein. This last hypothesis could explain the lack of activity that was observed in ceramide-NBD glycosyltransferase *in vitro* and *in vivo* assays.

Furthermore, the location of Bf3149-mCherry fusion in the strains expressing cytoplasmic protein (p2a33-BF3149-mCherry) is far from the location of the substrate, ceramide-NBD, which may be present in other membranes or the exterior of the cell membrane. This could also explain why *in vivo* assay did not show activity, since enzyme and substrate are not co-localized. In addition, it seems that Bf3149-mCherry might be in another compartment that is not the cytoplasm, as the distribution is not cell-wide, but localized (Figure 4.18). The fusion protein might have been transported to the vacuole for protein degradation^{145,146}, although, this hypothesis should be proved with dye or immunostaining targeting the vacuole.

Results for localization of fusion protein from p2a33-KRE2-BF3149-mCherry plasmid (Figure 4.19) are inconclusive, since they show that apparently both substrate and enzyme are co-localized but not in the GA. The image does not show the punctate pattern of Cer-NBD (specific green dots, not small, widespread dots) as in Figure 4.18, which does seem to match with GA. This could indicate that Cer-NBD was not internalized or not accumulated in the GA. However, comparing it to the location of the DAPI stain that indicates the presence of the nucleus, the red stain seems to be around it. Therefore, as explained above, this may indicate that the fusion Kre2-Bf3149-mCherry may be located in the endoplasmic reticulum. This result could mean that the fusion protein successfully entered the ER for protein synthesis, however, it could not leave the organelle to reach GA.

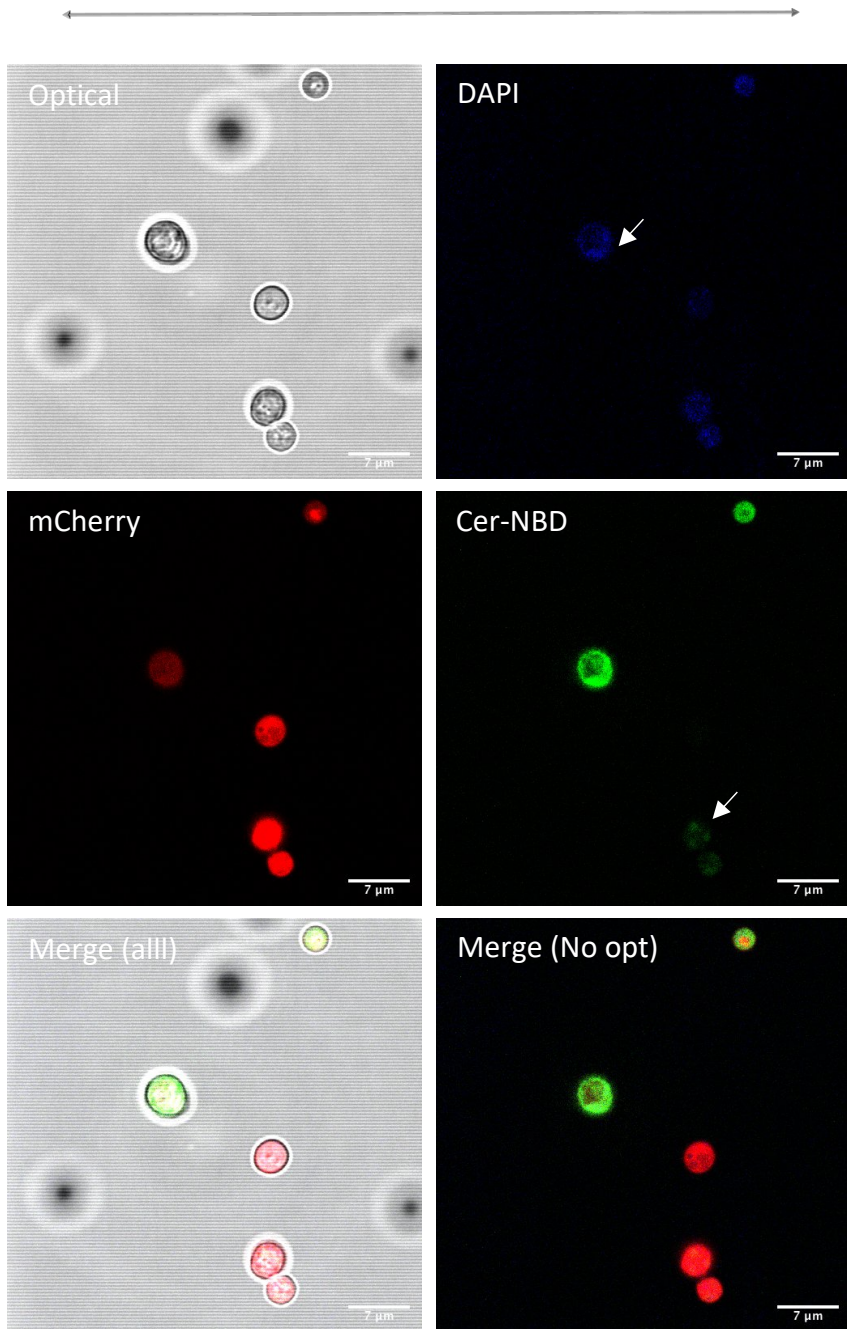


Figure 4.17 Confocal fluorescence microscopy of RH6082 + p2a33-mCherry strain. Figure shows the same image from different channels: optical (light, no fluorescence filter), DAPI (blue channel), mCherry (red channel), and Cer-NBD (green channel, excitation and emission set for ceramide-NBD). Merge (all) is a merged image of the four channels, which helps see the limits of the cells. Merge (No opt) is a merged image of the three fluorescent channels (blue, red, green) without the optical image, which shows the different stains of different colors in one picture. Note that some fluorescence can be higher than others and can overlay the other fluorescence. Arrows indicate possible organelles stained in the image: nucleus (DAPI, blue), cytosol (mCherry, red), Golgi apparatus (Cer-NBD, green).

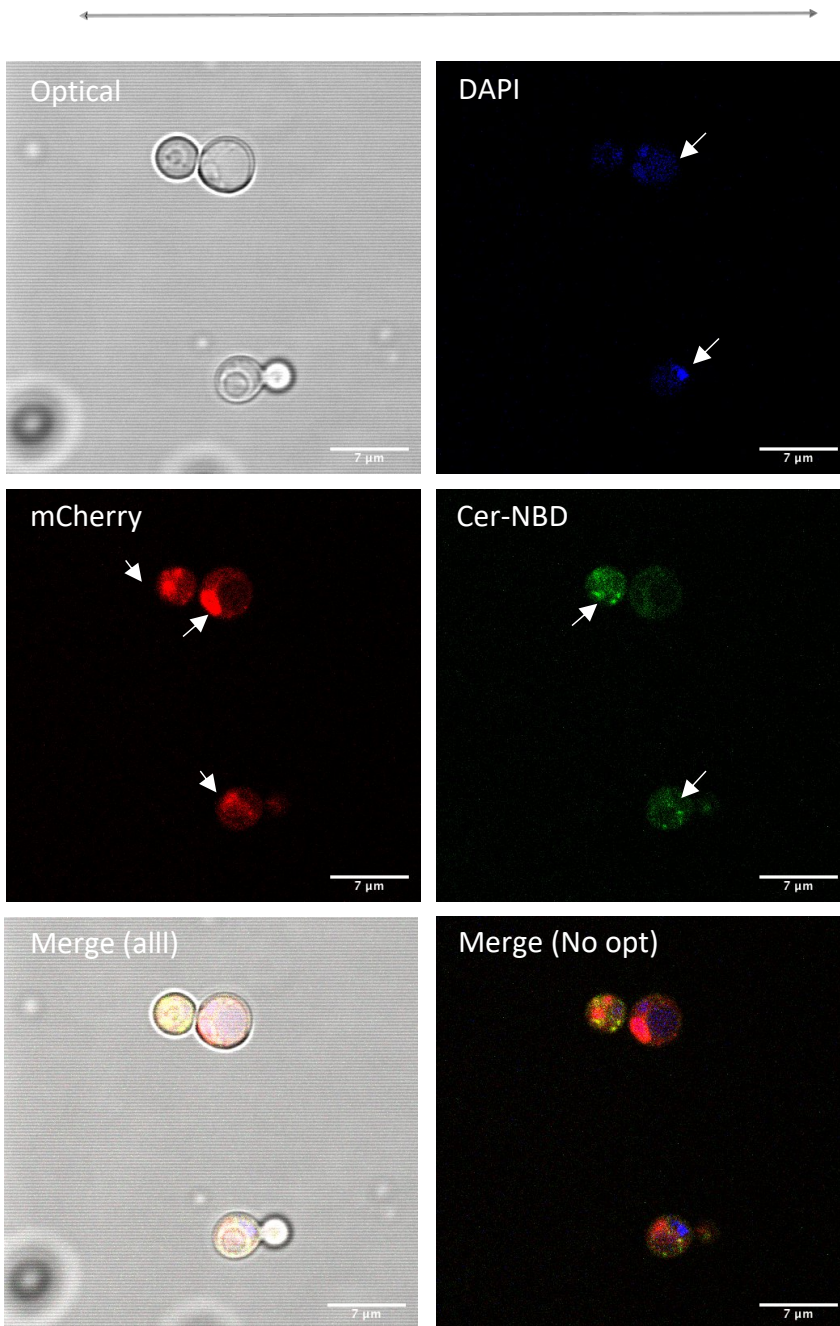


Figure 4.18 Confocal fluorescence microscopy of RH6082 + p2a33-BF3149-mCherry strain. Figure shows the same image from different channels: optical (light, no fluorescence filter), DAPI (blue channel, excitation and emission set for DAPI), mCherry (red channel, excitation and emission set for mCherry) and Cer-NBD (green channel, excitation and emission set for ceramide-NBD). Merge (all) is a merged image of the four channels, which helps see the limits of the cells. Merge (No opt) is a merged image of the three fluorescent channels (blue, red, green) without the optical image, which shows the different stains of different colors in one picture. Note that some fluorescence can be higher than others and can overlay the other fluorescence. Arrows indicate possible organelles stained in the image: nucleus (DAPI, blue), Golgi apparatus (Cer-NBD, green). Red stain indicate presence of BF3149-mCherry construct.

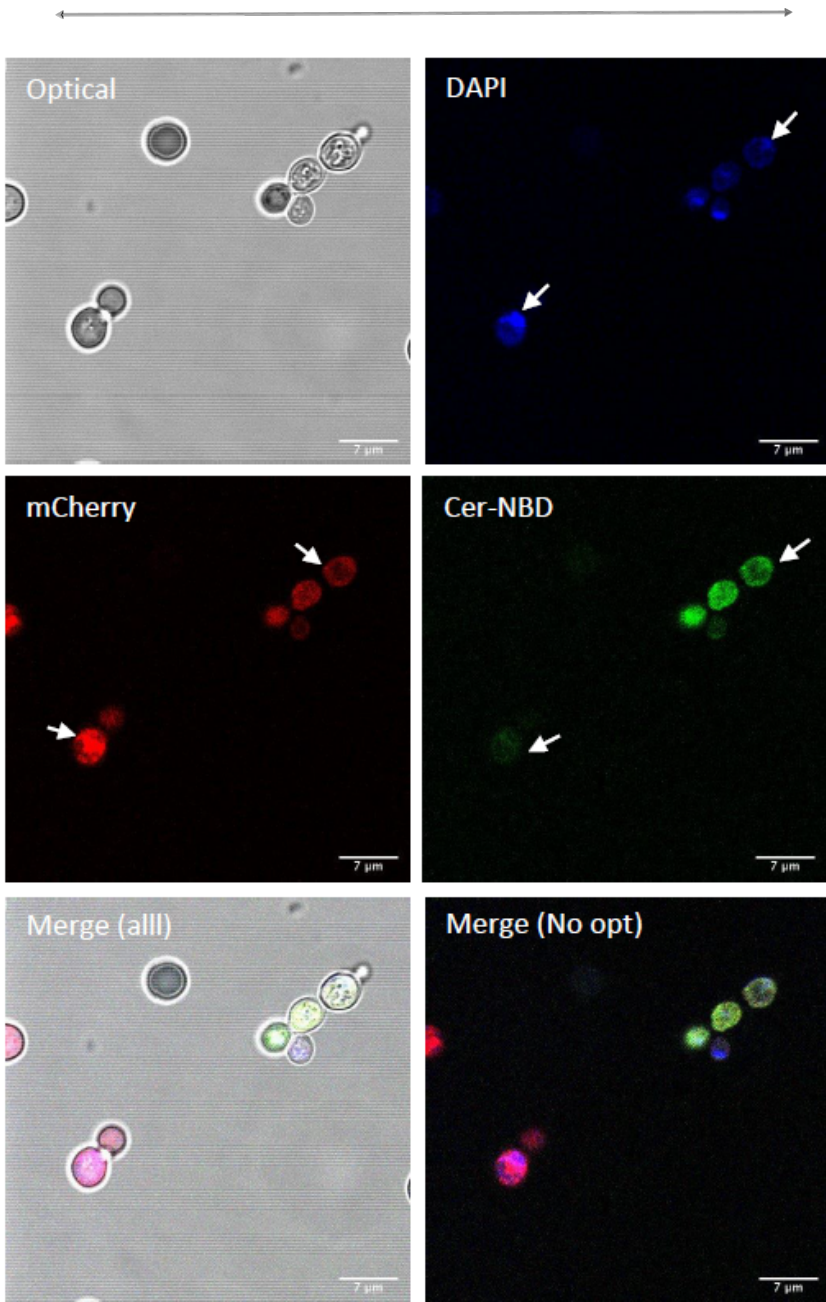


Figure 4.19 Confocal fluorescence microscopy of RH6082 + p2a33-Kre2-BF3149-mCherry strain. Figure shows the same image from different channels: optical (light, no fluorescence filter), DAPI (blue channel, excitation and emission set for DAPI), mCherry (red channel, excitation and emission set for mCherry) and Cer-NBD (green channel, excitation and emission set for ceramide-NBD). Merge (all) is a merged image of the four channels, which helps see the limits of the cells. Merge (No opt) is a merged image of the three fluorescent channels (blue, red, green) without the optical image, which shows the different stains of different colors in one picture. Note that some fluorescence can be higher than others and can overlay the other fluorescence. Arrows indicate possible organelles stained in the image: nucleus (DAPI, blue), endoplasmic reticulum (mCherry, red), Golgi apparatus (Cer-NBD, green). Red stain indicates presence of Kre2-Bf3149-mCherry construct.



These results show the need to further expand the knowledge on Bf3149 α -GalCer synthase. The complications found on these yeast strains could be caused by the poor stability of the enzyme. Which may lead to low protein content and, thus, to low activity. This was also observed in *E. coli*, where low protein stability and low protein levels were obtained.

Knowing the existence of a protein that could confer stability to Bf3149 or improving the enzyme would also be helpful in its expression, since it could stabilize and increase α -GalCer synthase activity. As a conclusion, other relocation approaches should be tested since the attempts completed in this thesis were not successful, even though the possibility that these results might have been affected by protein low stability must be kept into consideration.



4.2 Summary

For Bf3149 expression in *S. cerevisiae*, two set of different vectors were used to clone Bf3149 sequence. pSF-TPI1-URA3 derived vectors with Bf3149, mKate2 and fusion protein Bf3149-mKate2 for different locations, were successfully synthesized, however, no red colonies appeared in yeast strains expressing single mKate2, which indicated that construct was not expressed.

Then, it was decided to construct p2a33-mCherry derived plasmids, with fusion of Bf3149-Cherry and for cytoplasm and Golgi apparatus localization. The new plasmids were successfully synthesized, and colonies from control yeast strain with p2a33-mCherry did appear as reddish colonies, which indicated that mCherry, was expressed and constructs were viable. These strains expressing Bf3149 were analyzed for ceramide-NBD galactosyltransferase activity.

Ceramide-NBD *in vitro* assays did not show any activity of Bf3149 in engineered strains, since no galactosylceramide product was identified after HPLC analysis. Furthermore, *in vivo* analysis of the same engineered strains did not show activity neither.

Finally, confocal microscopy was used to analyze Bf3149, fused with mCherry, expression and location in the cell. Microscopy images showed that Bf3149 content in the cell was very low, which could be caused by low expression or low protein stability. Furthermore, Bf3149 did not seem to co-locate with its substrate, ceramide-NBD, which could also explain the lack of activity. Last, Kre2-Bf3149-mCherry was not present in the GA as it was expected and appeared to be in the ER. Therefore, location approach in further experiments should be reconsidered in order to improve the location of Bf3149. Another hypothesis considered was that these constructs were transported to the vacuole for degradation, however, this explanation could not be proved.

4.3 Annexes Chapter 4

Table S4.1 List of primers used in Chapter 4 and their sequences. Table indicates name of the primer as mentioned in the chapter, sequence, and length (Lngh). Underlined nucleotides indicate the added homology region for assembly.

Name	Sequence	Lngh
cyt_bb_fw	<u>TCAAGCGTAACGGCAACAAGT</u> GACTGACTGATACAATCGATTTC	44
cyt_bb_rv	<u>TCATTAACAAACCAATCAT</u> GGTTGGTACCTCCTTTGAATTC	42
cyt_BfGT_fw	<u>ATTCAAAGGAGGTACCAACCAT</u> GATTGGTTTGTTAATGATTGC	44
cyt_BfGT_rv	<u>TCGATTGTATCAGTCAGTCACT</u> GTTGCCGTTACGCTTG	39
ER_bb_fw	<u>GCAACAAGCATGATGAATTG</u> TGACTGACTGATACAATCGATTTC	44
ER_bb_rv	<u>AAGATGCTTGAAATCTCAT</u> GGTTGGTACCTCCTTTGAATTC	42
ER_aMF_fw	<u>ATTCAAAGGAGGTACCAACCAT</u> GAGATTTCCAAGCATCTTC	41
ER_aMF_rv	<u>TCATTAACAAACCAATCAT</u> TGCTTCTGCCTCTCTCTTC	39
ER_BfGT_fw	<u>AGAAGAGAGAGGCAGAAGCAAT</u> GATTGGTTTGTTAATGATTGC	44
ER_BfGT_rv	<u>GTCAGTCACAATTCATCAT</u> GTCTGTTGCCGTTACGCTTG	39
GA_bb_rv	<u>TTACTGAGAAAGAGGGCCAT</u> GGTTGGTACCTCCTTTGAATTC	42
GA_Kre2_fw	<u>ATTCAAAGGAGGTACCAACCAT</u> GGCCCTCTTCTCAGTAAG	41
GA_Kre2_rv	<u>TCATTAACAAACCAATCAT</u> CGGGGCATCTGCCTTTTC	38
GA_BfGT_fw	<u>CTGAAAAGGCAGATGCCCGAT</u> GATTGGTTTGTTAATGATTGC	44
cyt_bb_mK2_fw	<u>CTAGCAAACCTGGGGCACAGAT</u> GACTGACTGATACAATCGATTTC	44
cyt_bb_mK2_rv	<u>TTAATCAGCTCGCTCACCAT</u> CTTGTGCCGTTACGCTTG	39
cyt_mK2_fw	<u>TCAAGCGTAACGGCAACAAGAT</u> GGTGAGCGAGCTGATTAAG	41
cyt_mK2_rv	<u>TCGATTGTATCAGTCAGTCACT</u> GTGCCCCAGTTTGCTAG	40
ER_bb_mK2_fw	<u>CTAGCAAACCTGGGGCACAGACAT</u> GATGAATTGTGACTGACTG	42
ER_mK2_rv	<u>GTCAGTCACAATTCATCAT</u> GTCTGTGCCCCAGTTTGCTAG	40
ColPCR_pSF_fw	TTTACTATTTCCCTTCTTACG	22
ColPCR_pSF_rv	GAATGTAAGCGTGACATAAC	20
pSF_bb_half_rv	GTGCGCTCTATAATGCACTC	20
pSF-bb_half_fw	TCTAACCTTAACGGACCTAC	20
cyt_bb_mK2_rv	<u>TTAATCAGCTCGCTCACCAT</u> GGTTGGTACCTCCTTTGAATTC	42
mK2_cyt_fw	<u>ATTCAAAGGAGGTACCAACCAT</u> GGTGAGCGAGCTGATTAAG	41
GA_mK2_bb_rv	<u>TTAATCAGCTCGCTCACCAT</u> CGGGGCATCTGCCTTTTC	38
mK2_GA_fw	<u>CTGAAAAGGCAGATGCCCGAT</u> GGTGAGCGAGCTGATTAAG	41
ER_mK2_bb_rv	<u>TTAATCAGCTCGCTCACCAT</u> TGCTTCTGCCTCTCTCTTC	39
mK2_ER_fw	<u>AGAAGAGAGAGGCAGAAGCAAT</u> GGTGAGCGAGCTGATTAAG	41

p2a33_fw	<u>AGCGTAACGGCAACAAGTAAGGCGCGCCACTTCTAAATAAG</u>	41
p2a33_half_rv	CAATACGCAAACCGCCTCTC	20
p2a33_half_fw	TAATGAATCGGCCAACGC	18
p2a33_rv	<u>TCATTAACAAACCAATCATTTTGTAAATAAAACCTTAGATTAGATTGCTAT</u> G	52
p2a33_BfGT_fw	<u>ATCTAAGTTTTAATTACAAAATGATTGGTTTGTTAATGATTGCTTC</u>	47
p2a33_BfGT_rv	<u>TTATTTAGAAAGTGGCGCGCCCTTGTTGCCGTTACGCTTG</u>	39
p2a33_mCh_fw	GCTAGCGCAACCGGACGCCACCATGGTGAGCAAGGGCGAG	40
p2a33_mCh_rv	<u>TCATTAACAAACCAATCATTTTGTAAATAAAACCTTAGATTAGATTGCTAT</u> GC	53
BfGT_mCher_fw	<u>ATCTAAGTTTTAATTACAAAATGATTGGTTTGTTAATGATTGCTTC</u>	47
BfGT_mCher_rv	<u>TGGCGTCCGGTTGCGCTAGCGTCTTGTTGCCGTTACGCTTG</u>	41
p2a33_Kre2_rv	<u>TTACTGAGAAAGAGGGCCATTTTGTAAATAAAACCTTAGATTAGATTGC</u>	48
Kre2_BfGT_fw	<u>ATCTAAGTTTTAATTACAAAATGGCCCTCTTCTCAGTAAG</u>	41
p2a33_ColPCR_fw	ACGGTCTTCAATTTCTCAAG	20
tADH1	TGACCTACAGGAAAGAGTTAC	21

Table S4.2 mKate2 gene and protein sequence. Start codon is indicated with bold letters. Stop codon is indicated with underlined letters. End of protein is indicated by *.

