Expression and characterization of a human sodium glucose transporter (hSGLT1) in Pichia pastoris

Albert Suades Sala

Universitat Autònoma de Barcelona

Expression and characterization of a human sodium glucose transporter (hSGLT1) in *Pichia pastoris*

TESI DOCTORAL

Albert Suades Sala

2017

Universitat Autònoma de Barcelona

Expression and characterization of a human sodium glucose transporter (hSGLT1) in *Pichia pastoris*

Mèmoria presentada per Albert Suades Sala per optar al grau de doctor.

El treball presentat ha estat dirigit pel Dr.Josep Bartomeu Cladera Cerdà, el Dr. Joan Manyosa Ribatallada i el Dr. Alex Perálvarez Marín i realitzat en la Unitat de Biofísica del Departament de Bioquímica i de Biologia Molecular i al Centre d'Estudis en Biofísica (CEB) de la facultat de Medicina a la Universitat Autònoma de Barcelona

Vist i plau dels directors de la tesi:

Dr. Josep Bartomeu Cladera Dr. Joan Manyosa Ribatallada Dr. Alex Perálvarez Marín

Blessed are the forgetful, for they get the better even of their blunders

Friedrich Nietzsche

Agraïments

En primer lloc, m'agradaria agrair els meus directors de tesi; Joan, Pep i Alex per haver-me donat l'oportunitat i la confiança de realitzar la tesis. Al Joan, per tenir sempre preparada una broma sarcàstica sota la màniga i per animar-me de la forma més efectiva empleada a tot el món; una bona cervesa freda. Al Pep, per estar sempre disponible per qualsevol tipus de problema, qualitat poc habitual en la jefatura de doctorands. Al Alex, per haber-me ajudat a conèixer millor les cares mes interessants i enriquidores de la ciència conjuntament amb les més negres i obscures.

En segon lloc, m'agradaria nombrar a totes les persones que m'han ajudat cientificament en aquest llarg viatje. A la Elena, per a la seva col·laboració amb el projecte. AlRamon, per aguantar les meves consultes pesades de liposomes i els meus crits a l'hora de dinar. Al Josú, per estar sempre disponible a fer-me un cop de mà malgrat la seva infinitat de feina. A Fan-li, a Roger i Asrar per ajudar-me en les coses mes trivials de cultius de llevat. Al Oliver, per donar-me bons consells de laboratori i d'unitat. Al Ernie, per tractar-me amb molta proximitat i per ser una font de coneixement interminable. A Thorsten, per ajudar-me en un dels moments més complicats de la tesis. Al Jeff, per ajudar-me a integrar-me al laboratori al màxim, malgrat que fos molt difícil. A la Lidia, per ajudar-me a fer ITC tot i que fos un fracàs. Als meus primers estudiants de pràctiques i màster, Adrià i Esteban, per haver-me tret feina del damunt. A la Núria de microscòpia per ajudar-me sempre amb un somriure als llavis. Al Fernando, per permetrens l'oportunitat d'anar al sincrotró malgrat totes les dificultats que teníem. A la Bea, per donar-me consells sobre problemes de gels/westerns. A l'Antonio, per la seva atenció i dedicació a realitzar les mesures de Planar lipid membranes. Per últim i molt important, a la Neus, per ajudar-me al laboratori en general des de el principi i, en concret, per ser el millor GPS que un pot demanar. Per últim a la Elodia, per ajudar-me amb la infinitat de problemes tècnics que tenim a la unitat. Segur que em deixo alguna persona que m'ha ajudat i, si està llegint això, que ho dubto, lamento no recordar-la.