Gene Name	mKate2		
GenBank Accession number	-		
DNA sequence	Length	699	bp
ATG GTGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGAACAACCACCACTTCAA GTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATCAAGGCGGTCGAGGGCGGCCCT CTCCCCTTCGCTTCGACATCTGGCTACCAGCTTATGTACGGCAGCAAAACCTTATCAACCACACCCAGGGCATCCC CGACTTCTTAAGCAGTCCTTCCCCGAGGGCTTCACATGGGAGAGAGTCACCACATACGAAGACGGGGCGTGTCTGAC CGTACCCAGGACACCCAGCTCCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACCTCCCATCAA CGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTCCACCGAGACCCTGTACCCCGCTGACGGCGCCTGG AAGGCAGAGCCGACATGGCCCTGAAGCTCGTGGGCGGGGCCACCTGATCTGCAACTGAAGACCACATACAGATCC AAGAAACCCGCTAAGAACCTCAAGATGCCCGCGTCTACTATGTGGACAGAAGACTGGAAAGAATCAAGGAGGCCGA CAAAGAGACCTACGTCGAGCAGCAGAGGTGGCTGTGGCCAGATACTGCGACCTCCCTAGCAAACCTGGGGCAGAGAT <u>AA</u>			
Protein sequence	Length	233	aa
MVSELIKENMHMKLYMEGTVNHHFKCTSEGEKPYEGTQTMRIKAVEGGPLPFAFDILATSFMYGSKTFINHTQGPIDFF KQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFPNPGVPMQKTLGWEASTETLYPADGGLEGRADM ALKLVGGGHLICNLKTTYRSKPKAKNLKMPGVYVDRRLERIKEADKETYVEQHEVAARYCDLPSKLGHR*			



Table S4.3 Truncated *KRE2* DNA (300 bp) and protein sequence (100 residues). Start codon is indicated with bold letters.

Gene Name	<i>KRE2</i>		
GenBank Accession number	NM_001180791		
DNA sequence	Length	300	bp
ATGGCCCTCTTTCTCAGTAAGAGACTGTTGAGATTTACCGTCATTGCAGGTGCGGTTATTGTTCTCCTCCTAACATTGAA TTCCAACAGTAGAACTCAGCAATATATCCGAGTTCATCTCCGCTGCATTTGATTTTACCTCAGGATCTATATCCCCTGA ACAACAAGTCATCTCTGAGGAAAATGATGCTAAAAAATTAGAGCAAAGTGCTCTGAATTCAGAGGCAAGCGAAGACTC CGAAGCCATGGATGAAGAATCCAAGGCTCTGAAAGCTGCCGCTGAAAAGGCAGATGCCCCG			
Protein sequence	Length	100	aa
MALFLSKRLLRFTVIAGAVIVLLLTLNSNSRTQYIPSSISAAFDFTSGSISPEQQVISEENDAKKLEQSALNSEASEDSEAMDEE SKALKAAAEKADAP			

Table S4.4 Alpha Mating Factor sequence (183 bp) and protein sequence (61 residues). Start codon is indicated with bold letters.

Gene Name	Alpha Mating Factor		
GenBank Accession number	-		
DNA sequence	Length	183	bp
ATGAGATTTCCAAGCATCTTCACTGCTGTTTTGTTGCGCAGCGATTCTGCCCTAGCGGCACCCGTCAATACTACACAGA GGATGAGCTGGAGGGTGATTTTGATGTCGCCGCTTACCGTTTAGTGCATCAATCGCCGCTAAAGAAGAAGGAGTGTC ACTGGAGAAGAGAGAGGCAGAAGCA			
Protein sequence	Length	61	aa
MRFPSIFTAVLFAASSALAAPVNTTTEDELEGDFDVAVLPFSASIAAKEEGVSLEKREAEA			

Table S4.5 Linker DNA and protein sequence.

Gene Name	Linker sequence		
GenBank Accession number	-		
DNA sequence	Length	24	bp
ACGCTAGCGCAACCGGACGCCACC			
Protein sequence	Length	8	aa
TLAQPDAT			

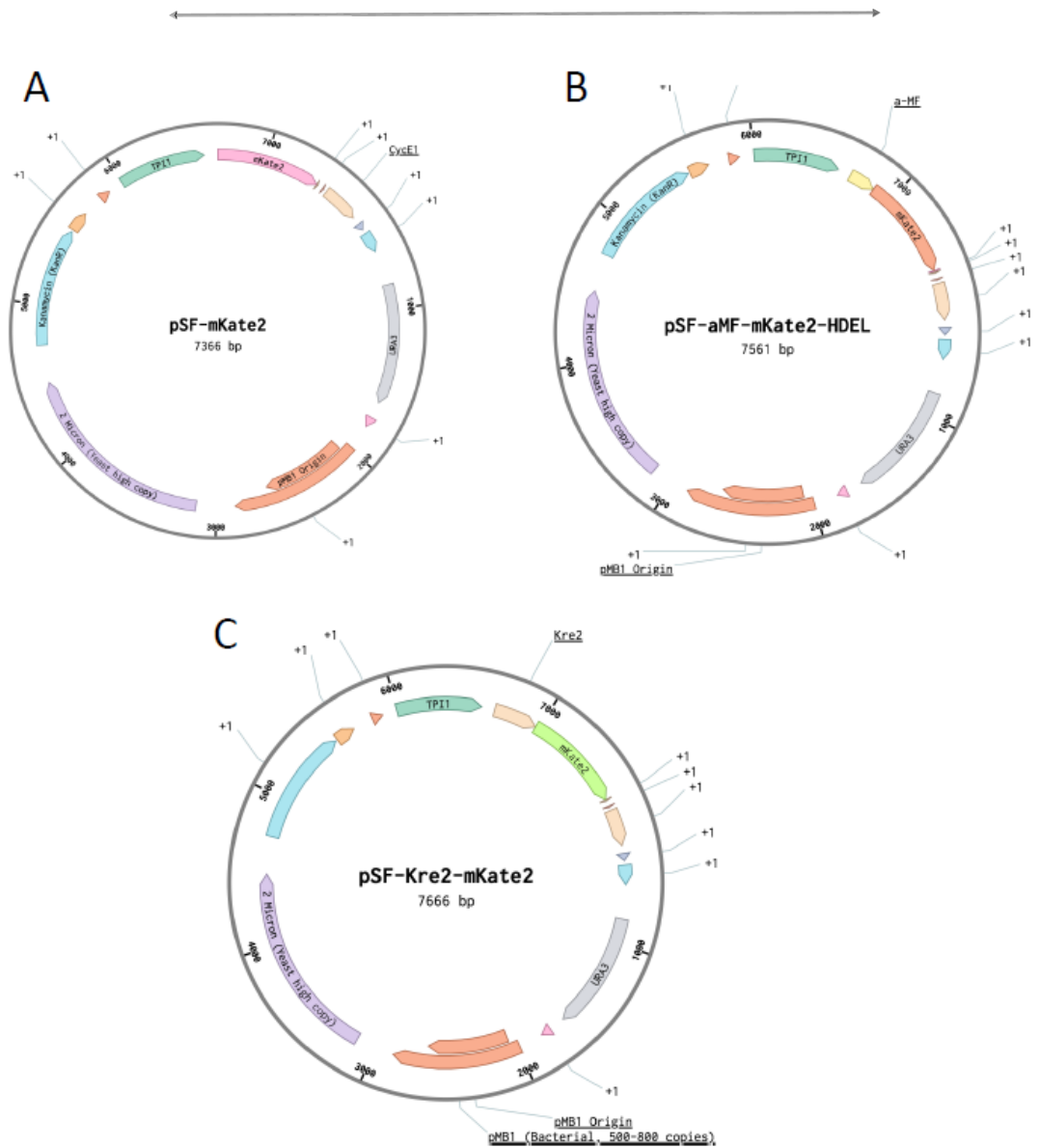


Figure S4.1 Vector map of pSF-mKate2 (A), pSF- α MF-mKate2-HDEL (B) and pSF-KRE2-mKate2 (C). In the center, the name of the vector and its size. Legend: Light blue, Kanamycin R. Orange, RnG Bacterial Terminator. Red (short), 5' HS1 BetaGlobulin Insulator. Green, *TPI1* promoter. Brown (short), Stop codon x3. Light brown, *CycE1* terminator. Blue (short), T7 terminator. Light blue, RnG Bacterial terminator. Grey, *URA3* selection marker. Pink, 3' HS1 BetaGlobulin Insulator. Red, *pMB1* origin. Purple, 2 μ origin. Yellow (B), α Mating Factor signal peptide. Light brown (C), *KRE2* sequence. Pink (A), Red (B) and Green (C), *mKate2* gene.



Final discussion



Final discussion



Final discussion



Final discussion

Before the conclusions of this thesis, here we discuss the main results obtained throughout the project, as well as the possibilities that are open for future work.

The idea of this project originated with the evidence that *Bacteroides fragilis* can produce α -galactosylceramide³⁷. The lack of synthetic alternatives of this compound opens an opportunity to use this knowledge to explore new approaches for α -GalCer production. This project planned to create a biological platform to synthesize the product using metabolic engineering. *Saccharomyces cerevisiae* is a good candidate as a host since it can produce the precursors for α -GalCer, UDP-galactose and phytoceramide, and many molecular biology techniques are available.

Some gene modifications involved in ceramide metabolism have been tested to increase or reduce yeast ceramide or glycosphingolipids in the cell^{88, 89}. In this project, three key genes were identified as key in redirecting metabolic flux towards phytoceramide synthesis: i) *SUR2* (sphingosine hydroxylase responsible of conversion of dihydroceramide and dihydrosphingosine into phytoceramide and phytosphingosine, respectively), ii) *SCS7* (very long chain fatty acid hydroxylase which catalyzes the α -hydroxylation of ceramide acyl chains) and iii) *ISC1* (phosphoinositide phospholipase C that performs the hydrolysis of complex sphingolipids to release ceramide). Therefore, three modifications were planned to forward metabolism to phytoceramide pool: overexpression of *SUR2* and *ISC1* genes, and deletion of *SCS7* gene. In addition, the glycosyltransferase that catalyzes α -GalCer synthesis was identified and characterized in this project and planned to be expressed in the *S. cerevisiae* to produce the product of interest.

After the design of the host strain, the modifications were carried out in the RH6082 *S. cerevisiae* strain. On one hand, *SUR2* and *ISC1* genes were cloned into two yeast expression vectors using constructs with promoter and terminator of similar strength. The constructs were successfully obtained, and gene expression levels were quantified by qPCR. *SUR2* expression was confirmed to be 144 times higher than that of the control strain, but *ISC1* could not be determined. Probably the primers were dimerizing and in future work the use of new primers is recommended. On the other hand, *SCS7* deletion was carried out using CRISPR/Cas9 technology. Although the plasmids required for the gene knockout were successfully designed and constructed (*SCS7* gRNA plasmid and *SCS7* donor DNA plasmid), no positive colonies appeared after transforming yeast with gRNA plasmid, Cas9 expression plasmid and the linear donor DNA. Therefore, future work is needed for *SCS7* gene knockout.

A breakthrough of this project was the identification of the glycosyltransferase from *B. fragilis* capable of synthesizing α -galactosylceramide: the α -GalCer synthase BF9343_3149. Although the discovery was reported by Okino *et al.*³⁸ during this project, our findings came to the same conclusion. Nonetheless, we approached the identification of candidate glycosyltransferase as it follows: *B. fragilis* GT4 sequences annotated in the CAZY database were pooled with two monoglucosyldiacylglycerol synthases (from *Acholeplasma laidlawii*^{116,117} and *Streptococcus pneumoniae*¹¹⁸) as ceramide and

diacylglycerol are similar structures and we hypothesized that the α -GalCer synthase would be similar to the diacylglycerol synthases. After multiple sequence alignment of all sequences (all *B. fragilis* GT4 + 2 monoglucosyldiacylglycerol synthases) and BLAST search, 4 candidates were identified. However, only BF9343_3149 showed glycosyltransferase activity and was producing α -GalCer from UDP-Gal and ceramide derivative.

Biochemical characterization of the α -GalCer synthase showed that the enzyme is a non-processive GT that maintains the α linkage in the final product. First reactions were carried out using the enzyme expressed in pET28a-BF3149 plasmid in *E. coli* but protein expression was low, and protein could not be recovered with His-tag purification. Therefore, next experiments were carried out using a new vector with the protein fused with SUMO at N-terminus^{123,124}, and Strep-tag for purification. Expression levels improved and 0.6 mg/L of pure enzyme were obtained from cultures. However, the enzyme was quite unstable and after 24h almost all activity was lost. Buffer containing glycerol helped maintaining activity quite stable in the cell-free extract fraction. Further work needs to focus on optimizing protein expression and improve enzyme stability since it was the main obstacle to run the reactions for the biochemical characterization (reactions were always carried out within 2 hours after purification, which was a window in which the enzyme activity remained stable).

Further characterization revealed that the enzyme activity was higher at pH 7.3 and temperature range between 30-35°C. Enzyme is active in a pH range between 5 and 9.5, outside that range activity was greatly reduced. As for temperature, at 55°C enzyme was inactive, which indicates a poor thermostability. Since some GT usually have metal requirements^{42, 125}, metallodependence was studied. EDTA did not reduce the activity and no metal cations improved enzyme activity. The results were not expected, as other GT-B enzymes that do not have de DXD signature still show metallodependent activities, such as GlcAT from *Z. mobilis* and α -monoglucosyldiacylglycerol synthase from *A. laidlawii*^{38,117}. Although evidence suggests that there is no metal requirement for BF3149 glycosyltransferase, it is still possible that the metal is strongly bound to the enzyme and other cations or EDTA cannot displace the bound cation.

Substrate preference was also analyzed: UDP-galactose is preferred over UDP-glucose, showing a 14.5-fold higher specific activity. This agrees with the results from Okino *et al.*³⁸, in which those were the only substrate accepted by α -GalCer synthase, tested together with UDP-GalNAc, UDP-GlcA, UDP-GlcNAc and GDP-Man. Michaelis-Menten parameters were determined for UDP-galactose and also ceramide. The enzyme presents a K_M of 218 μ M for UDP-galactose, with a slight substrate inhibition at concentration above 2 mM ($K_I = 3.78$ mM); as for ceramide, calculated K_M was 4.6 μ M and substrate inhibition was higher, with an accented reduction in enzyme velocity at concentrations above 50 μ M ($K_I = 161$ μ M).

As other glycolipid synthases such as α -monoglucosyldiacylglycerol synthase from *A. laidlawii* show activation by anionic lipids, BF3149 α -GalCer synthase was also tested in the presence of anionic and neutral lipids such as DOPG and DOPC respectively. For that, the ceramide acceptor was solubilized in vesicles containing DOPC, DOPG or a mixture. Results showed no activation by anionic lipids and the

activity was inhibited in the presence of detergents (in addition to DOPG and DOPC, CHAPS was previously tested and showed enzyme inhibition). Optimal activity was obtained when the lipid acceptor is solubilized with BSA. These results may indicate that the enzyme takes the acceptor when presented by a lipid-protein complex rather than from membranes. Further experiments can be performed to elucidate the mechanism of acceptor interaction.

Finally, structural analysis on the α -GalCer synthase showed the enzyme has the common GT-B fold structure and the typical EX₇E motif essential for activity of GT4 enzymes¹²⁷, but with a substitution of an Asp for the first Glu. Moreover, the three-dimensional model of α -GalCer synthase with UDP-Gal allowed to locate the main residues that interact with the donor substrate. This model could open a new approach of protein engineering to improve the enzyme.

Finally, the BF3149 α -GalCer synthase gene was expressed in the yeast strain RH6082. Three approaches were considered to redirect the enzyme to specific compartments. First, expression in the cytoplasm would allow the enzyme to be potentially able to find and use the substrates. Second, redirection to the Golgi apparatus would allow the enzyme to find UDP-galactose and ceramide, since in this compartment glycosylation of other molecules and complex sphingolipid synthesis occurs. Third, the endoplasmic reticulum allocates the synthesis of ceramide, therefore, relocation in this compartment would place the enzyme near a high availability of its substrate.

The first set of plasmids were constructed with pSF-TPI1-URA3 derived vectors with BF3149, mKate2 and fusion protein BF3149-mKate2 for different locations. For cytoplasm expression, no modification was added. For Golgi apparatus redirection, the N-terminus of Kre2 protein (a native Golgi apparatus protein) was fused to N-terminus of the construct to anchor the protein in Golgi membrane^{136, 137}. Finally, for endoplasmic reticulum, α mating factor signal peptide was fused to the N-terminus of the construct, and the retaining sequence HDEL was added to the C-terminus^{130,133,134}. Nonetheless, after construction of the plasmids, no expression was observed for reasons that are not clear, and a new set of vectors using p2a33-mCherry plasmid as backbone were designed. Constructs with Bf3149 fused and non-fused to mCherry were synthesized for expression in cytoplasm and Golgi apparatus.

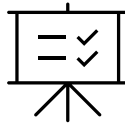
After transforming *S. cerevisiae* with the new plasmids, glycosyltransferase activity was analyzed. First, strains expressing BF3149 were tested using soluble and non-soluble fractions of the cultures after lysis, however, no product formation was observed. Next, an *in vivo* glycosyltransferase assay was carried out using the same strains expressing BF3149 with the different constructs in the presence of ceramide-NBD, as it can be internalized in the cell¹³⁹. However, no activity was observed when extracts from the reactions were analyzed by HPLC.

Finally, construct location of Bf3149 fused with mCherry was observed in the different engineered strains (cytoplasm and Golgi apparatus expression) by confocal microscopy. Bf3149 had a lower presence compared to mCherry, which could indicate low expression or low protein stability. Moreover, the construct did not co-locate with the substrate ceramide-NBD, which could explain the lack of activity in these strains. From the images, it appeared that the construct could be in the vacuole,

Conclusions



Conclusions



Conclusions





Conclusions

1. After a bibliography study of the metabolic routes for glycosphingolipid synthesis in the yeast *Saccharomyces cerevisiae* and the modifications carried out in these pathways, three enzymes, *SUR2*, *SCS7* and *ISC1*, were identified to be key for increasing ceramide levels and redirect metabolic flux towards the synthesis of phytoceramide.
2. Two vectors were successfully constructed for yeast ceramide metabolic engineering approach: pTDH3-SUR2 and pTEF2-ISC1. Expression analysis only proved *SUR2* to be overexpressed in the yeast host strain, whereas *ISC1* remains to be checked. Although *SCS7* knockout plasmids were successfully synthesized, the attempts showed that *SCS7* gene knockout was not performed.
3. The microorganism *Bacteroides fragilis* was found to be a natural producer of α -galactosylceramide. Multiple sequence alignment of two known retaining glycosyltransferases from Family 4 (non-processive monoglucosyldiacylglycerol synthases from *Acholeplasma laidlawii* and *Streptococcus pneumoniae*) with *B. fragilis* GT4 sequences, and BLAST search (including non-classified GT4 sequences), resulted in four candidate sequences to putatively have α -galactosylceramide synthase activity: BF9343_3149, and BF9343_1306, BF9343_3589 and BF9343_0585. Only the previously reported BF9343_3149 (also named as BF3149 or α -GalCer synthase/GT) proved to be able to produce α -galactosylceramide.
4. Bf3149 GT showed common features of family 4 of glycosyltransferases: GT-B structure with two clear domains was observed using AlphaFold model. Furthermore, it conserves the typical EX₇E motif essential for activity in GT4 enzymes, but with an Asp for the first Glu.
5. Bf3149 expression was low, and protein remained predominantly in the insoluble extract after disruption. Fusion with SUMO protein and use of Strep tag helped in increasing its solubilization and purification to about 0.6 mg per liter. Protein stability was observed to be poor, being virtually inactive at 24h in phosphate or Tris buffer. Activity assays were carried out within a window of 2h after purification.
6. Bf3149 α -GalCer synthase proved to be a non-processive GT since a single product was obtained. Preferred donor is UDP-Gal over UDP-Glc (14.5-fold higher specific activity for the first). Maximal activity is obtained at pH 7.3 (conserving >70% of activity within pH 6-8.5) and between 30-35°C, being inactive at 55°C, indicating a poor thermostability.



7. Although GT enzymes with GT-A fold are metallodependent enzymes with the conserved DXD motif that binds metal cation, GT-B enzymes do not have the DXD motif but they still can show metallodependence activity. BF3149 enzyme, as a GT-B enzyme, does not show the DXD signature, and no evidence indicated a cation requirement. Addition of Mg^{2+} or Mn^{2+} did not affect activity, but presence of Zn^{2+} inactivated the enzyme. Furthermore, EDTA at 200 mM did not remove activity from the enzyme.
8. Ceramide-C6-NBD was used as lipid acceptor. Although conversion with this acceptor was near 100% at high protein concentrations, it was observed that presentation of the lipid acceptor by a solubilizing protein (BSA) was critical for enzyme activity. When Cer-NBD was solubilized in DOPC and/or DOPG mixed cells, activity was highly reduced and no activation by anionic lipids was observed. CHAPS presence in lysis buffer also inactivated the enzyme, confirming previous results with other detergents.
9. A set of plasmids based on pSF-TPI1-URA vector was constructed with different strategies for Bf3149 enzyme relocation in yeast cytosol, Golgi apparatus and endoplasmic reticulum, and with fusion with mKate2 fluorescent protein. After transformation in yeast, constructs appeared not to be expressed, so new vectors were constructed.
10. A different set of vectors using p2a33-mCherry as backbone were constructed, including relocation strategies for yeast cytoplasm and Golgi apparatus, as well as fusion of Bf3149 with mCherry. After transformation, constructs appeared to be expressed as red colonies were observed for strains expressing mCherry.
11. Analysis of five RH6082 yeast strains with each constructed plasmid (p2a33-mCherry, p2a33-BF3149, p2a33-BF3149-mCherry, p2a33-KRE2-BF3149 and p2a33-KRE2-BF3149-mCherry) by Cer-NBD *in vitro* and *in vivo* assays did not show galactosyltransferase activity in strains expressing Bf3149.
12. Confocal microscopy of strains expressing mCherry and Bf3149-mCherry fusions showed that all constructs were expressed. However, cellular localization of Bf3149-mCherry and KRE2-Bf3149-mCherry did not match with the expected organelles and appeared to not colocalize with the ceramide-NBD substrate.
13. Future work should include (1) the construction of the final strain with optimized ceramide production, (2) design new approaches for protein relocation (using Bf3149 or other ceramide GT), and (3) improve Bf3149 enzyme to enhance its activity.

Conclusions



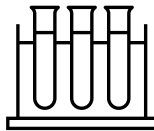
14. Future perspectives of this thesis open new possibilities on identifying new α -GalCer synthases in other organisms, engineer yeast metabolism to enhance C26-phytoceramide or other structures, and use new α -GalCer synthases with improved biochemical features to enhance α -GalCer production in the host strain.

Conclusions





Material and methods



Material and methods



Basic biochemistry and molecular biology protocols and materials

Strains and media

DH5 α *E. coli* cells were used for cloning and molecular biology work. Cells were cultured in LB broth with the appropriate antibiotic at 37°C, either on liquid culture or agar plates. BL21(DE3)Star *E. coli* cells were used for protein expression. *E. coli* cells were cultured in liquid or solid media (with agar) with Luria-Bertani broth supplemented with the appropriate antibiotic (ampicillin at 100 μ g/mL or kanamycin 50 μ g/mL).

As for yeast strains. RH6082 strain was provided by Dr. Howard Riezman and used as host for the metabolic engineering approach. In addition, yeast strain YM4271 and Sc1170, available in the Laboratory of Biochemistry (IQS-URL), were used for qPCR analysis and genomic DNA extraction, respectively. *S. cerevisiae* were cultured in liquid or solid media (with agar) with YPD (Yeast Extract-Peptide-Dextrose) supplemented with Uracil and Adenine (also named as YPUAD). For selection media, SD (Synthetic Defined) media was prepared with Yeast Nitrogen Base (YNB) and glucose. For each selection marker, the appropriate amount of amino acids were added except the auxotrophy marker. Adenine and Uracil were also supplemented to SD media, except when URA3 selection marker was used.

Plasmids

pET28a and pET22b were commercial plasmids acquired from Novagen. pET22b-Strep-SUMO-mKate2 was already available in the Laboratory of Biochemistry and constructed in previous studies. pHES834 plasmid was acquired from Addgene. pTDH3-ScoMG517 was available in the Laboratory of Biochemistry from IQS, synthesized in collaboration with the group of Prof. Marjan de Mey of the Center for Synthetic Biology from Ghent University. pSF-TPI1-URA3 plasmid was purchased from Merck. p2a33-mCherry was provided by Dr. Marjan de Mey from the Center for Synthetic Biology from Ghent University.

Reagents

Ceramide-C6-NBD (Cer-NBD) (N-[6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl]-D-erythro-sphingosine) was provided by the synthesis facility at the Institut de Química Avançada de Catalunya, Barcelona, Spain. DAPI solution (4',6-diamidino-2-phenylindol) was kindly provided by Dra. Cristina Fornaguera from the Material Engineering group (GEMAT) from IQS-URL. UDP-Galactose sodium salt, DOPG (1,2-Dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol)) sodium salt and DOPC (1,2-Dioleoyl-sn-glycero-3-phosphocholine) were obtained from Merck. PMSF (phenylmethylsulfonyl fluoride) and d-desthiobiotin were purchased from Merck. Lyticase from *Arthrobacter luteus* was purchased from Merck. TRIzol™ reagent was purchased from ThermoFisher.

All restriction enzymes used in this project were purchased from New England Biolabs (NEB).

Biochemistry protocols and molecular biology kits

- *DNA transformation*: *E. coli* was transformed following the standard heat shock transformation protocol. *S. cerevisiae* was transformed following the Gietz LiAc/SS carrier DNA/PEG method¹⁰⁷.
- *DNA polymerases*: iProof DNA Polymerase MasterMix 2x (Bio-Rad) and PrimeStarGXL DNA polymerase Premix 2x (TakaraBio) were used in PCR amplification, the second polymerase was only used for pTDH3-SUR2, pTEF2-ISC1 and p2a33 vectors. REDTaq ReadyMix PCR Reaction Mix (Merck) was used for Colony PCR analysis.
- *DNA handling kits*: Plasmid extraction was carried out using High Purity Plasmid Miniprep Kit from NeoBiotech and following the indicated protocol. PCR purification was performed using GeneJet PCR Purification Kit (ThermoFisher) and following the indicated protocol. DNA extraction from agarose gel was carried out using GeneElute Gel Extraction Kit (Merck). Genomic DNA extraction was carried out using Quick Yeast Genomic DNA Extraction Kit (Neo-Biotech) and following the indicated protocol.
- *DNA/Protein quantification*: QUBIT dsDNA HS/BR Assay kit were used for DNA quantification with desktop QUBIT fluorimeter. Protein was measured with Bradford (Merck) or BCA Pierce (ThermoFisher) assay protocols, depending on the use of detergents (i.e. CHAPS) in the buffer.
- *qPCR kits*: RNA extraction was carried out RNeasy Mini Kit (QIAGEN) and following the indicated steps. cDNA was obtained using NZY First-Strand cDNA Synthesis Kit.

In silico analysis.

CAZy database^{115,147} was used to search for glycosyltransferase families in *Bacteroides fragilis*. Protein sequences were obtained from UniProt database. Jalview was used for sequence alignments using MAFFT and Toffee algorithms using Jalview¹²⁹ with a neighbor-joining phylogenetic tree calculated using BLOSUM62 scores as distance metric. Protein homology search was carried out using BLAST tools from NCBI and UNIPROT.

Cloning, protein expression and purification in *E. coli*

Cloning of *B. fragilis* GT sequences into *E. coli* expression vector.

Cloning in pET28a

Protein sequences were obtained from UniProt database and then translated to DNA code optimized for *E. coli*. Synthetic sequences were obtained from GeneArt (ThermoScientific) and cloned using restriction enzymes NcoI and XhoI into pET28a vector to obtain the expression vectors pET28a-HisTag-BF9343_XXXX. Positive transformants were selected on LB plates with 50 µg/mL kanamycin. Successful plasmid constructions were verified by Sanger sequencing.

Cloning in pET22b-Strep-SUMO vector

Sequence BF9343_3149 was later cloned into a pET22b vector containing a N-terminal fused Strep-tag and SUMO protein. Cloning was carried out by Circular Polymerase Extension cloning (CPEC) cloning¹⁴⁸ amplifying the linear vector and the insert containing BF9343_3149 with primers sharing homologous regions, and CPEC reaction to obtain pET22b-SUMO-BF9343_3149-StrepTag. Positive transformants were selected on LB plates with 100 µg/mL ampicillin. Successful plasmid construction was verified by Sanger sequencing.

Expression of GT candidates and α -GalCer_{GT} in *E. coli*.

pET28a-HisTag-BF9343_XXXX

BL21(DE3)Star cells harboring pET28a containing *B. fragilis* GT sequences was cultured in 300 mL LB containing kanamycin at a final concentration of 50 µg/mL, inoculated at an initial OD of 0.05 and incubated at 37°C until OD 1. At this point, protein expression was induced by adding IPTG to a final concentration of 1 mM and cultures were incubated at 25°C for 4h. Cells were then harvested at 5,000 x g for 25 min and washed with 15 mL of NaCl 0.9%. Cell pellets were kept at -20°C until use.

pET22b-SUMO-BF9343_3149-StrepTag

BL21(DE3)Star cells harboring pET22b-SUMO-BF9343_3149-StrepTag were grown in 600 mL LB containing ampicillin at a final concentration of 100 µg/mL, inoculated at initial OD of 0.05 and incubated at 37°C until OD = 1. At this point, protein expression was induced by adding IPTG to a final concentration of 0.1 mM and cultures were incubated at 30°C for 4h. Cells were then harvested at 5,000 x g for 25 min and washed with 15 mL of NaCl 0.9%. Cell pellets were kept at -20°C until use.

Protein purification

HisTrap affinity chromatography

Frozen cell pellets from 300 mL cultures were thawed to room temperature and resuspended in 15 mL of lysis buffer (50 mM Na₂HPO₄, 150 mM NaCl, pH 7.5) supplemented with 1mM PMSF (phenylmethylsulfonyl fluoride). Cells were disrupted by sonication in a Soniprep 150 sonifier at 4°C (10 min, 10 s ON/ 20 s OFF, 50% amplitude). After sonication, soluble fraction was filtered with 0.45 µm filter, and loaded into HisTrap 1 mL column (GE Healthcare) and washed with lysis buffer. The protein was eluted using a gradient from 0 to 50% for 50 min, with lysis buffer with imidazole at 500 mM. Samples were collected and dialyzed with lysis buffer to remove imidazole using dialysis membranes of 10 kDa. Protein was quantified using the Bradford assay, or BCA assay if CHAPS was present in the lysis buffer.

Strep-tag affinity chromatography

Frozen cell pellets from 600 mL cultures were thawed to room temperature and resuspended in 15 mL of lysis buffer (50 mM Na₂HPO₄, 150 mM NaCl, pH 7.5) supplemented with 1mM PMSF (phenylmethylsulfonyl fluoride). Cells were disrupted by sonication in a Soniprep 150 sonifier at 4°C (10 min, 10 s ON/ 20 s OFF, 50% amplitude). The lysate was centrifuged at 25,000 x g for 60 min. The supernatant was recovered and filtered (0.45µm) before purification with a StrepTrap HP 1 mL column (GE Healthcare). The protein was eluted with lysis buffer with 2.5 mM d-desthiobiotin. Eluted fractions were combined and dialyzed using Amicon Ultra-15 10 kDa (Milli-pore) with lysis buffer. The final retained fraction was recovered, and volume was adjusted to 1 mL with lysis buffer. Protein was quantified using the Bradford assay.

Galactosylceramide transferase activity assay

Glycolipid synthase activity from fresh cell-free extracts or purified enzyme was measured with Cer-NBD as acceptor and UDP-Gal (or UDP-Glc) as donor by HPLC with fluorescence detector (HPLC 1200 Agilent with fluorescence detector, excitation wavelength at 458 nm and emission at 530 nm, Nova-pak C18 column, flowrate of 1 mL·min⁻¹, eluent 75% acetonitrile in water) as described in Orive *et al.* 2020⁴⁷. Cer-NBD was solubilized with BSA at an equimolar concentration or using mixed vesicles with DOPC or/and DOPG as explained below. Product (α-GalCer-NBD) was quantified from the chromatograms from the relative areas of substrate and product:

$$[\alpha\text{-GalCer-NBD}] = [\text{Cer-NBD}]_0 \times \text{Area}_{\alpha\text{-GalCer-NBD}} / (\text{Area}_{\text{Cer-NBD}} + \text{Area}_{\alpha\text{-GalCer-NBD}}).$$



Ceramide-NBD with BSA

Reactions were performed with 25 μM Cer-NBD, 25 μM BSA, 1.25 mM UDP-Gal, phosphate buffer (50 mM Na_2HPO_4 150 mM NaCl, pH 7.5) and the appropriate amount of enzyme in 200 μl final reaction volume. It was preincubated at 37°C for five minutes and the reaction was started by adding UDP-Gal. Sampling time was adjusted depending on enzyme concentration. Aliquots were withdrawn and mixed with methanol (MeOH:sample 8:2) and centrifuged to eliminate debris. Samples were then analyzed by HPLC. Chromatographic peaks were assigned by co-injection with independent standards. Initial rates were obtained from the linear progress curve of product formation. Initial rates were determined from triplicate assays.

Kinetic parameters


UDP-Gal (0.1-2.5 mM), ceramide-NBD (5-300 μM) with equimolar amount of BSA were individually varied maintaining constant the other components. Initial rates vs. donor or acceptor concentrations were fitted to a Michaelis-Menten equation with substrate inhibition by non-linear regression using Prism 8 software (GraphPad) from which the kinetic parameters were derived.

Effect of metal cations

Tris buffer was used when analyzing metal binding to avoid precipitation of di-valent cations. Cell pellet was resuspended with 20 mM Tris buffer with 200 mM NaCl, 1 mM EDTA 1 mM DTT pH 7.5 supplemented with 1 mM PMSF. Enzyme purification was carried out following the same protocol as before but with lysis buffer 20 mM Tris buffer with 200 mM NaCl pH 7.5. Reactions were performed with 25 μM Cer-NBD, 25 μM BSA, 1.25 mM UDP-Gal, the appropriate amount of enzyme, Tris buffer (20 mM Tris 150 mM NaCl, pH 7.5) with or without EDTA (200 mM) and addition of metal ions (0.25-1 mM) in 200 μL final reaction volume. It was preincubated at 37°C for five minutes and the reaction was started by adding UDP-Gal.

Ceramide-NBD in vesicles

Ceramide-NBD, DOPG and/or DOPG (matrix lipid) were dissolved in chloroform in a glass vial using the desired amounts for each experiment as indicated. Chloroform was evaporated under a N_2 stream, and the lipid mixture dried under vacuum for 2 h. When reactions were performed, the lipid film was solubilized to homogeneity in 80 μL of phosphate buffer solution (50 mM Na_2HPO_4 150 mM NaCl, pH 7.5) by extensive vortexing and ultrasound bath for 5 min and kept on ice for 1 min. This cycle was repeated 8 times. Vesicle formation was analyzed using ZetaSizer (Malvern Panalytical). For enzymatic reaction, 25 μL of enzyme were added to 60 μL of freshly prepared vesicles solution. The mixture was kept for 30 min on ice and preincubated for 5 min at 37 °C. Reaction was started by adding 12.5 μL UDP-Gal and incubated at 37°C. Reaction conditions were: 1.25 mM UDP-Gal, 25 μM Cer-NBD, DOPC/DOPG vesicles (condition 1: 1 mM DOPC; 2: 1 mM total DOPC and DOPG (40 mol % DOPG); 3: 1



mM DOPG, phosphate buffer (50 mM Na₂HPO₄ 150 mM NaCl, pH 7.5), and enzyme (0.92-2.75 μM) in 200 μL reaction volume. Aliquots were withdrawn every 30 min for 90 min, stopped as before and analyzed by HPLC.

Modeled 3D structure and ligand docking.

The AlfaFold model of *B. fragilis* α-GalCer_GT was obtained from Uniprot A0A380YRQ3. UDP-Gal structure was taken from PDB 5M7D. The complex α-GalCer_GT·UDP-Gal was generated by docking with AUTODOCK VINA¹⁴⁹. Both the protein and ligand structure were first parametrized with AutoDockTools¹⁵⁰: polar hydrogens were added, Auto-Dock4.2 atom typing was used, and Gaisteger partial charges were computed. All rotatable bonds of the ligands were considered free during the docking calculations, whereas the whole protein structure was kept fixed. The docking search space was confined in a box centered in the active site. Exhaustiveness level was set to 24, and 20 binding modes of the ligands were generated. Only low energy binding poses were considered for analysis. Pictures of the complexes were generated with VMD¹⁵¹.

Yeast adapted RNA extraction protocol for qPCR analysis

RNA extraction was carried out by culturing cells with the appropriate selection media and then inoculating cultures with YPUAD media. Cells were incubated at 30°C in agitation until exponential phase (OD was measured continuously until reached a value around 1), taking an appropriate volume for a total of 10⁸ cells, and treating the cells with lyticase (1 mg/mL) for 30 min, and then adding 1 mL of TRIzol to the mixture. After treatment, the mixture was then used to follow the RNeasy Mini Kit extraction protocol. The thermocycler used for qPCR analysis QuantStudio™ 3 Real-Time PCR System, 96-well, 0.2 mL.

Yeast Bf3149 expression analysis

In vitro galactosylceramide transferase activity assay

Cell pellets were then resuspended in 9.5 mL lysis buffer (50 mM Na₂HPO₄ 150 mM NaCl pH 7.5 with 1 mM PMSF). Finally, samples were disrupted by addition of glass beads and vortexed for 5 min. When needed, OD was checked, and sample were vortexed again. Glass beads were removed after cell disruption. *E. coli* cells were disrupted by sonication (10s ON 20s OFF, ON total time 5 min at 50% amplitude).

From each yeast culture, two reactions were prepared with different protein source: one with suspension after disruption (containing both soluble and insoluble parts) and one with cell-free extract

(soluble part after centrifugation at 12,000 rpm for 30 min). Each reaction mix contained: 100 μ L of protein source (concentration was not measured), 25 μ M ceramide-NBD solubilized in equimolar amount with BSA, phosphate buffer (50 mM Na_2HPO_4 150 mM NaCl pH 7.5) and 1.25 mM UDP-galactose (added to start reaction). Samples were taken at 30, 60 and 120 minutes by mixing 20 μ L of reaction with 80 μ L of methanol to stop reaction.