En tercer lloc, m'agradaria recordar a tota la gent amb qui, a nivell personal, ha sigut un plaer treballar i passar bons moments tan fora com dins la feina. He vist passar molta gent i molt diferent; però, tots ells, han tingut importància en el seu moment concret. Primer, recordar als membres originaris del "coke" club; Asrar i Fan-li. Vau aconseguir que em rigués com mai degut a les discussions absurdes de ciència, llengua i cultura, sobretot, perqué Fan-li tenia l'habilitat de treure de pollaguera amb gràcia a Asrar. Recordar també als altres membres no originaris del "coke" club: Ping i Li-ying, com diria Asrar; la triada xina. Tots ells sempre es van mostrar molt hospitalaris amb mi i n'estaré sempre agrait: shēng rì kuài lè! Fer un esment final cap a Asrar, ja que va ser el meu primer professor d'angles particular que no em cobrava i amb qui compartíem molt més del que als dos ens agradaria reconéixer, I still miss your dark humor and the pointless discussions we had! Recordar també a tots els estudiants que han passat per aquí deixant emprenta. Al Adrià, al meu primer minion amb qui estava més còmode que un upgraded zealot

matant zerlings i, al Esteban, pels bons partits de squash i per ser molt divertit i entranyable. A Erika, per permetrem conèixer Roma i per les sessions de cine clandestines amb Asrar. A Lucia, per les interesants i divertides discussions sense haver de controlar el "tono" malgrat que no ens entenguéssim ni a la de tres.

Recordar també altres persones, com la Neus, per escoltar les meves desgràcies i aguantar la meva cara de malhumorat i, inclús així, esperar-me per anar a esmorzar quan feia el mandrós. A Kendrick i Nabill, thank you guys for making my stay in USA easier than I thought it was going to be. A la Bea, per poder parlar d'escatologia sense sentir-me com un alien fora d'aquest planeta. A la Maribel que, tot i que costes temps conèixe'ns, malgrat les circumstàncies, vam acabar entenent-nos molt més del que hagués imaginat. A la txell, per tenir aquest caràcter tan alegre llastima per això que "quiera romper España!" A Willy Toledo, per no jutjarme pels meus comentaris de retard al principi. A Jordi, el diplomàtic, gràcies per permetren's conèixer la part mes "cultural" de Terrasa. A Tzvetana, per ser una persona molt amable i encantadora i tractar a tothom correctament independent de qui fos. Per últim i amb més importància: Aida i Josu. A Josu, per trencar tots els esquemes i estereotips que tenia i per ser, junt amb Maribel, l'ùnic company de principi a fi amb qui sempre podria parlar de qualsevol cosa: problemes personals, jocs, series, animes, pelis i, com no, de ties. Josu, sempre expandiré els conceptes de "ginter" "fordo/a" i els petits consells que semblaven absurd però no ho eren, com el del modo "commando". A Aida, per la seva bogeria innata, el seu humor tan ridículament idiota (i tan atípic en el gènere femini) i pels bons moments a Ventalló. Ahora te toca a ti sola ser la buffon oficial de la unidad. Probablement, podria dir moltes més coses de les persones anomenades però els qui han sigut importants ja ho saben.

En quart lloc i, per últim, als meus pares i al meu germà. Al meu germà per compartir el "freakisme" en les venes malgrat que no estigués físicament present. A la meva mare, per aguantar-me i estimar-me quan de pitjor humor estava malgrat que som molt diferents i iguals a la vegada. Al meu pare, màrtir, mentor, amic i pare a la vegada. Amb ell podia parlar de ciència en general, de problemes de doctorands, de futur, de cine, política i, malgrat que em pogués treure de pollaguera, ell sempre aguantava el tipus, fet que encara, a dia d'avui, s'em fa insòlit, gràcies.

Abbreviations

2-NBDG: 2-(N-(7-Nitrobenz-2-oxa-1,3-diazol-4-yl)Amino)-2-Deoxyglucose)).

AOX1: Alcohol oxidase I.

ATP: Adenosine triphosphate.

BB: Breaking Buffer.

BGY: Buffered glycerol complex medium.

BMG: Buffered minimal glycerol medium.

BMM: Buffered minimal methanol medium.

BMY: Buffered methanol complex medium.

BSA: Bovine serum albumin.

CHEMS: Cholesterol hemisuccinate.

CIP: Alkaline phosphatase.