In vivo galactosylceramide transferase activity assay

For the *in vivo* assay, Levine *et al.*¹³⁹ protocol was adapted, the same strains were tested, and 5 mL of cultures were inoculated from a 24h precultures in the same selection media. Starting OD was 0.1-0.15 and cells were incubated at 30°C in agitation until OD 1. At this point, cells were centrifuged at low speed to avoid cell disruption. Cell pellet was then resuspended with 750 μ L of selective medium, and 250 μ L of ceramide-NBD solubilized in BSA was added to a final concentration of 25 μ M of ceramide and 5 mg/mL of BSA. These cultures with Cer-NBD were then incubated at 30°C for another 30 min in agitation. Finally, 500 μ L samples were taken at 15 and 30 minutes and centrifuged at low speed. Cells were then washed with chilled phosphate buffer (50 mM Na_2HPO_4 150 mM NaCl pH 7.5) and centrifuged again. Pellet was then resuspended in 500 μ L of a mix of chloroform:methanol:1M HCl (5:12:4), 2 spoonful of glass beads were added and samples were vortexed for 5 min, then centrifuged and transferred to new fresh tubes. Next, 125 μ L of chloroform was added into the mix and centrifuged again. finally, the lower phase (chloroform) was recovered, added into fresh tubes, and dried under N_2 stream. For HPLC analysis, lipid film was resuspended in 100 μ L of methanol.

In vivo staining of *S. cerevisiae* with C6-Ceramide-NBD and DAPI for confocal microscopy

Yeast cells were incubated in the appropriate selection media (SD -X, where X is the selection marker amino acid) for 24h. This culture was used to inoculate a fresh culture, which was incubated at 30°C in agitation from OD 0.2 to 1. Then, cultures are centrifuged at 3,000 $\times g$ for 20 min. Cell pellet is resuspended in 0.75 mL of SD complete media (SD media containing all amino acids). 0.25 mL of this BSA solution at 100 μ M are added to the 0.75 mL of resuspended cells. Cell solution with Cer-NBD is incubated at 30°C for 30 min in agitation. Then cells are centrifuged again at 3,000 $\times g$ for 20 min, and cell pellet is resuspended in 1 mL of chilled phosphate buffer (50 mM Na_2HPO_4 150 mM NaCl, pH 7.5). Then, cells are resuspended with a solution 1/1000 of DAPI (from a stock solution of 10 mg/mL in phosphate buffer), incubated for 10 min at 30°C, and then washed with chilled phosphate buffer and resuspending final pellet with 500 μ L of chilled phosphate buffer. Cells were then stored in the fridge until microscopy analysis.

Cells were observed under confocal microscope (Leica Laser Scanning Confocal Microscope TCS SP8), observing each strain under the microscope at 63x the appropriate conditions for fluorescence detection of mCherry (excitation/emission 587/610 nm), ceramide-NBD (ex/em. 466/536 nm) and DAPI (ex/em. 358/461 nm), as well as optical mode to see cell outline.

Material and methods



Bibliography



Bibliography



Bibliography



Bibliography

1. Gory Lecollinet G-, Gulik A, Mackenzie G, Goodby JW, Benvegna T, Plusquellec D. *Supramolecular Self-Assembling Properties of Membrane-Spanning Archaeal Tetraether Glycolipid Analogues*.
2. Faivre V, Rosilio V. Interest of glycolipids in drug delivery: from physicochemical properties to drug targeting. *Expert Opin Drug Deliv*. 2010;7(9):1031-1048. doi:10.1517/17425247.2010.511172
3. Abdel-Mawgoud AM, Stephanopoulos G. Simple glycolipids of microbes: Chemistry, biological activity and metabolic engineering. *Synth Syst Biotechnol*. 2018;3(1):3-19. doi:10.1016/j.synbio.2017.12.001
4. Leray C. Complex glycolipids. In: *Introduction to Lipidomics*. CRC Press; 2012:247e94.
5. Leray C. Simple lipids with two different components. In: *Introduction to Lipidomics*. CRC Press; 2012:169e210.
6. Leray Claude. *Introduction to Lipidomics : From Bacteria to Man*. CRC Press; 2013. Accessed September 3, 2022. <https://www.perlego.com/book/1604012/introduction-to-lipidomics-from-bacteria-to-man-pdf>
7. Hölzl G, Dörmann P. Structure and function of glycoglycerolipids in plants and bacteria. *Prog Lipid Res*. 2007;46(5):225-243. doi:10.1016/j.plipres.2007.05.001
8. Bagnat M, Simons K. Lipid Rafts in Yeast Protein Sorting. *Biol Chem*. 2002;383:1475-1480.
9. Mollinedo F. Lipid raft involvement in yeast cell growth and death. *Front Oncol*. 2012;2. doi:10.3389/fonc.2012.00140
10. Douglas LM, Konopka JB. Fungal Membrane Organization: The Eisosome Concept. *Annu Rev Microbiol*. 2014;68(1):377-393. doi:10.1146/annurev-micro-091313-103507
11. Hölzl G, Dörmann P. Structure and function of glycoglycerolipids in plants and bacteria. *Prog Lipid Res*. 2007;46(5):225-243. doi:10.1016/j.plipres.2007.05.001
12. Banat IM, Makkar RS, Cameotra SS. *Potential Commercial Applications of Microbial Surfactants*.
13. Chong H, Li Q. Microbial production of rhamnolipids: Opportunities, challenges and strategies. *Microb Cell Fact*. 2017;16(1):1-12. doi:10.1186/s12934-017-0753-2
14. Bahia FM, de Almeida GC, de Andrade LP, et al. Rhamnolipids production from sucrose by engineered *Saccharomyces cerevisiae*. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-018-21230-2
15. Corti M, Cantù L, Brocca P, del Favero E. Self-assembly in glycolipids. *Curr Opin Colloid Interface Sci*. 2007;12(3):148-154. doi:10.1016/j.cocis.2007.05.002
16. He H, Lu Y, Qi J, Zhu Q, Chen Z, Wu W. Adapting liposomes for oral drug delivery. *Acta Pharm Sin B*. 2019;9(1):36-48. doi:https://doi.org/10.1016/j.apsb.2018.06.005
17. Muthusamy KK, Gopalakrishnan S, Ravi T, Sivachidambaram P. Biosurfactants : Properties, commercial production and application. *Curr Sci*. 2008;94:736-747.

18. Liu Z, Guo J. NKT-cell glycolipid agonist as adjuvant in synthetic vaccine. *Carbohydr Res.* 2017;452:78-90. doi:<https://doi.org/10.1016/j.carres.2017.10.006>
19. Ghosh B, Lai YH, Shih YY, Pradhan TK, Lin CH, Mong KKT. Total Synthesis of a Glycoglycerolipid from *Meiothermus taiwanensis* through a One-Pot Glycosylation Reaction and Exploration of its Immunological Properties. *Chem Asian J.* 2013;8(12):3191-3199. doi:<https://doi.org/10.1002/asia.201300933>
20. Schnaar RL, Sandhoff R, Tiemeyer M, Kinoshita. T. Chapter 11 Glycosphingolipids. In: Varki A, Cummings R, Esko J, et al., eds. *Essentials of Glycobiology*. 4th Editio. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
21. Zhang T, de Waard AA, Wuhrer M, Spaapen RM. The Role of Glycosphingolipids in Immune Cell Functions. *Front Immunol.* 2019;10(January):1-22. doi:10.3389/fimmu.2019.00090
22. Ichikawa S, Sakiyama H, Suzuki G, Hidari KIPJ, Hirabayashi Y. Expression cloning of a cDNA for human ceramide glucosyltransferase that catalyzes the first glycosylation step of glycosphingolipid synthesis. *Proc Natl Acad Sci.* 1996;93:4638-4643.
23. Sorge J, West C, Westwood B, Beutler E. Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. *Proc Natl Acad Sci.* 1985;82(November):7289-7293.
24. Shayman JA. Targeting Glucosylceramide Synthesis in the Treatment of Rare and Common Renal Disease. *Semin Nephrol.* 2018;38(2):183-192. doi:10.1016/j.semnephrol.2018.01.007.Targeting
25. Bendelac A, Savage PB, Teyton L. The Biology of NKT Cells. *Annu Rev Immunol.* 2007;25:297-336. doi:10.1146/annurev.immunol.25.022106.141711
26. Natori T, Koezuka Y, Higa T. Agelasphins, novel α -galactosylceramides from the marine sponge *Agelas mauritanus*. *Tetrahedron Lett.* 1993;34(35):5591-5592. doi:[https://doi.org/10.1016/S0040-4039\(00\)73889-5](https://doi.org/10.1016/S0040-4039(00)73889-5)
27. Natori T, Morita M, Akimoto K, Koezuka Y. Agelasphins, novel antitumor and immunostimulatory cerebrosides from the marine sponge *Agelas mauritanus*. *Tetrahedron.* 1994;50(9):2771-2784. doi:[https://doi.org/10.1016/S0040-4020\(01\)86991-X](https://doi.org/10.1016/S0040-4020(01)86991-X)
28. Morita M, Motoki K, Akimoto K, et al. Structure-Activity Relationship of α -Galactosylceramides against B16-Bearing Mice. *J Med Chem.* 1995;38(12):2176-2187. doi:10.1021/jm00012a018
29. Kohlgruber AC, Donado CA, LaMarche NM, Brenner MB, Brennan PJ. Activation strategies for invariant natural killer T cells. *Immunogenetics.* 2016;68(8):649-663. doi:10.1007/s00251-016-0944-8
30. McEwen-Smith RM, Salio M, Cerundolo V. CD1d-dependent endogenous and exogenous lipid antigen presentation. *Curr Opin Immunol.* 2015;34:116-125. doi:10.1016/j.coi.2015.03.004
31. Birkholz AM, Kronenberg M. Antigen specificity of invariant natural killer T-cells. *Biomed J.* 2015;38(6):470-483. doi:10.1016/j.bj.2016.01.003

32. Carreño LJ, Saavedra-Ávila NA, Porcelli SA. Synthetic glycolipid activators of natural killer T cells as immunotherapeutic agents. *Clin Transl Immunology*. 2016;5(4):e69. doi:10.1038/cti.2016.14
33. Artiaga BL, Yang G, Hackmann TJ, et al. α -Galactosylceramide protects swine against influenza infection when administered as a vaccine adjuvant. *Sci Rep*. 2016;6(January):23593. doi:10.1038/srep23593
34. Pérez K, Osorio M, Hernández J, Carr A, Fernández LE. NGcGM3/VSSP vaccine as treatment for melanoma patients. *Hum Vaccin Immunother*. 2013;9(6):1237-1240. doi:10.4161/hv.24115
35. Zhang L, Donda A. Alpha-galactosylceramide/CD1d-antibody fusion proteins redirect invariant natural killer T cell immunity to solid tumors and promote prolonged therapeutic responses. *Front Immunol*. 2017;8(NOV):4-6. doi:10.3389/fimmu.2017.01417
36. Macho-Fernandez E, Chekkat N, Ehret C, et al. Solubilization of α -galactosylceramide in aqueous medium: Impact on Natural Killer T cell activation and antitumor responses. *Int J Pharm*. 2017;530(1-2):354-363. doi:10.1016/j.ijpharm.2017.07.054
37. Wieland Brown LC, Penaranda C, Kashyap PC, et al. Production of α -Galactosylceramide by a Prominent Member of the Human Gut Microbiota. *PLoS Biol*. 2013;11(7). doi:10.1371/journal.pbio.1001610
38. Okino N, Li M, Qu Q, et al. Two bacterial glycosphingolipid synthases responsible for the synthesis of glucuronosylceramide and α -galactosylceramide. *Journal of Biological Chemistry*. 2020;295(31):10709-10725. doi:10.1074/jbc.RA120.013796
39. Yunoki K, Sato M, Seki K, Ohkubo T, Tanaka Y, Ohnishi M. Simultaneous Quantification of Plant Glyceroglycolipids Including Sulfoquinovosyldiacylglycerol by HPLC–ELSD with Binary Gradient Elution. *Lipids*. 2009;44(1):77-83. doi:https://doi.org/10.1007/s11745-008-3248-4
40. Anastas PT, Kirchhoff MM. Origins, Current Status, and Future Challenges of Green Chemistry. *Acc Chem Res*. 2002;35(9):686-694. doi:10.1021/ar010065m
41. Sinnott ML. *Catalytic Mechanisms of Enzymic Glycosyl Transfer*. Vol 90.; 1990.
42. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyl transferases: Structures, functions, and mechanisms. *Annu Rev Biochem*. 2008;77:521-555. doi:10.1146/annurev.biochem.76.061005.092322
43. Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG. Mechanistic evidence for a front-side, S_Ni-type reaction in a retaining glycosyltransferase. *Nat Chem Biol*. 2011;7(9):631-638. doi:10.1038/nchembio.628
44. McArthur JB, Chen X. Glycosyltransferase engineering for carbohydrate synthesis. *Biochem Soc Trans*. 2016;44:129-142. doi:10.1042/BST20150200
45. Hancock SM, Vaughan MD, Withers SG. Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol*. 2006;10(5):509-519. doi:10.1016/j.cbpa.2006.07.015
46. Mackenzie LF, Wang Q, Antony Warren RJ, Withers SG. *Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis*. Vol 26.; 1998.

47. Orive-milla N, Delmulle T, Mey M de, Fajjes M, Planas A. Metabolic engineering for glycolipid production in *E. coli*: Tuning phosphatidic acid and UDP-glucose pathways. *Metab Eng.* 2020;61(December 2019):106-119. doi:10.1016/j.ymben.2020.05.010
48. Rodríguez-Villalón A, Pérez-Gil J, Rodríguez-Concepción M. Carotenoid accumulation in bacteria with enhanced supply of isoprenoid precursors by upregulation of exogenous or endogenous pathways. *J Biotechnol.* 2008;135(1):78-84. doi:10.1016/j.jbiotec.2008.02.023
49. Erb TJ, Jones PR, Bar-Even A. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr Opin Chem Biol.* 2017;37:56-62. doi:10.1016/j.cbpa.2016.12.023
50. Shiue E, Prather KLJ. Improving d-glucaric acid production from myo-inositol in *E. coli* by increasing MIOX stability and myo-inositol transport. *Metab Eng.* 2014;22:22-31. doi:10.1016/j.ymben.2013.12.002
51. Gallego-Jara J, de Diego T, del Real Á, Écija-Conesa A, Manjón A, Cánovas M. Lycopene overproduction and in situ extraction in organic-aqueous culture systems using a metabolically engineered *Escherichia coli*. *AMB Express.* 2015;5(1). doi:10.1186/s13568-015-0150-3
52. Xie W, Liu M, Lv X, Lu W, Gu J, Yu H. Construction of a controllable β -carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol Bioeng.* 2014;111(1):125-133. doi:10.1002/bit.25002
53. Li XR, Tian GQ, Shen HJ, Liu JZ. Metabolic engineering of *Escherichia coli* to produce zeaxanthin. *J Ind Microbiol Biotechnol.* 2015;42(4):627-636. doi:10.1007/s10295-014-1565-6
54. Park SY, Binkley RM, Kim WJ, Lee MH, Lee SY. Metabolic engineering of *Escherichia coli* for high-level astaxanthin production with high productivity. *Metab Eng.* 2018;49:105-115. doi:10.1016/j.ymben.2018.08.002
55. Weissman KJ. Genetic engineering of modular PKSs: From combinatorial biosynthesis to synthetic biology. *Nat Prod Rep.* 2016;33(2):203-230. doi:10.1039/c5np00109a
56. Wittgens A, Santiago-Schuebel B, Henkel M, et al. Heterologous production of long-chain rhamnolipids from *Burkholderia glumae* in *Pseudomonas putida*—a step forward to tailor-made rhamnolipids. *Appl Microbiol Biotechnol.* 2018;102(3):1229-1239. doi:10.1007/s00253-017-8702-x
57. Jezierska S, Claus S, Ledesma-Amaro R, van Bogaert I. Redirecting the lipid metabolism of the yeast *Starmerella bombicola* from glycolipid to fatty acid production. *J Ind Microbiol Biotechnol.* 2019;46(12):1697-1706. doi:10.1007/s10295-019-02234-x
58. Cao X, Lv YB, Chen J, Imanaka T, Wei LJ, Hua Q. Biotechnology for Biofuels Metabolic engineering of oleaginous yeast *Yarrowia lipolytica* for limonene overproduction. *Biotechnol Biofuels.* Published online 2016:1-11. doi:10.1186/s13068-016-0626-7
59. Niehus X, Crutz-Le Coq AM, Sandoval G, Nicaud JM, Ledesma-Amaro R. Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. *Biotechnol Biofuels.* 2018;11(1):1-10. doi:10.1186/s13068-018-1010-6

Bibliography

60. Bajhaiya AK, Ziehe Moreira J, Pittman JK. Transcriptional Engineering of Microalgae: Prospects for High-Value Chemicals. *Trends Biotechnol.* 2017;35(2):95-99. doi:10.1016/j.tibtech.2016.06.001
61. Niu FX, Lu Q, Bu YF, Liu JZ. Metabolic engineering for the microbial production of isoprenoids: Carotenoids and isoprenoid-based biofuels. *Synth Syst Biotechnol.* 2017;2(3):167-175. doi:10.1016/j.synbio.2017.08.001
62. Shin KS, Kim S, Lee SK. Improvement of free fatty acid production using a mutant acyl-CoA thioesterase I with high specific activity in *Escherichia coli*. *Biotechnol Biofuels.* 2016;9(1):1-10. doi:10.1186/s13068-016-0622-y
63. Runguphan W, Keasling JD. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng.* 2014;21:103-113. doi:10.1016/j.ymben.2013.07.003
64. Ogawa T, Tamoi M, Kimura A, et al. Enhancement of photosynthetic capacity in *Euglena gracilis* by expression of cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase leads to increases in biomass and wax ester production. *Biotechnol Biofuels.* 2015;8(1). doi:10.1186/s13068-015-0264-5
65. Phale PS, Mohapatra B, Malhotra H, Shah BA. Eco-physiological portrait of a novel *Pseudomonas* sp. CSV86: an ideal host/candidate for metabolic engineering and bioremediation. *Environ Microbiol.* 2022;24(6):2797-2816. doi:10.1111/1462-2920.15694
66. Mangwani N, Shukla SK, Rao TS, Das S. Calcium-mediated modulation of *Pseudomonas mendocina* NR802 biofilm influences the phenanthrene degradation. *Colloids Surf B Biointerfaces.* 2014;114:301-309. doi:10.1016/j.colsurfb.2013.10.003
67. Liang H, Hu Z, Zhang T, et al. Production of a bioactive unnatural ginsenoside by metabolically engineered yeasts based on a new UDP-glycosyltransferase from *Bacillus subtilis*. *Metab Eng.* 2017;44(April):60-69. doi:10.1016/j.ymben.2017.07.008
68. Haro MA, de Lorenzo V. *Metabolic Engineering of Bacteria for Environmental Applications: Construction of Pseudomonas Strains for Biodegradation of 2-Chlorotoluene*. Vol 85.; 2001. www.elsevier.com/locate/jbiotec
69. Zelcbuch L, Lindner SN, Zegman Y, et al. Pyruvate formate-lyase enables efficient growth of *Escherichia coli* on acetate and formate. *Biochemistry.* 2016;55(17):2423-2426. doi:10.1021/acs.biochem.6b00184
70. Sarma H, Prasad MNV. Metabolic Engineering of Rhizobacteria Associated With Plants for Remediation of Toxic Metals and Metalloids. In: *Transgenic Plant Technology for Remediation of Toxic Metals and Metalloids*. Elsevier; 2018:299-318. doi:10.1016/B978-0-12-814389-6.00014-6
71. Dangi AK, Sharma B, Hill RT, Shukla P. Bioremediation through microbes: systems biology and metabolic engineering approach. *Crit Rev Biotechnol.* 2019;39(1):79-98. doi:10.1080/07388551.2018.1500997

72. van Meer G, Voelker D, Feigenson GW. Membrane lipids: where they are and how they behave. *Nature Reviews*. 2008;9. doi:10.1038/nrm2330TriacylglycerolA
73. Moseley JB. Eisosomes. *Current Biology*. 2018;28(8):R376-R378. doi:10.1016/j.cub.2017.11.073
74. Epstein S, Castillon GA, Qin Y, Riezman H. An essential function of sphingolipids in yeast cell division. *Mol Microbiol*. 2012;84(6):1018-1032. doi:10.1111/j.1365-2958.2012.08087.x
75. Morimoto Y, Tani M. Synthesis of mannosylinositol phosphorylceramides is involved in maintenance of cell integrity of yeast *Saccharomyces cerevisiae*. *Mol Microbiol*. 2015;95(4):706-722. doi:10.1111/mmi.12896
76. Ren Jihui and Hannun YA. Metabolism and Roles of Sphingolipids in Yeast *Saccharomyces cerevisiae*. In: Geiger O, ed. *Biogenesis of Fatty Acids, Lipids and Membranes*. Springer International Publishing; 2017:1-21. doi:10.1007/978-3-319-43676-0_21-1
77. Malhotra R, Singh B. Glycolipids of *Saccharomyces cerevisiae* Cell. *Journal of Biological Sciences*. 2005;5(3):253-256.
78. Johnson B, Nelson SJ, Brown CM. *Influence of Glucose Concentration on the Physiology and Lipid Composition of Some Yeasts*. Vol 38.; 1972.
79. Lian J, Zhao H. Reversal of the β -oxidation cycle in *Saccharomyces cerevisiae* for production of fuels and chemicals. *ACS Synth Biol*. 2015;4(3):332-341. doi:10.1021/sb500243c
80. Zhou YJ, Buijs NA, Zhu Z, Qin J, Siewers V, Nielsen J. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat Commun*. 2016;7(May):11709. doi:10.1038/ncomms11709
81. Vickers CE, Williams TC, Peng B, Cherry J. Recent advances in synthetic biology for engineering isoprenoid production in yeast. *Curr Opin Chem Biol*. 2017;40(Figure 1):47-56. doi:10.1016/j.cbpa.2017.05.017
82. Arendt P, Miettinen K, Pollier J, de Rycke R, Callewaert N, Goossens A. An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. *Metab Eng*. 2017;40(January):165-175. doi:10.1016/j.ymben.2017.02.007
83. Suástegui M, Shao Z. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. *J Ind Microbiol Biotechnol*. 2016;43(11):1611-1624. doi:10.1007/s10295-016-1824-9
84. Andrés E, Martínez N, Planas A. Expression and Characterization of a *Mycoplasma genitalium* Glycosyltransferase in Membrane Glycolipid Biosynthesis. *The Journal of Biological Chemistry*. 2012;286(41):35367-35379.
85. Mora-buye N, Faijes M, Planas A. An engineered *E. coli* strain for the production of glycolipids. *Metab Eng*. 2012;14:551-559.
86. Dickson RC, Sumanasekera C, Lester RL. Functions and metabolism of sphingolipids in *Saccharomyces cerevisiae*. *Prog Lipid Res*. 2006;45(6):447-465. doi:10.1016/j.plipres.2006.03.004

87. Tani M. Structure–Function Relationship of Complex Sphingolipids in Yeast. *Trends in Glycoscience and Glycotechnology*. 2016;28(164):E109-E116. doi:10.4052/tigg.1509.1E
88. Murakami S, Shimamoto T, Nagano H, Tsuruno M. Producing human ceramide-NS by metabolic engineering using yeast *Saccharomyces cerevisiae*. *Nature Publishing Group*. Published online 2015:1-11. doi:10.1038/srep16319
89. Kim SK, Noh YH, Koo JR, Yun HS. Effect of expression of genes in the sphingolipid synthesis pathway on the biosynthesis of ceramide in *Saccharomyces cerevisiae*. *J Microbiol Biotechnol*. 2010;20(2):356-362. doi:10.4014/jmb.0909.09041
90. Voynova NS, Vionnet C, Ejsing CS, Conzelmann A. A novel pathway of ceramide metabolism in *Saccharomyces cerevisiae*. *Biochem J*. 2012;114:103-114. doi:10.1042/BJ20120712
91. Hama H, Young DA, RAdding JA, et al. Requirement of sphingolipid alpha-hydroxylation for fungicidal action of syringomycin E. *FEBS Lett*. 2000;478:26-28.
92. Herrero AB, Astudillo AM, Balboa MA, Cuevas C, Balsinde J, Moreno S. Levels of SCS7/FA2H-mediated fatty acid 2-hydroxylation determine the sensitivity of cells to antitumor PM02734. *Cancer Res*. 2008;68(23):9779-9787. doi:10.1158/0008-5472.CAN-08-1981
93. Schorling S, Barz WP, Riezman H, Oesterhelt D. Lag1p and Lac1p Are Essential for the Acyl-CoA – dependent Ceramide Synthase Reaction in *Saccharomyces cerevisiae*. *Mol Biol Cell*. 2001;12:3417-3427.
94. Nagiec MM, Nagiec EE, Baltisberger JA, Wells GB, Lester RL, Dickson RC. Sphingolipid synthesis as a target for antifungal drugs. Complementation of the inositol phosphorylceramide synthase defect in a mutant strain of *Saccharomyces cerevisiae* by the AUR1 gene. *Journal of Biological Chemistry*. 1997;272(15):9809-9817. doi:10.1074/jbc.272.15.9809
95. Cerantola V, Guillas I, Roubaty C, et al. Aureobasidin A arrests growth of yeast cells through both ceramide intoxication and deprivation of essential inositolphosphorylceramides. *Mol Microbiol*. 2009;71(6):1523-1537. doi:10.1111/j.1365-2958.2009.06628.x
96. Megyeri M, Riezman H, Schuldiner M, Futerman AH. *Making Sense of the Yeast Sphingolipid Pathway*. Vol 428. Elsevier B.V.; 2016. doi:10.1016/j.jmb.2016.09.010
97. Tani M, Toume M. Alteration of complex sphingolipid composition and its physiological significance in yeast *Saccharomyces cerevisiae* lacking vacuolar ATPase. *Microbiology (United Kingdom)*. 2015;161(12):2369-2383. doi:10.1099/mic.0.000187
98. Oh CS, Toke DA, Mandala S, Martin CE. ELO2 and ELO3, homologues of the *Saccharomyces cerevisiae* ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation. *Journal of Biological Chemistry*. 1997;272(28):17376-17384. doi:10.1074/jbc.272.28.17376
99. Yu T, Zhou YJ, Wenning L, et al. Metabolic engineering of *Saccharomyces cerevisiae* for production of very long chain fatty acid-derived chemicals. *Nat Commun*. 2017;8(May):15587. doi:10.1038/ncomms15587

Bibliography

100. Kajiwara K, Watanabe R, Pichler H, et al. Yeast ARV1 Is Required for Efficient Delivery of an Early GPI Intermediate to the First Mannosyltransferase during GPI Assembly and Controls Lipid Flow from the Endoplasmic Reticulum. *Mol Biol Cell*. 2008;19:2069-2082. doi:10.1091/mbc.E07-08
101. Kajiwara K, Muneoka T, Watanabe Y, Karashima T, Kitagaki H, Funato K. Perturbation of sphingolipid metabolism induces endoplasmic reticulum stress-mediated mitochondrial apoptosis in budding yeast. *Mol Microbiol*. 2012;86(5):1246-1261. doi:10.1111/mmi.12056
102. Karim AS, Curran KA, Alper HS. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res*. 2013;13(1):107-116. doi:10.1111/1567-1364.12016
103. Curran KA, Karim AS, Gupta A, Alper HS. Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng*. 2013;19:88-97. doi:10.1016/j.ymben.2013.07.001
104. Curran KA, Morse NJ, Markham KA, Wagman AM, Gupta A, Alper HS. Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth Biol*. 2015;4(7):824-832. doi:10.1021/sb5003357
105. Redden H, Morse N, Alper HS. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Res*. 2015;15(1):1-10. doi:10.1111/1567-1364.12188
106. Cankorur-Cetinkaya A, Dereli E, Eraslan S, Karabekmez E, Dikicioglu D, Kirdar B. A novel strategy for selection and validation of reference genes in dynamic multidimensional experimental design in yeast. *PLoS One*. 2012;7(6). doi:10.1371/journal.pone.0038351
107. Gietz RD, Schiestl RH. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*. 2007;2(1):38-41. doi:10.1038/nprot.2007.15
108. von Gerichten J, Lamprecht D, Opálka L, et al. Bacterial immunogenic α -galactosylceramide identified in the murine large intestine: Dependency on diet and inflammation. *J Lipid Res*. 2019;60(11):1892-1904. doi:10.1194/jlr.RA119000236
109. Burdin N, Brossay L, Koezuka Y, et al. Selective Ability of Mouse CD1 to Present Glycolipids: α -Galactosylceramide Specifically Stimulates $V\alpha 14+$ NK T Lymphocytes. *The Journal of Immunology*. 1998;161(7):3271-3281. <https://www.jimmunol.org/content/161/7/3271>
110. Shimizu K, Kurosawa Y, Taniguchi M, Steinman RM, Fujii SI. Cross-presentation of glycolipid from tumor cells loaded with α -galactosylceramide leads to potent and long-lived T cell-mediated immunity via dendritic cells. *Journal of Experimental Medicine*. 2007;204(11):2641-2653. doi:10.1084/jem.20070458
111. Mattarollo SR, West AC, Steegh K, et al. NKT cell adjuvant-based tumor vaccine for treatment of myc oncogene-driven mouse B-cell lymphoma. *Blood*. 2012;120(15):3019-3029. doi:10.1182/blood-2012-04-426643

112. Gableh F, Saeidi M, Hemati S, et al. Combination of the toll like receptor agonist and α -Galactosylceramide as an efficient adjuvant for cancer vaccine. *J Biomed Sci.* 2016;23(1):16. doi:10.1186/s12929-016-0238-3
113. Plettenburg O, Bodmer-Narkevitch V, Wong CH. Synthesis of α -Galactosyl Ceramide, a Potent Immunostimulatory Agent. *J Org Chem.* 2002;67(13):4559-4564. doi:10.1021/jo0201530
114. Janssens J, Decruy T, Venken K, et al. Efficient Divergent Synthesis of New Immunostimulant 4"-Modified α -Galactosylceramide Analogues. *ACS Med Chem Lett.* 2017;8(6):642-647. doi:10.1021/acsmchemlett.7b00107
115. Drula E, Garron M line, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database : functions and literature. 2022;50(November 2021):571-577.
116. Karlsson OP, Dahlqvist A, Vikström S, Wieslander Å. Lipid dependence and basic kinetics of the purified 1,2-diacylglycerol 3-glucosyltransferase from membranes of *Acholeplasma laidlawii*. *Journal of Biological Chemistry.* 1997;272(2):929-936. doi:10.1074/jbc.272.2.929
117. Berg S, Edman M, Li L, Wikstro M, Wieslander Å. Sequence Properties of the 1 , 2-Diacylglycerol 3-Glucosyltransferase from *Acholeplasma laidlawii* Membranes Sequence Features of MGlcDAG Synthase. *The Journal of Biological Chemistry.* 2001;276(25):22056-22063. doi:10.1074/jbc.M102576200
118. Edman M, Berg S, Storm P, et al. Structural features of glycosyltransferases synthesizing major bilayer and nonbilayer-prone membrane lipids in *Acholeplasma laidlawii* and *Streptococcus pneumoniae*. *Journal of Biological Chemistry.* 2003;278(10):8420-8428. doi:10.1074/jbc.M211492200
119. Sonnhammer ELL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. In: *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology.* ; 1998:175-182.
120. Krogh A, Larsson È, Heijne G von, Sonnhammer ELL. Predicting Transmembrane Protein Topology with a Hidden Markov Model : Application to Complete Genomes. *J Mol Biol.* 2001;305:567-580. doi:10.1006/jmbi.2000.4315
121. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420-423. doi:10.1038/s41587-019-0036-z
122. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. *Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia Coli.* Vol 64.; 1998. <https://journals.asm.org/journal/aem>
123. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Stephen D, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics.* 2004;5:75-86.

124. Thiyagarajan N, Pham TTK, Stinson B, et al. Structure of a metal-independent bacterial glycosyltransferase that catalyzes the synthesis of histo-blood group A antigen. *Sci Rep.* 2012;2. doi:10.1038/srep00940
125. Albesa-Jové D, Giganti D, Jackson M, Alzari PM, Guerin ME. Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology.* 2014;24(2):108-124. doi:10.1093/glycob/cwt101
126. Guerin ME, Kordulakova J, Schaeffer F, et al. Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from mycobacteria. *Journal of Biological Chemistry.* 2007;282(28):20705-20714. doi:10.1074/jbc.M702087200
127. Kocev A, Melamed J, Torgov V, Danilov L, Veselovsky V, Brockhausen I. The wclY gene of Escherichia coli serotype O117 encodes an α 1,4-glycosyltransferase with strict acceptor specificity but broad donor specificity. *Glycobiology.* 2020;30(12):9003-9014. doi:10.1093/glycob/cwaa045
128. Rodrigo-Unzueta A, Ghirardello M, Urresti S, et al. Dissecting the Structural and Chemical Determinants of the “Open-to-Closed” Motion in the Mannosyltransferase PimA from Mycobacteria. *Biochemistry.* 2020;59(32):2934-2945. doi:10.1021/acs.biochem.0c00376
129. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033
130. Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Published online 2014:1-12. doi:10.1186/s12934-014-0125-0
131. Brake AJ, Merryweather JP, Coit DG, et al. *Biochemistry A-Factor-Directed Synthesis and Secretion of Mature Foreign Proteins in Saccharomyces cerevisiae (Epidermal Growth Factor/Protein Fusion/Peptide Processing/Yeast Mating Type/in Vitro Mutagenesis)*. Vol 81.; 1984.
132. Caplan S, Green R, Rocco J, Kurjanlt J. *Glycosylation and Structure of the Yeast Mfotd A-Factor Precursor Is Important for Efficient Transport through the Secretory Pathway*. Vol 173.; 1991.
133. Pelham HRB. The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem Sci.* 1990;15(12):483-486. doi:10.1016/0968-0004(90)90303-S
134. Rossanese OW, Reinke CA, Bevis BJ, et al. *A Role for Actin, Cdc1p, and Myo2p in the Inheritance of Late Golgi Elements in Saccharomyces cerevisiae*. Vol 153.; 2001. <http://www.jcb.org/cgi/content/full/153/1/47>
135. Kainuma M, Chiba Y, Takeuchi M, Jigami Y. Overexpression of HUT1 gene stimulates in vivo galactosylation by enhancing UDP – galactose transport activity in *Saccharomyces cerevisiae*. *Ye.* 2001;18:533-541.

136. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, de Vusser K, Contreras R. In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. *Appl Environ Microbiol*. 2004;70(5):2639-2646. doi:10.1128/AEM.70.5.2639-2646.2004
137. Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N. Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. 2009;4(1). doi:10.1038/nprot.2008.213
138. Partow S, Siewers V, Bjørn S, Nielsen J, Maury J. Characterization of different promoters for designing a new expression vector in *Sa. Yeast*. 2010;27:955-964. doi:10.1002/yea
139. Levine TP, Wiggins CAR, Munro S. Inositol phosphorylceramide synthase is located in the Golgi apparatus of *Saccharomyces cerevisiae*. *Mol Biol Cell*. 2000;11(7):2267-2281. doi:10.1091/mbc.11.7.2267
140. Woo CH, Gao C, Yu P, et al. Conserved function of the lysine-based KXD/E motif in Golgi retention for endomembrane proteins among different organisms. *Mol Biol Cell*. 2015;26(23):4280-4293. doi:10.1091/mbc.E15-06-0361
141. Canals D, Salamone S, Hannun YA. Visualizing bioactive ceramides. *Chem Phys Lipids*. 2018;216:142-151. doi:10.1016/j.chemphyslip.2018.09.013
142. Sinha A. Fluorescence dyes and their usefulness in yeast cell research. *ResearchgateNet*. 2018;(November). <https://www.researchgate.net/publication/329070216>
143. Koning AJ, Lum PY, Williams JM, Wright R. DiOC6 staining reveals organelle structure and dynamics in living yeast cells. *Cell Motil Cytoskeleton*. 1993;25(2):111-128. doi:10.1002/cm.970250202
144. Yang HC, Simon V, Swayne TC, Pon L. *Visualization of Mitochondrial Movement in Yeast.*; 2001.
145. Hong E, Davidson AR, Kaiser CA. A Pathway for Targeting Soluble Misfolded Proteins to the Yeast Vacuole. *J Cell Biol*. 1996;135(3):623-633.
146. Holkeri H, Makarow M. Different degradation pathways for heterologous glycoproteins in yeast. *FEBS Lett*. 1998;429:162-166.
147. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. 2009;37(October 2008):233-238. doi:10.1093/nar/gkn663
148. Quan J, Tian J. Circular Polymerase Extension Cloning. In: Valla S, Lale R, eds. *DNA Cloning and Assembly Methods*. Humana Press; 2014:103-117. doi:10.1007/978-1-62703-764-8_8
149. Trott O, Olson AJ. Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455-461. doi:<https://doi.org/10.1002/jcc.21334>
150. Morris GM, Ruth H, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785-2791. doi:10.1002/jcc.21256
151. Humphrey W, Dalke A, Schulten K. *VMD: Visual Molecular Dynamics.*; 1996.

Bibliography

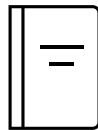


152. Koch M, Stronge VS, Shepherd D, et al. The crystal structure of human CD1d with and without α -galactosylceramide. *Nat Immunol*. 2005;6(8):819-826. doi:10.1038/ni1225
153. Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology*. 2018;28(1):3-8. doi:10.1093/glycob/cwx089

Publications



Publications



Publications





Article

Characterization of a Glycolipid Synthase Producing α -Galactosylceramide in *Bacteroides fragilis*

Marc Caballé, Magda Faijes * and Antoni Planas *

Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, 08017 Barcelona, Spain

* Correspondence: magda.faijes@iqs.url.edu (M.F.); antoni.planas@iqs.url.edu (A.P.)

Abstract: Glycolipids are complex molecules involved in important cellular processes. Among them, the glycosphingolipid α -galactosylceramide has proven to be of interest in biomedicine for its immunostimulatory capabilities. Given its structural requirements, the use of ceramide glycosyltransferase enzymes capable of synthesizing this molecule under in vivo or in vitro conditions is a potential production strategy. Several GT4 enzymes from *Bacteroides fragilis* were considered as potential candidates in addition to the known BF9343_3149, but only this one showed glycolipid synthase activity. The enzyme was expressed as a SUMO fusion protein to produce soluble protein. It is a non-processive glycosyltransferase that prefers UDP-Gal over UDP-Glc as a donor substrate, and maximum activity was found at pH 7.3 and around 30–35 °C. It does not require metal cations for activity as other GT4 enzymes, but Zn^{2+} inactivates the enzyme. The reaction occurs when the ceramide lipid acceptor is solubilized with BSA (100% conversion) but not when it is presented in mixed micelles, and anionic lipids do not increase activity, as in other membrane-associated glycolipid synthases. Further protein engineering to increase stability and activity can make feasible the enzymatic synthesis of α -GalCer for biomedical applications.

Keywords: glycosyltransferase; glycosphingolipids; galactosylceramide



Citation: Caballé, M.; Faijes, M.; Planas, A. Characterization of a Glycolipid Synthase Producing α -Galactosylceramide in *Bacteroides fragilis*. *Int. J. Mol. Sci.* **2022**, *23*, 13975. <https://doi.org/10.3390/ijms232213975>

Academic Editor: Tomohisa Ogawa

Received: 18 October 2022

Accepted: 8 November 2022

Published: 12 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Glycosphingolipids (GSLs) are glycoconjugates that are widely distributed among eukaryotic cells but less common in bacteria. These amphiphilic molecules consist of an acylated sphingoid base (ceramide, Cer) linked to a hydrophilic carbohydrate moiety. They are structurally diverse with regard to the acyl chain (length and modifications) and the sugar moiety attached to the ceramide, from monosaccharides to complex glycans (such as in gangliosides) [1]. GSLs are components of eukaryotic plasma membranes that are not uniformly distributed but clustered in “lipid rafts” (small lateral microdomains of self-associating membrane molecules) and play important roles in modulating receptor proteins in signal transduction pathways [1,2]. Bacterial GSLs are less common, mainly found in α -proteobacteria. Whereas eukaryotic GSL biosynthesis pathways are well characterized [1], bacterial GSL biosynthesis is less studied. A serine-palmitoyl transferase (SPT), which catalyzes the first step of ceramide synthesis, was identified from *Sphingomonas paucimobilis* [3], and some glycosyltransferases that transfer sugar to ceramide acceptor have been reported in some species, such as Sgt1 and Sgt2 in *Caulobacter crescentus* [4]. Some GSL-containing bacteria lack lipopolysaccharides, suggesting that bacterial GSL may be a functional replacement in their outer membrane [5].