CMC: Critical micelle concentration.

CV: Colum Volume.

DDM: *N*-Dodecyl β-D-maltoside.

DLS: Dynamic light scattering.

DM: Lauryl-beta-D-maltoside.

DPHPC: 1,2-diphytanoyl-sn-glycero-3-phosphocholine.

DTT: Dithiothreitol.

E.coli*: Escherichia coli.*

EDTA: Ethylenediaminetetraacetic acid.

E: Elution step.

F12: Foscholine 12.

FLAG: Peptide sequence DYKDDDDK

FSEC: Fluorescence size exclusion chromatography.

FT-IR: Fourier transform infrared spectroscopy.

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase.

GGM: Glucose-galactose-malabsorption.

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

HPLC: High pressure liquid chromatography.

hSGLT1: Human sodium glucose cotransporter 1.

IMAC: Immobilized metal ion affinity chromatography.

KCl: Potassium chloride

LDAO: Lauryldimethylamine N-oxide.

MCS: Multiclonning site.

MD: Minimal dextrose medium.

MG: Minimal glycerol medium.

MM: Minimal methanol medium.

Ni-NTA: Niquel-nitrilotriacetic acid.

NaCl: Sodium chloride.

OD: Optical density.

OG: *N*-octyl glucoside.

ORT: Oral rehydration therapy.

PAGE: Polyacrylamide Gel Electrophoresis

PBS: Potassium buffered sodium.

PC: Phosphatidylcholine.

PCR: Polymerase chain reaction.

P. pastoris: Pichia pastoris.

PS: Phosphatidylserine.

RT: Room temperature.

S. cerevisae: *Saccharomyces cerevisae.*

SDS: Sodium dodecyl sulfate.

SEC: Size exclusion chromatography.

SLC: Solute carriers.

SM: Starting material.

SUV: Small-Unilamellar Vesicles.

TAE: Acetic tris buffer with EDTA.

TE buffer: Tris EDTA buffer.

Teflon: Polytetrafluoroethylene.

TGKCL: Potassium chloride Tris buffer with glycerol.

TKCL: Potassium chloride Tris buffer.

TTBS: Tris saline buffer with tween.

TX-100: Triton X-100.

VDAC: Voltage dependent anion cannel.

vSGLT: *Vibrio parahaemolyticus* Sodium Glucose cotransporter.

W: Wash step.

WT: Wild type.

YPD: Yeast extract peptone dextrose media.

Index

1. Introduction

1.1 Proteins: The life machines

Proteins are present in all organisms functioning as a big interactive network capable of performing many different roles and functions towards making cell life possible.

All proteins are composed of the amino acid polypeptide chain, which variety arises different size, structure and chemistry properties. Considering the cell as a binary system of hydrophilic and hydrophobic compartments, there are two kinds of proteins: soluble and membrane proteins. Soluble proteins amino acid composition defines exposed hydrophilic surfaces, and therefore are soluble in water. These proteins are mainly present in water soluble environments, such as the extracellular space, the cytoplasm, or the inner space of organelles. Membrane proteins, which are not soluble in water because the amino acid composition, define hydrophobic exposed surfaces that need to be balanced by hydrophobic environments, such as cellular membranes, or non-polar solvents. Although membrane proteins are less than one third of the total proteome (**1**), membrane proteins account for key functions, such as transport across membrane, energy and signal transduction processes, cell-cell recognition, etc. Thus, membrane proteins are the first therapeutic target for pharmacological intervention, and so far, 60% of present drug targets are membrane proteins (**2**).

The greater goal for protein structural biologist is to elucidate the structure-function dynamics of proteins with as much detail as possible, being this one of the greatest challenges because the many hindrances present in the structural biology pathway. Regarding membrane proteins the challenge is greater, because the need to find optimal conditions to keep these proteins soluble and avoiding aggregation. In addition, membrane protein expression levels in native systems are minimal, which makes mandatory the development and implementation of membrane proteins expression and purification systems (**3**).