In eukaryotic GSLs, the first sugar attached to the ceramide is typically a β -linked galactose (β GalCer) or glucose (β GlcCer). Several glycosyltransferases (GT) responsible for the first transfer to ceramide have been characterized (i.e., the ceramide glucosyltransferase UGCG [6]) as transferring the UDP-sugar donor to the ceramide acceptor with inversion of configuration. A relevant function of GSL in the human immune system is the activation of invariant natural killer T (iNKT) cells. iNKT cells are involved in the suppression of

autoimmune reactions, cancer metastasis, and the graft-rejection response. Some GSLs presented by CD1d (an MHC class I molecule) of dendritic cells can activate a T-cell antigen receptor expressed on the cell surface of iNKT cells, in order to trigger the secretion of several cytokines (such as interleukin IL-2, IL-4 and interferon- γ) as important effectors in the subsequent immune response [7].

iNKT cells can also be activated by non-mammalian α GalCer, with an α -glycosidic linkage not found in mammals. Indeed, agelasphin, first isolated in 1993 as an anti-tumor agent in the marine sponge *Agelas mauritianus*, was identified as α GalCer [8,9]. Chemical optimization of the ceramide portion afforded a α GalCer congener with C18-phytosphingosine and cerotic acid (26:0), named KRN700, as a candidate for clinical applications [10]. More recently, α GalCer was identified as a lipid antigen for iNKT cells when presented by CD1d [11,12]. In the search of other natural sources, α GalCer was isolated from *Bacteroides fragilis*, a common component of the human colon microbiota; it is generally commensal but can become pathogenic, causing infections if displaced into the bloodstream or surrounding tissues following trauma, surgery or disease [13]. α GalCer and other bacterial GSLs are exogenous ligands of CD1d and play important roles in the host's immune system. The bacterial origin of α GalCer supports the hypothesis that the identified congeners in the sponge *Agelas* were actually produced by symbiotic bacteria [7].

Recently, Okino et al. [14] identified a glycosyltransferase from *Bacteroides fragilis* that can catalyze the reaction of α GalCer synthesis. This protein showed homology to ceramide UDP-glucuronosyltransferase (Cer-GlcAT) from *Z. mobilis*, another transferase that uses ceramide as an acceptor. They studied different substrates for the *B. fragilis* GT, showing that ceramide, but apparently not diacylglycerol, was the acceptor, and UDP-galactose or UDP-glucose (with preference for the first) were the donor substrates, but not UDP-GalNAc, UDP-GlcA, UDP-GlcNAc or GDP-Man. The α configuration of the glycosidic linkage confirmed the enzyme as a retaining glycosyltransferase synthesizing α GalCer. Other congeners have also been identified in other gut bacteria, such as *Bacteroides vulgatus* and *Prevotella capri* [15]. The *Bacteroides* α GalCer is thought to be a critical signaling molecule in gut physiology. Studies with mice identified a different congener of α GalCer with a β -hydroxylated palmitic acid and C18-sphinganine (α GalCerMLI) in the mice colon that was not detected in germ-free mice [16]. Since its production was dependent on diet and inflammation, it was suggested that its effector function through iNKT cells is important in gut energy and immunity. Although the synthetic pathway for this unique α GalCer is unknown, the identified *Bacteroides* GT may be involved in its production.

α -GalCer has gained attention as a vaccine adjuvant for the immunotherapy of tumors, by the induction of potent natural killer cell-dependent anti-tumor cytotoxic responses [17–20]. Since the chemical synthesis of α -GalCer is tedious with low yields [21,22], biotechnological production by biocatalysis and cell factory approaches are of current interest, thus, demanding better knowledge of its biosynthetic enzymes. To obtain further insights into the biosynthesis of α GalCer in *B. fragilis*, here, we search for other potential GTs with GSL synthase activity in the *B. fragilis* genome, and biochemically characterize the GT recently identified by Okino et al. [14].

2. Results

2.1. Bioinformatics Search of GT Candidates in the *B. fragilis* Genome

Gene BF9343_3149 from *B. fragilis* was recently identified as encoding a glycolipid synthase (α CerGal_GT) by Ito and collaborators (Okino et al. [14]). The protein sequence shares 23.8% identity with *Z. mobilis* glucuronyl glycosyltransferase, and it was proven to catalyze the reaction between UPD-Gal and ceramide with a retaining mechanism. To obtain insight into glycolipid synthases in *B. fragilis*, we searched for GTs from this microorganism that could also show galactosylceramide transferase activity using the CAZy database, which compiles identified and annotated GTs [23]. A total of 83 putative GT genes from *B. fragilis* NCTC9343 are annotated in this database; however, only retaining GTs (families 3, 4, 5, 8 and 35) were considered in the analysis. GT3, 5 and 35 members were dis-

carded, since these families contain enzymes involved in glycogen and starch metabolism. Families 4 and 8 contain enzymes that use lipids as acceptors, but only GT4 enzymes were considered since it is the only family with reported GT activities using ceramide or diacylglycerol as acceptors. The 22 annotated GT4 sequences from the *B. fragilis* genome were aligned with two biochemically characterized GT4 glycolipid synthases: the non-processive monoglucosyldiacylglycerol synthases from *Acholeplasma laidlawii* [24,25] and *Streptococcus pneumoniae* [25]. The results from the multiple sequence alignment (Supplementary Figure S1) are shown in the form of a neighbor-joining tree in Figure 1. Among the *B. fragilis* GT4 sequences, only two were clustered with the query sequences: the already identified BF9343_3149 (UniProt A0A380YRQ3), and BF9343_1306 (UniProt Q5L7S6). Next, the other four GT genes from *B. fragilis* among the non-classified GT sequences in CAZY were also considered: BF9343_3589 (UniProt Q5L962), BF9343_0008 (UniProt Q5LJ89), BF9343_0585 (UniProt Q5LHL6) and BF9343_0009 (UniProt A0A380YSP3). BLAST searches with these six candidate sequences were performed against the Swiss-Prot, RefSeq and TrEMBL databases. Candidates with high homology ($\geq 50\%$) to other enzymes with known activity were discarded. For example, BF9343_0008 and BF9343_0009 showed high homology with D-inositol 3-phosphate glycosyltransferase (57.5% and 74.2%, respectively), and thus, were rejected. By contrast, BF9343_3589 and BF9343_0585 did not show any significant hits, and BF9343_1306 and BF9343_3149 showed homology (40% and 70%, respectively) to uncharacterized enzymes involved in cell wall synthesis or capsular polysaccharide biosynthesis. Therefore, the GT4 sequence BF9343_1306 and non-classified GT sequences BF9343_3589 and BF9343_0585 were considered as potential candidates with ceramide glycosyltransferase activity, in addition to the known BF9343_3149 (α CerGal_GT).

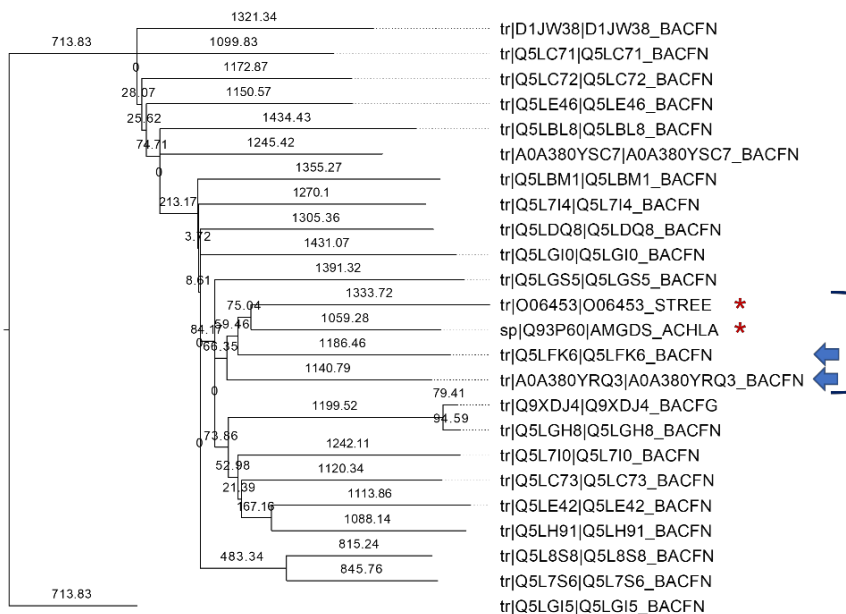


Figure 1. Neighbor-joining phylogenetic tree of annotated GT4 enzymes from *Bacteroides fragilis* and query sequences of monoglucosyldiacylglycerol synthases from *Acholeplasma laidlawii* (UniProt Q93P60) and *Streptococcus pneumoniae* (UniProt O06453) (marked with *). The closest *B. fragilis* sequences are indicated with arrows.

2.2. Identification of α -Galactosylceramide Transferase Activity in *B. fragilis* GT Candidates

The three GT candidates and BF9343_3149 were cloned into pET28a fused with C-terminal His-tag and expressed in *E. coli* cells for activity identification. Cultures were

induced with IPTG at 25 °C and cell pellets were resuspended in phosphate buffer and lysed. The cell extracts and pellets from BF9343_3149 and BF9343_0585 presented the expected new proteins with molecular masses of 45.0 and 63.8 kDa, respectively (Figure 2). By contrast, the expression of BF9343_3589 and BF9343_1306 proteins with theoretical masses of 40.4 and 45.5 kDa were not clear compared to the control (empty pET28a). Nevertheless, all four cell extracts were assayed with UDP-Gal and Cer-NBD (C6-ceramide with fluorescent label, see Materials and Methods) for galactosyltransferase activity on ceramide. In this assay, the Cer-NBD acceptor is solubilized with an equimolar amount of BSA, the cell extract is added, and reactions are initiated by the addition of UDP-Gal at pH 7.5, 37 °C, and monitored by HPLC. Only the cell extract expressing BF9343_3149 showed activity synthesizing the corresponding galactosylceramide (Figure 3).

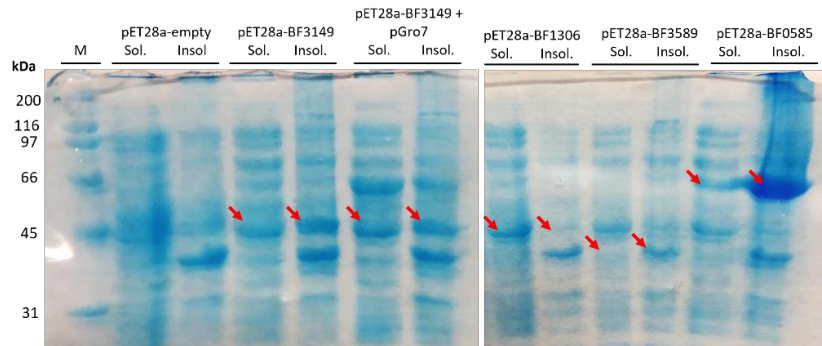


Figure 2. SDS-PAGE gel electrophoresis results of expression of *B. fragilis* GT candidates in *E. coli*. Labels indicate the soluble and insoluble fraction for each culture; M: protein molecular mass marker. Arrows show the expressed proteins. Expected MWs for BF3149, BF1306, BF3589 and BF0585 proteins were 45.0, 45.5, 40.4 and 63.8 kDa, respectively.

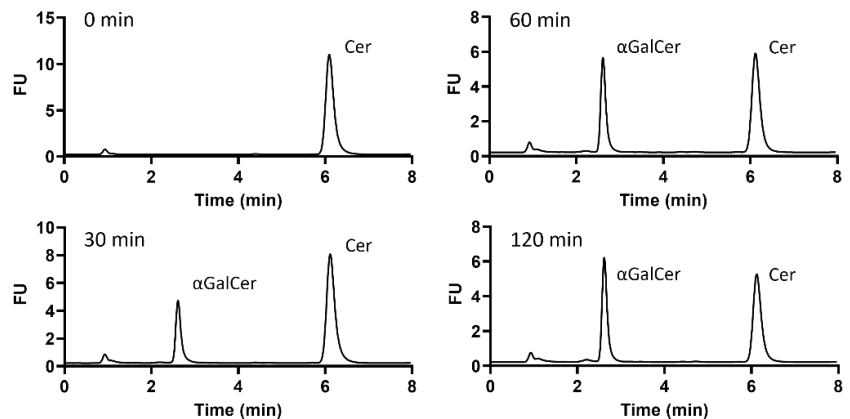


Figure 3. Time-course reaction monitoring of galactosyltransferase activity of a cell extract expressing BF9343_3149. The reaction was carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and, as enzyme source, a cell-free extract from *E. coli*/pET28a-BF3149 (\approx 4 mg/mL total protein) in phosphate buffer at pH 7.5 and 37 °C.

The results confirm that only BF9343_3149, but not BF9343_1306, BF9343_3589 and BF9343_0585, is a glycosylceramide synthase.

2.3. Expression and Purification of *B. fragilis* Glycosyltransferase

The BF9343_3149 amino acid sequence was analyzed to identify possible transmembrane domains or signal peptide sequences to be removed for recombinant protein expression. TMHMM 2.0 [26,27] showed no transmembrane domains in the sequence, and SignalP 5.0 [28] did not predict any signal peptides either. The full-length protein was expressed as a C-terminal His-tagged protein in a pET28a vector. Expression in *E. coli* was explored at different conditions, since yield with standard IPTG induction at 25 °C was low, and the protein mainly remained in the insoluble fraction after cell lysis. Induction at different temperatures and co-expressions with pGro7 vector [29] encoding for GroEL and GroES chaperones was assayed, but the expression yields did not improve. Therefore, preparative protein expression was performed with IPTG induction at 25 °C with no chaperone co-expression.

Solubilization of the cell extract to isolate solubilized protein for purification was further explored. The cells were lysed by sonication in different buffers. Phosphate buffer alone and containing CHAPS, Tris buffer with DTT, and HEPES buffer containing CHAPS and glycerol were assayed. Only cell extract with buffers without detergent presented activity, indicating that the detergent could interfere with the active enzyme.

Despite the presence of active enzyme in the soluble cell extract with phosphate buffer, the purification by metal affinity chromatography did not work. Since activity was only found in the flowthrough and not in the elution step with imidazole, it was thought that the His-tag was not exposed or was hydrolyzed. Then, the BF9343_3149 gene was subcloned into a pET22b expression vector fused with a different affinity tag, Strep-tag at the C-terminus, as well as the SUMO protein at the N-terminus trying to enhance protein solubility, as reported for other proteins [30,31]. Using this new vector, expression was performed with 0.1 mM of IPTG induction at 30 °C for 4 h. The cell-free extract in phosphate buffer showed 3.5 times higher activity than the previous one (Figure 4), proving the effectiveness of the SUMO fusion protein. In addition, purification using Strep-trap affinity chromatography was successful. The eluted protein after dialysis was about 0.39 mg/mL, with an overall yield of 0.6 mg per liter of culture.

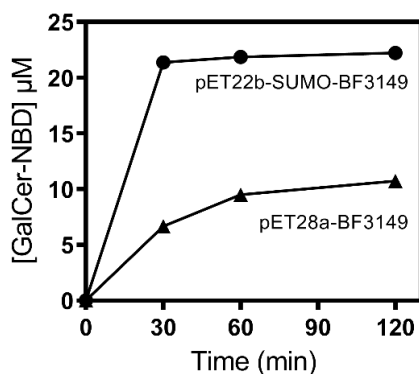


Figure 4. Ceramide-NBD activity assay of α GalCer_{GT} (BF9343_3149) cell-free extracts from different expression systems. Reactions were carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and cell-free extracts of pET28a-BF9343_3149 and pET22b-SUMO-BF9343_3149 cultures (total protein concentration in the assay 9.9 ± 0.4 mg/mL), in 50 mM of phosphate buffer and 150 mM of NaCl at pH 7.5 and 37 °C.

2.4. Kinetic Characterization

2.4.1. Specific Activity

Enzyme activity was determined at 1.25 mM of UDP-Gal donor and 25 μ M of Cer-NBD acceptor solubilized as a complex with BSA in equimolar concentration (Cer-NBD:BSA

1:1) in phosphate buffer, pH 7.5, 37 °C. Time-course monitoring by HPLC showed that the enzyme forms α GalCer as a unique product and the reaction is complete after 15 min (Figure 5A and Figure S2). The specific activity was calculated to be $3.27 \cdot 10^{-2} \text{ s}^{-1}$ from the linear dependence of initial velocity with enzyme concentration (Figure 5B).

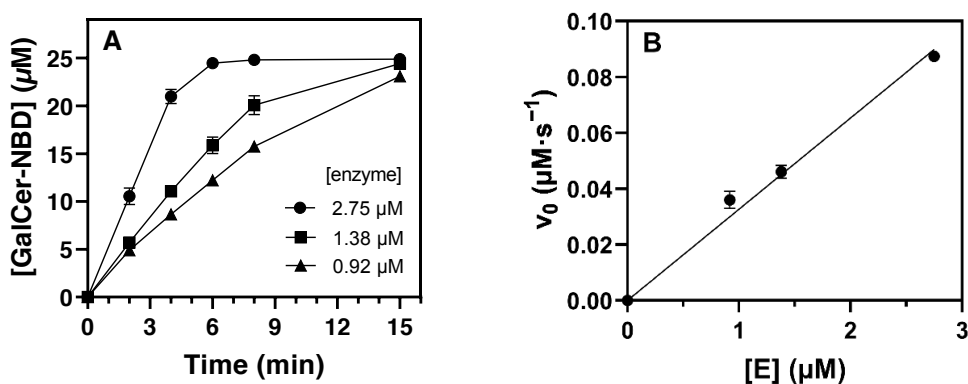


Figure 5. (A) Time course of α GalCer_{GT} reaction. (B) Linear dependence of initial rates with enzyme concentration. Reactions were carried out with 25 μM of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 1 to 3 μM enzyme in 50 mM of phosphate buffer and 150 mM of NaCl at pH 7.5 and 37 °C.

Unexpectedly, the stability of the purified enzyme was a key issue. Enzymatic activity was only completely maintained for 2 h, about 39% at 6 h and nearly lost at 24 h. This behavior was similar in Tris buffer and in the presence of 20% glycerol, while in the cell extract, the enzyme was stable until 5 h and kept 42% of activity after 24 h (Figure 6). Therefore, all activity data presented are for freshly expressed and purified protein (less than 2 h storage after purification).

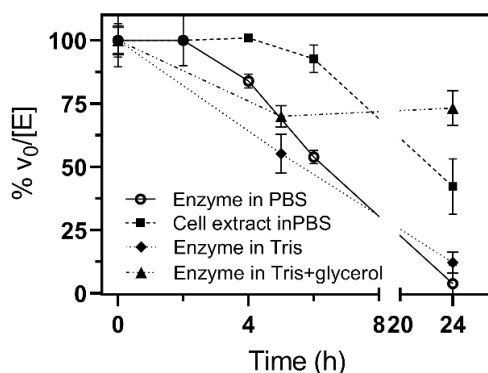


Figure 6. Stability of α GalCer_{GT}. The specific activity of freshly prepared enzyme stored at 4 °C in phosphate, pH 7.5, was measured at different times. Assay conditions: 25 μM of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 1.39 μM of purified enzyme or soluble cell extracts (0.77 μM of active enzyme), in 50 mM of phosphate or Tris buffer and 150 mM of NaCl at pH 7.5 and 37 °C.

2.4.2. Metal Binding

Since cation requirement is common in glycosyltransferases, metal binding was also studied. Buffers used for protein purification and activity assays did not contain added

metal cations. Treatment with EDTA did not reduce the activity relative to untreated and freshly purified enzyme, indicating that the enzyme either does not bind divalent cations or that the metal cation is strongly bound to the active site. Mg^{2+} and Mn^{2+} at 0.25 mM slightly increased the enzyme activity, but the effect was reduced at a higher concentration, with some inhibition at 1 mM of Mg^{2+} , which reduced the activity to 70%. Whereas Ca^{2+} at 0.25 mM did not significantly affect the enzyme activity, Zn^{2+} inactivated the enzyme (Figure 7).

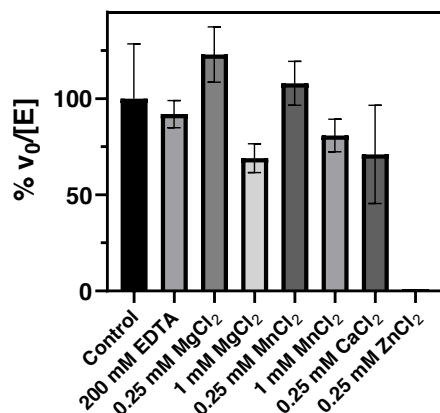


Figure 7. Effect of metal cations on α GalCer_GT activity. Specific activities ($v_0/[E]$) were compared and plotted as % activity relative to the activity of freshly purified enzyme (control). Reaction conditions: 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and 35–70 nM of enzyme in 20 mM of Tris buffer and 200 mM of NaCl at pH 7.5 and 37 $^{\circ}$ C.

2.4.3. Optimal pH and Temperature

The pH profile of specific activity using phosphate–citrate buffers follows a bell-shaped curve with the optimum catalytic efficiency at a pH of around 7.3 (Figure 8A). Between 6 and 8.5, the enzyme retained more than 70% activity, while it was essentially inactive at pH < 5 and >9.5. In addition, α CerGal_GT remained catalytically active from 25 $^{\circ}$ C to 40 $^{\circ}$ C with an optimal temperature around 30 $^{\circ}$ C, and it was inactivated at 50 $^{\circ}$ C (Figure 8B).

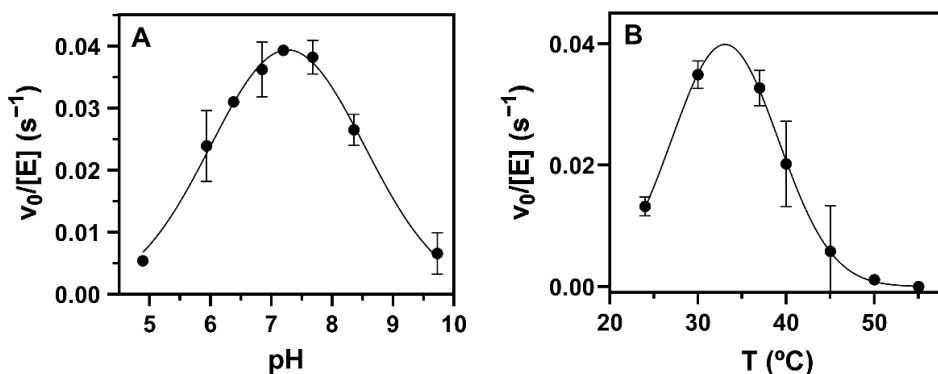


Figure 8. Optimal pH and temperature profiles of α GalCer_GT activity. Reactions were carried out with 25 μ M of Cer-NBD in the presence of equimolar BSA, 1.25 mM of UDP-Gal and different enzyme concentrations. (A) pH profile between 5 and 9.5 with 50 mM/50mM of citrate/phosphate buffers adjusted at each pH, 37 $^{\circ}$ C and 830 nM of enzyme. (B) Temperature profile between 25 and 55 $^{\circ}$ C in phosphate buffer at 7.5 and 1.1–3.2 μ M of enzyme.

2.4.4. Kinetic Parameters for Donor and Acceptor Substrates

Kinetics varying UDP-Gal at a saturating Cer-NBD concentration (25 μM) solubilized with BSA, and varying Cer-NBD concentration at a saturating UDP-Gal concentration (1.25 mM) at pH 7.5 and 37 $^{\circ}\text{C}$, obeyed Michaelis–Menten kinetics with some substrate inhibition (Figure 9). K_M for UDP-Gal is 218 μM , with a catalytic efficiency of $0.22\text{ mM}^{-1}\cdot\text{s}^{-1}$, and a slight donor substrate inhibition can be observed above 2 mM. For the NBD-ceramide acceptor at saturating donor conditions (1.25 mM), kinetics show substrate inhibition with maximum activity at 25 μM of Cer-NBD, K_M of 4.6 μM and K_I of 161 μM .

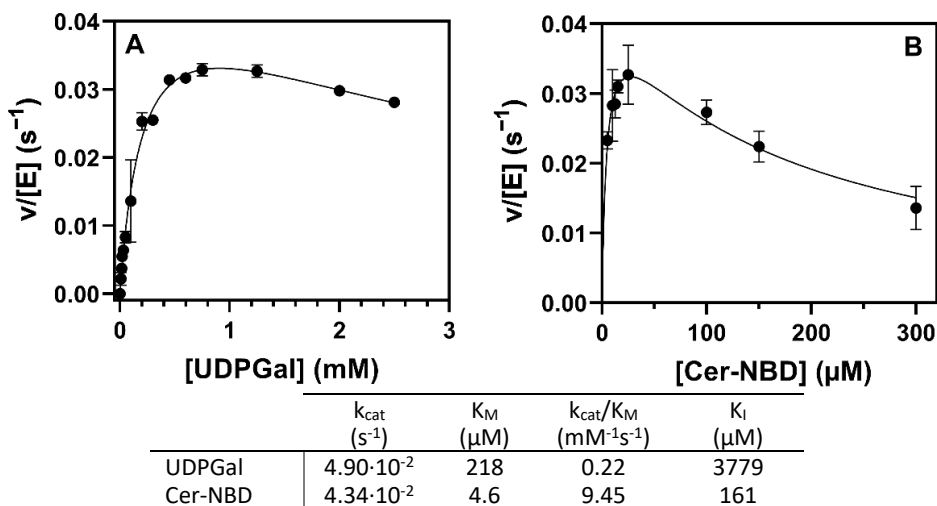


Figure 9. Kinetics of $\alpha\text{GalCer_GT}$. (A) Michaelis–Menten curve for UDP-Gal. Reactions conditions: 10–2500 μM of UDP-Gal, 25 μM of Cer-NBD in the presence of equimolar BSA, 0.2–0.5 μM of enzyme, at pH 7.5 and 37 $^{\circ}\text{C}$. (B) Michaelis–Menten curve for Cer-NBD. Reactions conditions: 1.25 mM of UDP-Gal, 5–300 μM of Cer-NBD in the presence of equimolar BSA, 0.35–0.75 μM of enzyme, at pH 7.5 and 37 $^{\circ}\text{C}$.

In addition, UDP-Gal presents 14.5-fold higher specific activity than UDP-Glc, using NBD-ceramide as an acceptor at a 1.25 mM/25 μM donor/acceptor ratio (Figure S3). Assuming saturating conditions for both donors and acceptor, UDP-Gal is a better donor than UDP-Glc in terms of k_{cat} .

2.4.5. Activity in Mixed Micelles and in the Presence of Anionic Lipids

The $\alpha\text{CerGal_GT}$ kinetic characterization described above was performed with Cer-NBD solubilized with BSA. Since other bacterial and plant glycolipid synthases present activity on lipid micelles and require anionic phospholipids for activation [24,32,33], the effect of presenting the acceptor in micelles was evaluated using Cer-NBD with DOPC and/or DOPG mixed vesicles (DOPC, dioleoylphosphatidylcholine as neutral lipid, and DOPG, dioleoylphosphatidylglycerol as anionic lipid). After the incubation of Cer-NBD in micelles with the enzyme for 30 min, UDP-Gal was added. As can be observed in Figure 10, enzyme activity decreases considerably in ceramide/DOPC or DOPG micelles compared with CerNBD solubilized with BSA. Neither the presence of neutral or anionic lipids in vesicles activates the enzyme, but rather inhibits it, since the addition of Cer-NBD solubilized with BSA to the enzyme preincubated with DOPC vesicles only results in 22% of the original activity.

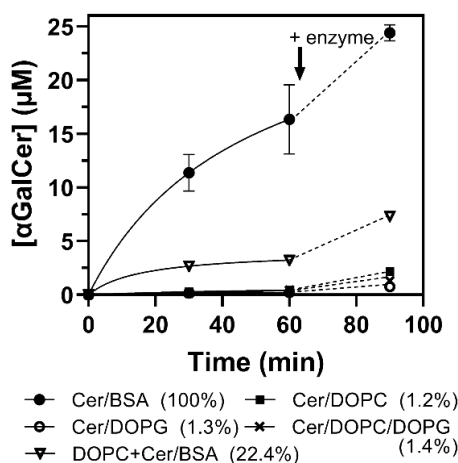


Figure 10. Reaction monitoring of α CerGal_{GT} activity using ceramide in mixed micelles as acceptor. Reactions were carried out with 25 μ M of Cer-NBD, 1.25 mM of UDP-Gal, 0.25 μ M of enzyme. Ceramide was solubilized with BSA or in mixed micelles with DOPC (1 mM) or DOPC/DOPG 0.6/0.4 (total 1 mM) or DOPG (1 mM). Moreover, activity was measured with Cer-NBD in BSA in the presence of DOPC (1 mM). Legend (in parenthesis): % activity at 30 min. At 60 min reaction, a fresh aliquot of enzyme was added (indicated with an arrow).

2.5. Modeled α GalCer_{GT} Structure and Ligand Binding

Based on the three-dimensional AlphaFold model of the free enzyme, a preliminary structural analysis was performed to evaluate the topology of the active site and identify residues interacting with the substrate. The α GalCer_{GT} amino acid sequence was aligned with GT4 enzymes with a solved crystal structure (see Supporting Information, Table S1 and Figure S4). Likewise, the AlphaFold model of α GalCer_{GT} was structurally aligned with the 3D structures of solved GT4 enzymes (Figure S5). α GalCer_{GT} does not differ from the typical GT-B fold of GT4 enzymes, comprising two separate $\beta/\alpha/\beta$ Rossmann-fold domains that form an inter-domain substrate-binding crevice. To locate the UDP-Gal donor binding site in the enzyme 3D model structure, docking experiments were performed, highlighting first-shell amino acid residues (Figure 11).

The closest protein with a solved X-ray structure is the phosphatidylinositol mannosyl-transferase (PimA) from *Mycobacterium smegmatis* (Uniprot: A0QWG6) with 42.4% similarity and 24.4% identity (Figure S6). The structural alignment of the modeled α GalCer_{GT} with UDP-Gal ligand with the 3D structure of a complex PimA with bound GDP-Man substrate [34] allowed a comparison of the similar binding site topology (Figure S7) and first-shell amino acid residues in the binding site, although it should be taken as a tentative model due to the relatively low sequence identity between both proteins. Residues of α GalCer_{GT} interacting with the UDP nucleotide moiety comprise Val209, Arg269, Val294 and Glu297 (uracil), Lys216, Phe284 and Leu293 (ribose), whereas Thr19, His117, and Asp289 interact with the galactose unit (Figure 11). Acceptor docking experiments have not been attempted because the 3D enzyme model corresponds to the free enzyme, and a domain closure is expected upon acceptor binding, as is often observed in GT enzymes with a GT-B fold [35].

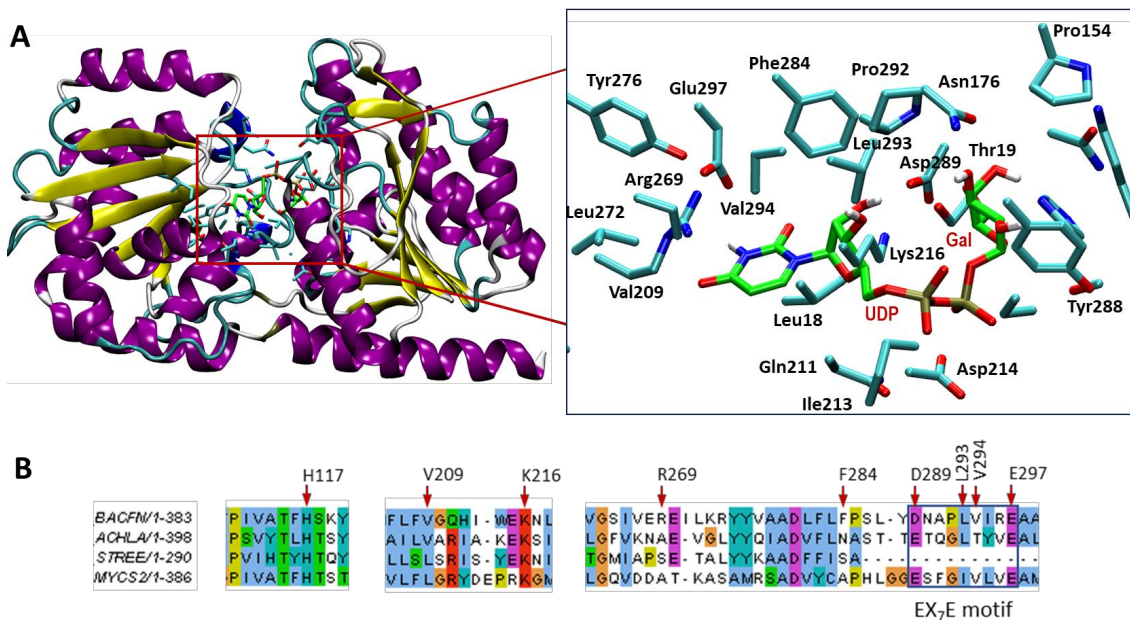


Figure 11. (A) Structural model of α CerGal_GT with bound UDP-Gal, generated by docking UDP-Gal into the AlphaFold model of the free enzyme (retrieved from Uniprot A0A380YRQ3). Right, magnification of the donor binding site, showing the amino acid side chains of first-shell residues interacting with UDP-Gal. (B) Sequence alignment of regions surrounding donor binding residues, GT4 glycolipid synthases from *Bacteroides fragilis* (α CerGal_GT, top), *Acholeplasma laidlawii*, *Streptococcus pneumoniae* and *Mycobacterium smegmatis*.

3. Discussion

The *B. fragilis* genome contains 83 putative glycosyltransferases annotated in the CAZY database, from which GT sequences belonging to the GT4 family were selected, since it is the only retaining GT family with characterized glycolipid synthases using ceramide or diacylglycerol as acceptors. Only BF9343_3149, previously identified by Ito and coworkers [14], proved to have galactosylceramide transferase activity.

Family 4 of glycosyltransferases is a broad family that comprises many different activities. The common features of the family are the retaining mechanism of action and the GT-B fold structure. α GalCer_GT has been reported to retain the α linkage of UDP-Gal in the α GalCer product, which matches with the mechanism of action of this family [14]. Likewise, sequence analysis shows two clear domains, which represent both domains of a GT-B structure. α GalCer_GT conserves the typical EX7E motif essential for activity in GT4 enzymes [36], but with an Asp for the first Glu (Figure 11B).

The full-length protein was mainly expressed as an insoluble protein, but fused with SUMO it significantly improved the solubility. Purification by affinity chromatography (Step-tagged protein) afforded pure enzyme in about 0.6 mg per liter of culture. The purified protein, however, proved to be rather unstable, with a loss of activity in 24 h (Figure 6), and all enzymatic reactions were carried out within a time window of 2 h after enzyme purification. This instability is an important issue that should be addressed in further work to improve stability and enable applications in enzymatic synthesis.

α GalCer_GT was shown to be a non-processive GT, since a single product was obtained with 100% conversion (Figure S2). It prefers UDP-Gal over UDP-Glc as a donor substrate (14.5-fold higher specific activity for UDP-Gal, Figure S3) and has no activity with UDP-GalNAc, UDP-GlcA, UDP-GlcNAc or GDP-Man, as reported [14]. It shows maximal

activity at pH 7.3 and 30–35 °C and is fully inactive at 55 °C (Figure 8), indicating poor thermostability, in agreement with the observed instability upon storage even at 4 °C.

GT enzymes with a GT-A fold are metallo-dependent enzymes and exhibit the conserved DXD motif that binds a metal cation, which facilitates leaving group departure by coordinating the phosphate groups of the sugar nucleotide donor. In contrast, GT-B enzymes do not have the DXD signature but instead use positively charged side chains and/or hydroxyls and helix dipoles to facilitate leaving group departure during catalysis [37]. Metal cations are not essential for enzymatic activity, but in some cases reaction rates are accelerated by certain metal cations [38]. In α GalCer_GT, there is no evidence of metal cation requirement. EDTA did not reduce enzyme activity, the addition of Mg^{2+} or Mn^{2+} did not significantly affect the activity, but Zn^{2+} inactivated the enzyme (Figure 7). However, related GT4 glycolipid synthases such as GlcAT from *Z. mobilis* and α -monoglucosyldiacylglycerol synthase from *A. laidlawii* did show metallo-dependent activity, despite the lack of a DXD motif [14,39].

As for lipid acceptors, Okino et al. [14] showed that ceramide with a C12 chain length (C12-Cer-NBD) was the preferred acceptor, whereas 50% activity was observed with C6-Cer-NBD and no activity was detected with diacylglycerol (DG-NBD). However, the conditions that explain how the lipidic acceptor substrate was presented were not detailed. Other glycolipid synthases are membrane-associated proteins, i.e., monoglucosyldiacylglycerol synthases from *A. laidlawii* [39] and *Streptococcus pneumoniae* [25], processive MG517 from *Mycoplasma genitalium* [32] or PimA from *M. smegmatis* [34], which require anionic lipids for activation and being active with the lipid acceptor presented in mixed micelles. Our results indicate that α GalCer_GT is fully active with the ceramide acceptor solubilized with BSA, but activity was largely reduced when using ceramide in lipid vesicles, and no activation by anionic lipids was observed (Figure 10). Furthermore, when ceramide was solubilized with BSA but there were DOPC vesicles present in the reaction, enzyme activity was also substantially reduced. It seems that the enzyme is not membrane-associated, since the presence of detergents in the lysis buffer when extracting the protein did not improve protein solubilization. Altogether, these results suggest that the enzyme takes the lipidic acceptor presented by a lipid–protein complex rather than from membranes, but it requires further studies to elucidate the mechanism.

In conclusion, we report the biochemical characterization of the first recently discovered GT enzyme synthesizing α -galactosylceramide, a relevant non-mammalian glycolipid with great interest as an effector in immune responses and with applications as a vaccine adjuvant.

4. Materials and Methods

4.1. In Silico Analysis

The CAZy database [23] was used to search for glycosyltransferase families in *Bacteroides fragilis*. Protein sequences were obtained from the UniProt database. Jalview was used for sequence alignments using MAFFT and TCOFFEE algorithms [40], with a neighbor-joining phylogenetic tree calculated using BLOSUM62 scores as a distance metric. A protein homology search was carried out using BLAST tools from NCBI and UNIPROT.

4.2. Strains, Plasmids and Reagents

DH5 α *E. coli* cells were used for cloning and molecular biology work. The cells were cultured in LB broth with the appropriate antibiotic at 37 °C, either on liquid culture or agar plates. BL21(DE3) Star *E. coli* cells were used for protein expression. pET28a and pET22b were from Novagen (Merck, Darmstadt, Germany). Ceramide-C6-NBD (Cer-NBD) (N-[6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl]-D-erythro-sphingosine) was provided by the synthesis facility at the Institut de Química Avançada de Catalunya, Barcelona, Spain. UDP-Galactose sodium salt was obtained from Merck.

4.3. Cloning of *B. fragilis* GT Sequences into *E. coli* Expression Vector

- (a) pET28a vector-Protein sequences were obtained from the UniProt database and then translated to DNA code optimized for *E. coli*. Synthetic sequences were obtained from GeneArt (ThermoFischer Scientific, Waltham, MA, USA) and cloned using restriction enzymes NcoI and XhoI into pET28a vector to obtain the expression vectors pET28a-HisTag-BF9343_XXXX. Positive transformants were selected on LB plates with 50 µg/mL of kanamycin. Successful plasmid constructions were verified by Sanger sequencing.
- (b) pET22b-Strep-SUMO vector-Sequence BF9343_3149 was later cloned into a pET22b vector containing a N-terminal fused Strep-tag and SUMO protein. Cloning was carried out by Circular Polymerase Extension cloning (CPEC), amplifying the linear vector and the insert containing BF9343_3149 with primers sharing homologous regions, and CPEC reaction to obtain pET22b-SUMO-BF9343_3149-StrepTag. Positive transformants were selected on LB plates with 100 µg/mL of ampicillin. Successful plasmid construction was verified by Sanger sequencing.