1.2 Membrane Proteins

The cell membrane is a fluid mosaic mostly composed of proteins and lipids. The lipid component accounts for the lipid bilayer, which is built up of lipids, amphipathic molecules with a hydrophobic tail and a polar head. This lipid bilayer spontaneously forms in aqueous solution, providing a suitable environment for membrane proteins. A wide number of membrane proteins are expressed in the cell and are distributed heterogeneously, and this distribution is multifactorial (cell type, protein half-life, protein regulation, etc.). In the human genome, it is estimated that membrane proteins comprise around 15-39% of all coded proteins, but those values vary

depending on the organism, for example *S. cerevisae* membrane proteome is around 28% (**5**). This data probes the importance of membrane proteins in all cell organisms from human to lower organisms.

Membrane proteins can be categorized into 3 main groups: lipid-anchored proteins, peripheral membrane proteins, and integral membrane proteins (**4**). Lipid-anchored proteins are proteins that in close contact with the lipid bilayer interface through a lipid modification of the protein. Peripheral membrane proteins are amphipathic proteins interacting with the surface of the membrane due to hydrophobic and electrostatic patches. These proteins are not inserted in the membrane and the interaction can be transient. Integral membrane proteins are proteins that are embedded in the hydrophobic region of the lipid bilayer. Depending on the type of insertion and the number of segments of the protein spanning across the membrane, there are monotopic, bitopic, and polytopic membrane proteins (figure 1.2). Integral membrane will be the focus of our work.

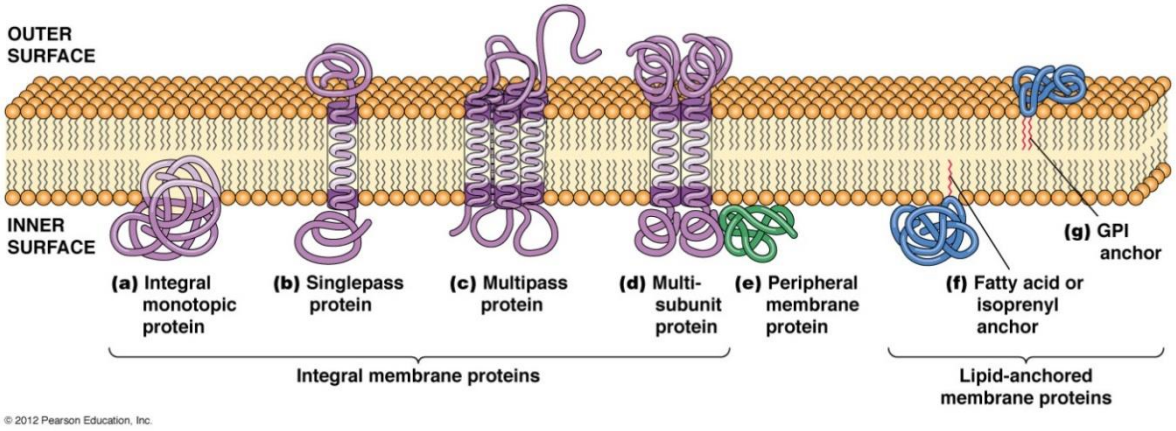


Figure 1.2: Representation of the different types of membrane protein. Majority of the membrane proteins are found in nature in the integral membrane protein group (**6**).

Cells need to adapt to a constantly changing environment in order to survive; membrane proteins are the mechanism of the cell to communicate with the extracellular environment (**7**). In eukaryotic cells, compartmentalization of the inner space in specialized organelles and nucleus, adds a new perspective to biological membranes, and to membrane proteins as a mechanism of communication between these compartments. Membrane proteins functional diversity allows the cell to respond to physic-chemical stimuli, but also to obtain resources and/or energy from the environment, where membrane transporters are the key. Membrane protein transporters are present in all domains of life and there are many tasks for membrane transporters such as: transport of carbon, nitrogen and sulfur sources; the regulation of catabolic and anabolic substances; secretion and recovery of drugs; immunity response to pathogens; transport of bigger molecules like neurotransmitters, hormones, antibodies or smaller molecules like ions.