4.4. Expression of GT Candidates and α GalCer_{GT} in *E. coli*

pET28a-HisTag-BF9343_XXXX.-BL21(DE3)Star cells harboring pET28a containing *B. fragilis* GT sequences were cultured in 300 mL of LB containing kanamycin at a final concentration of 50 µg/mL, inoculated at an initial OD of 0.05 and incubated at 37 °C until OD 1. At this point, protein expression was induced by adding IPTG to a final concentration of 1 mM, and cultures were incubated at 25 °C for 4 h. The cells were then harvested at 5000× *g* for 25 min and washed with 15 mL of NaCl 0.9%. The cell pellets were kept at −20 °C until use.

pET22b-SUMO-BF9343_3149-StrepTag.-BL21(DE3)Star cells harboring pET22b-SUMO-BF9343_3149-StrepTag were grown in 600 mL of LB containing ampicillin at a final concentration of 100 µg/mL, inoculated at an initial OD of 0.05 and incubated at 37 °C until OD = 1. At this point, protein expression was induced by adding IPTG to a final concentration of 0.1 mM, and cultures were incubated at 30 °C for 4 h. The cell pellets were obtained as previously described.

4.5. Protein Purification by Strep-Tag Affinity Chromatography

Frozen cell pellets from the 600 mL cultures were thawed to room temperature and resuspended in 15 mL of lysis buffer (50 mM of Na₂HPO₄, 150 mM of NaCl, pH 7.5) supplemented with 1mM of PMSF (phenylmethylsulfonyl fluoride). The cells were disrupted by sonication in a Soniprep 150 sonifier at 4 °C (10 min, 10 s ON/20 s OFF, 50% amplitude). The lysate was centrifuged at 25,000× *g* for 60 min. The supernatant was recovered and filtered (0.45 µm) before purification with a StrepTrap HP 1 mL column (GE Healthcare, Chicago, IL, USA). The protein was eluted with lysis buffer with 2.5 mM of d-desthiobiotin. Eluted fractions were combined and dialyzed using Amicon Ultra-15 10 kDa (Millipore) with lysis buffer. The final retained fraction was recovered, and volume was adjusted to 1 mL with lysis buffer. Protein was quantified using the Bradford assay.

4.6. Galactosylceramide Transferase Activity Assay

Glycolipid synthase activity from fresh cell-free extracts or purified enzyme was measured with Cer-NBD as an acceptor and UDP-Gal (or UDP-Glc) as a donor by HPLC with fluorescence detector (HPLC 1200 Agilent with fluorescence detector, excitation wavelength at 458 nm and emission at 530 nm, Nova-pak C18 column, flowrate of 1 mL·min^{−1}, eluent 75% acetonitrile in water), as described by Orive et al. 2020 [41]. Cer-NBD was solubilized with BSA at an equimolar concentration or using mixed vesicles with DOPC or/and DOPG, as explained below. Product (α GalCer-NBD) was quantified from the chromatograms from the relative areas of substrate and product:

$$[\alpha\text{GalCer-NBD}] = [\text{Cer-NBD}]_0 \times \text{Area}_{\alpha\text{GalCer-NBD}} / (\text{Area}_{\text{Cer-NBD}} + \text{Area}_{\alpha\text{GalCer-NBD}}).$$

- (a) Ceramide-NBD with BSA: Reactions were performed with 25 μM of Cer-NBD, 25 μM of BSA, 1.25 mM of UDP-Gal, phosphate buffer (50 mM of Na_2HPO_4 , 150 mM of NaCl, pH 7.5) and the appropriate amount of enzyme in a 200 μL final reaction volume. It was preincubated at 37 $^\circ\text{C}$ for five minutes and the reaction was started by adding UDP-Gal. Sampling time was adjusted depending on enzyme concentration. Aliquots were withdrawn and mixed with methanol (MeOH:sample 8:2) and centrifuged to eliminate debris. The samples were then analyzed by HPLC. Chromatographic peaks were assigned by co-injection with independent standards. Initial rates were obtained from the linear progress curve of product formation. Initial rates were determined from triplicate assays.
- (b) Ceramide-NBD in vesicles: Ceramide-NBD, DOPG and/or DOPG (matrix lipid) were dissolved in chloroform in a glass vial using the desired amounts for each experiment, as indicated. Chloroform was evaporated under a N_2 stream and the lipid mixture dried under vacuum for 2 h. When the reactions were performed, the lipid film was solubilized to homogeneity in 80 μL of phosphate buffer solution (50 mM of Na_2HPO_4 , 150 mM of NaCl, pH 7.5) by extensive vortexing and ultrasound bath for 5 min, and kept on ice for 1 min. This cycle was repeated 8 times. Vesicle formation was analyzed using ZetaSizer (Malvern Panalytical, Malvern, UK). For the enzymatic reaction, 25 μL of enzyme was added to 60 μL of freshly prepared vesicles solution. The mixture was kept for 30 min on ice and preincubated for 5 min at 37 $^\circ\text{C}$. The reaction was started by adding 12.5 μL of UDP-Gal and incubated at 37 $^\circ\text{C}$. The reaction conditions were: 1.25 mM of UDP-Gal, 25 μM of Cer-NBD, DOPC/DOPG vesicles (condition 1:1 mM of DOPC; 2:1 mM of total DOPC and DOPG (40 mol % DOPG); 3:1 mM of DOPG, phosphate buffer (50 mM of Na_2HPO_4 , 150 mM of NaCl, pH 7.5), and enzyme (0.92–2.75 μM) in a 200 μL reaction volume. Aliquots were withdrawn every 30 min for 90 min, stopped as before and analyzed by HPLC.

4.7. Kinetic Parameters

UDP-Gal (0.1–2.5 mM) and ceramide-NBD (5–300 μM) with an equimolar amount of BSA were individually varied, maintaining constant the other components. Initial rates vs. donor or acceptor concentrations were fitted to a Michaelis–Menten equation with substrate inhibition by non-linear regression, using Prism 8 software (GraphPad, San Diego, CA, USA), from which the kinetic parameters were derived.

4.8. Effect of Metal Cations

Tris buffer was used when analyzing metal binding to avoid precipitation of divalent cations. The cell pellet was resuspended with 20 mM of Tris buffer with 200 mM of NaCl, 1 mM of EDTA and 1 mM of DTT at pH 7.5 supplemented with 1 mM of PMSF. Enzyme purification was carried out following the same protocol as before, but with lysis buffer, 20 mM of Tris buffer and 200 mM of NaCl at pH 7.5. The reactions were performed with 25 μM of Cer-NBD, 25 μM of BSA, 1.25 mM of UDP-Gal, the appropriate amount of enzyme, Tris buffer (20 mM of Tris, 150 mM of NaCl, pH 7.5) with or without EDTA (200 mM), and the addition of metal ions (0.25–1 mM) in a 200 μL final reaction volume. It was preincubated at 37 $^\circ\text{C}$ for five minutes, and the reaction was started by adding UDP-Gal.

4.9. Modeled 3D Structure and Ligand Docking

The AlphaFold model of *B. fragilis* $\alpha\text{GalCer_GT}$ was obtained from Uniprot A0A380YRQ3. UDP-Gal structure was taken from PDB 5M7D. The complex $\alpha\text{GalCer_GT}\cdot\text{UDP-Gal}$ was generated by docking with AUTODOCK VINA [42]. Both the protein and ligand structure were first parametrized with AutoDockTools4 [43]: polar hydrogens were added, Auto-Dock4.2 atom typing was used, and Gasteiger partial charges were computed. All the rotatable bonds of the ligands were considered free during the docking calculations, whereas the whole protein structure was kept fixed. The docking search space was confined in a box centered in the active site. The exhaustiveness level was set to 24, and 20 binding

modes of the ligands were generated. Only low-energy binding poses were considered for analysis. Pictures of the complexes were generated with VMD [44].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms232213975/s1>, Reference [45] is cited in the Supplementary Material.

Author Contributions: M.C., experimental investigation, data analysis and writing first draft; M.F., conceptualization, experimental design, data analysis and reviewing the manuscript; A.P. conceptualization, data analysis, supervision, writing and reviewing the manuscript, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: Work supported by grant PID2019-104350RB-I00 from Ministry of Science and Innovation (MICINN), Spain (to A.P.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: M.C. acknowledges a predoctoral contract from AGAUR, Generalitat de Catalunya. We thank Aitor Vega from the Laboratory of Biochemistry for the docking experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Quinville, B.M.; Deschenes, N.M.; Ryckman, A.E.; Walia, J.S. A Comprehensive Review: Sphingolipid Metabolism and Implications of Disruption in Sphingolipid Homeostasis. *Int. J. Mol. Sci.* **2021**, *22*, 5793. [CrossRef] [PubMed]
2. Hannun, Y.A.; Obeid, L.M. Sphingolipids and their metabolism in physiology and disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 175–191. [CrossRef] [PubMed]
3. Ikushiro, H.; Hayashi, H.; Kagamiyama, H. A Water-soluble Homodimeric Serine Palmitoyltransferase from *Sphingomonas paucimobilis* EY2395 T Strain complex from Chinese hamster ovary cells. *J. Biol. Chem.* **2001**, *276*, 18249–18256. [CrossRef] [PubMed]
4. Stankeviciute, G.; Guan, Z.; Goldfine, H.; Klein, E.A. *Caulobacter crescentus* Adapts to Phosphate Starvation by Synthesizing Anionic Glycoglycerolipids and a Novel Glycosphingolipid. *mBio* **2019**, *10*, e00107-19. [CrossRef]
5. Kawasaki, S.; Moriguchi, R.; Sekiya, K.; Nakai, T.; Ono, E.; Kume, K.; Kawahara, K. The cell envelope structure of the lipopolysaccharide-lacking gram-negative bacterium *Sphingomonas paucimobilis*. *J. Bacteriol.* **1994**, *176*, 284–290. [CrossRef]
6. Health, R. Expression cloning of a cDNA for human ceramide glucosyltransferase that catalyzes the first glycosylation step of glycosphingolipid synthesis. *Proc. Nat. Acad. Sci. USA* **1996**, *93*, 4638–4643.
7. Bendelac, A.; Savage, P.B.; Teyton, L. The Biology of NKT Cells. *Annu. Rev. Immunol.* **2007**, *25*, 297–336. [CrossRef]
8. Natori, T.; Koezuka, Y.; Higa, T. Agelasphins, novel α -galactosylceramides from the marine sponge *Agelas mauritanus*. *Tetrahedron Lett.* **1993**, *34*, 5591–5592. [CrossRef]
9. Natori, T.; Morita, M.; Akimoto, K.; Koezuka, Y. Agelasphins, novel antitumor and immunostimulatory cerebrosides from the marine sponge *Agelas mauritanus*. *Tetrahedron* **1994**, *50*, 2771–2784. [CrossRef]
10. Morita, M.; Motoki, K.; Akimoto, K.; Natori, T.; Sakai, T.; Sawa, E.; Yamaji, K.; Koezuka, Y.; Kobayashi, E.; Fukushima, H. Structure-Activity Relationship of α -Galactosylceramides against B16-Bearing Mice. *J. Med. Chem.* **1995**, *38*, 2176–2187. [CrossRef]
11. McEwen-Smith, R.M.; Salio, M.; Cerundolo, V. CD1d-dependent endogenous and exogenous lipid antigen presentation. *Curr. Opin. Immunol.* **2015**, *34*, 116–125. [CrossRef] [PubMed]
12. Kohlgruber, A.C.; Donado, C.A.; LaMarche, N.M.; Brenner, M.B.; Brennan, P.J. Activation strategies for invariant natural killer T cells. *Immunogenetics* **2016**, *68*, 649–663. [CrossRef] [PubMed]
13. Brown, L.C.W.; Penaranda, C.; Kashyap, P.C.; Williams, B.B.; Clardy, J.; Kronenberg, M.; Sonnenburg, J.L.; Comstock, L.E.; Bluestone, J.A.; Fischbach, M.A. Production of α -Galactosylceramide by a Prominent Member of the Human Gut Microbiota. *PLoS Biol.* **2013**, *11*, e1001610. [CrossRef]
14. Okino, N.; Li, M.; Qu, Q.; Nakagawa, T.; Hayashi, Y.; Matsumoto, M.; Ishibashi, Y.; Ito, M. Two bacterial glycosphingolipid synthases responsible for the synthesis of glucuronosylceramide and α -galactosylceramide. *J. Biol. Chem.* **2020**, *295*, 10709–10725. [CrossRef] [PubMed]
15. von Gerichten, J.; Schlosser, K.; Lamprecht, D.; Morace, I.; Eckhardt, M.; Wachten, D.; Jennemann, R.; Gröne, H.-J.; Mack, M.; Sandhoff, R. Diastereomer-specific quantification of bioactive hexosylceramides from bacteria and mammals. *J. Lipid Res.* **2017**, *58*, 1247–1258. [CrossRef] [PubMed]

16. von Gerichten, J.; Lamprecht, D.; Opálka, L.; Soulard, D.; Marsching, C.; Pilz, R.; Sencio, V.; Herzer, S.; Galy, B.; Nordström, V.; et al. Bacterial immunogenic α -galactosylceramide identified in the murine large intestine: Dependency on diet and inflammation. *J. Lipid Res.* **2019**, *60*, 1892–1904. [CrossRef] [PubMed]
17. Hayakawa, K.; Smiley, M.T.K.S.; Grusby, M.J.; Gui, M.; Burdin, M.N.; Brossay, L.; Koezuka, Y. Selective Ability of Mouse CD1 to Present Glycolipids: α -Galactosylceramide Specifically Stimulates Va14+ NK T Lymphocytes. *J. Immunol.* **1998**, *161*, 3271–3281. Available online: <http://www.jimmunol.org/content/161/7/3271> (accessed on 18 September 2022).
18. Shimizu, K.; Kurosawa, Y.; Taniguchi, M.; Steinman, R.M.; Fujii, S.-i. Cross-presentation of glycolipid from tumor cells loaded with α -galactosylceramide leads to potent and long-lived T cell-mediated immunity via dendritic cells. *J. Exp. Med.* **2007**, *204*, 2641–2653. [CrossRef]
19. Mattarollo, S.R.; West, A.C.; Steegh, K.; Duret, H.; Paget, C.; Martin, B.; Matthews, G.M.; Shortt, J.; Chesi, M.; Bergsagel, P.L.; et al. NKT cell adjuvant-based tumor vaccine for treatment of myc oncogene-driven mouse B-cell lymphoma. *Blood* **2012**, *120*, 3019–3029. [CrossRef]
20. Gableh, F.; Saeidi, M.; Hemati, S.; Hamdi, K.; Soleimanjahi, H.; Gorji, A.; Ghaemi, A. Combination of the toll like receptor agonist and α -Galactosylceramide as an efficient adjuvant for cancer vaccine. *J. Biomed. Sci.* **2016**, *23*, 1–11. [CrossRef]
21. Plettenburg, O.; Bodmer-Narkevitch, A.V.; Wong, C.-H. Synthesis of α -Galactosyl Ceramide, a Potent Immunostimulatory Agent. *J. Org. Chem.* **2002**, *67*, 4559–4564. [CrossRef] [PubMed]
22. Janssens, J.; Decruy, T.; Venken, K.; Seki, T.; Krols, S.; Van der Eycken, J.; Tsuji, M.; Elewaut, D.; Van Calenbergh, S. Efficient Divergent Synthesis of New Immunostimulant 4'-Modified α -Galactosylceramide Analogues. *ACS Med. Chem. Lett.* **2017**, *8*, 642–647. [CrossRef] [PubMed]
23. Drula, E.; Garron, M.-L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Res.* **2022**, *50*, D571–D577. [CrossRef]
24. Karlsson, O.P.; Dahlqvist, A.; Vikström, S.; Wieslander, Å. Lipid Dependence and Basic Kinetics of the Purified 1,2-Diacylglycerol 3-Glucosyltransferase from Membranes of *Acholeplasma laidlawii*. *J. Biol. Chem.* **1997**, *272*, 929–936. [CrossRef] [PubMed]
25. Edman, M.; Berg, S.; Storm, P.; Wikström, M.; Vikström, S.; Öhman, A.; Wieslander, Å. Structural Features of Glycosyltransferases Synthesizing Major Bilayer and Nonbilayer-prone Membrane Lipids in *Acholeplasma laidlawii* and *Streptococcus pneumoniae*. *J. Biol. Chem.* **2003**, *278*, 8420–8428. [CrossRef] [PubMed]
26. Sonnhammer, E.L.; Von Heijne, G.; Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. In Proceedings of the International Conference on Intelligent Systems for Molecular Biology, Montréal, QC, Canada, 28 June–1 July 1998; Volume 6, pp. 175–182.
27. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef]
28. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [CrossRef] [PubMed]
29. Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2. *Escherichia coli. Appl. Environ. Microbiol.* **1998**, *64*, 1694–1699. [CrossRef]
30. Malakhov, M.P.; Mattern, M.R.; Malakhova, O.A.; Drinker, M.; Weeks, S.; Butt, T.R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genom.* **2004**, *5*, 75–86. [CrossRef]
31. Thiagarajan, N.; Pham, T.T.K.; Stinson, B.; Sundriyal, A.; Tumbale, P.; Lizotte-Waniewski, M.; Brew, K.; Acharya, K.R. Structure of a metal-independent bacterial glycosyltransferase that catalyzes the synthesis of histo-blood group A antigen. *Sci. Rep.* **2012**, *2*, srep00940. [CrossRef]
32. Andrés, E.; Martínez, N.; Planas, A. Expression and Characterization of a *Mycoplasma genitalium* Glycosyltransferase in Membrane Glycolipid Biosynthesis: Potential Target against *Mycoplasma* Infections. *J. Biol. Chem.* **2011**, *286*, 35367–35379. [CrossRef] [PubMed]
33. Dubots, E.; Audry, M.; Yamaryo, Y.; Bastien, O.; Ohta, H.; Breton, C.; Maréchal, E.; Block, M.A. Activation of the Chloroplast Monogalactosyldiacylglycerol Synthase MGD1 by Phosphatidic Acid and Phosphatidylglycerol. *J. Biol. Chem.* **2010**, *285*, 6003–6011. [CrossRef] [PubMed]
34. Guerin, M.E.; Kordulakova, J.; Schaeffer, F.; Svetlikova, Z.; Buschiazzi, A.; Giganti, D.; Gicquel, B.; Mikusova, K.; Jackson, M.; Alzari, P.M. Molecular Recognition and Interfacial Catalysis by the Essential Phosphatidylinositol Mannosyltransferase PimA from *Mycobacteria*. *J. Biol. Chem.* **2007**, *282*, 20705–20714. [CrossRef] [PubMed]
35. Rodrigo-Unzueta, A.; Ghirardello, M.; Urresti, S.; Dello, I.; Giganti, D.; Anso, I.; Trastoy, B.; Comino, N.; Tera, M.; D'Angelo, C.; et al. Dissecting the Structural and Chemical Determinants of the “Open-to-Closed” Motion in the Mannosyltransferase PimA from *Mycobacteria*. *Biochemistry* **2020**, *59*, 2934–2945. [CrossRef] [PubMed]
36. Kocev, A.; Melamed, J.; Torgov, V.; Danilov, L.; Veselovsky, V.; Brockhausen, I. The wclY gene of *Escherichia coli* serotype O117 encodes an α 1,4-glucosyltransferase with strict acceptor specificity but broad donor specificity. *Glycobiology* **2020**, *30*, 9003–9014. [CrossRef] [PubMed]
37. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555. [CrossRef] [PubMed]

38. Albesa-Jové, D.; Giganti, D.; Jackson, M.; Alzari, P.M.; Guerin, M.E. Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology* **2014**, *24*, 108–124. [[CrossRef](#)] [[PubMed](#)]
39. Berg, S.; Edman, M.; Li, L.; Wikström, M.; Wieslander, Å. Wieslander, Sequence properties of the 1,2-diacylglycerol 3-glucosyltransferase from *Acholeplasma laidlawii* membranes. Recognition of a large group of lipid glycosyltransferases in eubacteria and archaea. *J. Biol. Chem.* **2001**, *276*, 22056–22063. [[CrossRef](#)]
40. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)]
41. Orive-Milla, N.; Delmulle, T.; de Mey, M.; Faijes, M.; Planas, A. Metabolic engineering for glycoacylglycerol production in *E. coli*: Tuning phosphatidic acid and UDP-glucose pathways. *Metab. Eng.* **2020**, *61*, 106–119. [[CrossRef](#)]
42. 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.* **2010**, *31*, 455–461. [[CrossRef](#)] [[PubMed](#)]
43. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)] [[PubMed](#)]
44. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
45. Campanella, J.J.; Bitincka, L.; Smalley, J. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinform.* **2003**, *4*, 1–4. [[CrossRef](#)] [[PubMed](#)]