1.3 Membrane Transport

The lipid bilayer of the cell membrane is permeable to some small molecules, mainly gases, like oxygen (O_2), Carbon dioxide (CO_2), Nitric Oxide (NO). These molecules can cross this barrier by simple diffusion because of their nature and, at some extent, other molecules, like water molecules, ethanol and Urea can also cross the membrane (**6**). Highly hydrophobic molecules like Vitamin D can cross the membrane by simple diffusion or stay in within like cholesterol. But the lipid bilayer is impermeable to most essential molecules in the surrounding extracellular fluid including ions (Na^+, K^+, Ca^{+2}, Cl^-), small hydrophilic molecules (Hexoses) and macromolecules (RNA and proteins). These molecules cannot cross the membrane by simple diffusion, creating the need for the membrane transport function (Fig 1.2).

Transport across membranes can be either passive or active, depending whether the transport is carried out downhill or uphill. Passive transport is a downhill type of transport mostly driven by the second law of thermodynamics, where the system moves in one direction or another trying to reach electrochemical equilibrium. Passive transport can be through simple diffusion, passive diffusion (channel) or facilitated diffusion (uniporter, Fig 1.3).

Active transport can be primary or secondary but in both cases energy expenditure is required. In primary active transport, adenosine triphosphate (ATP) is hydrolyzed by the protein performing transport in order to provide the free energy needed to transport against an electrochemical gradient. In secondary active transport, there is no direct hydrolysis of ATP, instead it relies upon the electrochemical potential difference created by pumping ions in/out of the cell (typically, Na^+ or H^+) (**8**). This transport couples the movement of the driving ion down its electrochemical gradient to the movement of another molecule. If the driving ion and the driven molecule move in or out in the same direction, the transport is referred as cotransporter or symporter, but, if the driving ion and the molecules move in opposite directions, transport is referred as antiporter (**9**) (Fig 1.3). Membrane cotransport of glucose will be the main focus of this thesis.

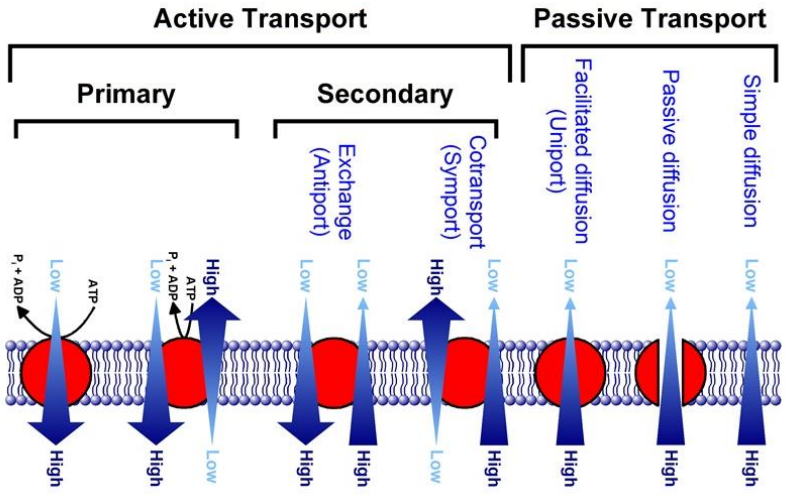


Figure 1.3: Representation of the different types of transporters in Nature with active and passive transporters represented (**6**).

1.4 Solute carrier (SLC) families of transporters

All human activities require glucose to provide energy. However, it is very important to maintain the blood glucose concentrations in the vascular system in a specific range, for example, low blood glucose concentrations in the brain can cause seizures, loss of consciousness and death but high glucose concentrations may cause diabetes and, as a consequence, renal failure, neuropathy, blindness and others. Glucose homeostasis in the vascular system is critical and this is where the glucose transporters have a very important role to play. The most well-known glucose transporters are the GLUT family but there are many other glucose transporters and all of them are members of the same genetic family: the solute carriers (SLC).

The SLC is a huge genetic family that includes a great number of transporters including passive transporters, symporters, antiporters and others (**10,11**), but we are interested in the glucose transporters (SLC2, SLC5, SLC50) (**15**). SLC2 is a family of facilitated transporters (GLUT); SLC5 is a family of active transporters (Co-transporters) and finally SLC50 which is a new group of uniporters (SWEET). All members of the SLC family share sequence similarity. A transporter is considered a member of the SLC family if it has at least 20-25% amino acid sequence identity to other member of that family (**12**).

	The HUGO SLC Family Series Total(REF to HUGO)	
SLC1	The high-affinity glutamate and neutral amino acid transporter family	7
SLC2	The facilitative GLUT transporter family	14
SLC3	The heavy subunits of the heteromeric amino acid transporters	2
SLC4	The bicarbonate transporter family	11
SLC5	The Na+/glucose cotransporter family	12
SLC6	The sodium- and chloride-dependent neurotransmitter transporter family	20
SLC7	The cationic amino acid transporter/glycoprotein-associated amino-acid transporter family	14
SLC8	The Na+/Ca2+ exchanger family	3
SLC9	The Na+/H+ exchanger family	13
SLC10	The sodium bile salt cotransport family	7
SLC11	The proton-coupled metal ion transporter family	2
SLC12	The electroneutral cation-Cl cotransporter family	9
SLC13	The human Na+-sulfate/carboxylate cotransporter family	5
SLC14	The urea transporter family	2
SLC15	The proton oligopeptide cotransporter family	4
SLC16	The monocarboxylate transporter family	14

SLC17	The vesicular glutamate transporter family	9
SLC18	The vesicular amine transporter family	3
SLC19	The folate/thiamine transporter family	3
SLC20	The type-III Na+-phosphate cotransporter family	2
SLC21	The organic anion transporting family	11
SLC22	The organic cation/anion/zwitterion transporter family	31
SLC23	The Na+-dependent ascorbic acid transporter family	4
SLC24	The Na+/(Ca2+-K+) exchanger family	6
SLC25	The mitochondrial carrier family	46
SLC26	The multifunctional anion exchanger family	11
SLC27	The fatty acid transport protein family	6
SLC28	The Na+-coupled nucleoside transport family	3
SLC29	The facilitative nucleoside transporter family	4
SLC30	The zinc efflux family	10
SLC31	The copper transporter family	2
SLC32	The vesicular inhibitory amino acid transporter family	1
SLC33	The acetyl-CoA transporter family	1
SLC34	The type-II Na+-phosphate cotransporter family	3
SLC35	The nucleoside-sugar transporter family	28
SLC36	The proton-coupled amino acid transporter family	4
SLC37	The sugar-phosphate/phosphate exchanger family	4
SLC38	The System A and N, sodium-coupled neutral amino acid transporter family	11
SLC39	The metal ion transporter family	14
SLC40	The basolateral iron transporter family	1
SLC41	The MgtE-like magnesium transporter family	3
SLC42	The Rh ammonium transporter family	3
SLC43	The Na+-independent, system-L-like amino acid transporter family	3
SLC44	The choline-like transporter	5
SLC45	The putative sugar transporter	4
SLC46	The folate transpoter	3
SLC47	The multidrug and toxin extrusion	2
SLC48	The heme transporter	1
SLC49	The transporters of the major facilitator superfamily	1
SLC50	The sugar efflux transporters of the SWEET family	1
SLC51	The transporters of steroid-derived molecules	1
SLC52	The riboflavin transpoter family RFVT/SLCt2	1
	TOTAL	385

In general the genes are named using the root symbol SLC, followed by a number (e.g., SLC1 solute carrier family 1), the letter A (which acts as a divider between the numerals) and finally the number of the individual transporter (E.g., SLC5A1, sodium/glucose cotransporter protein). These general rules of SLC gene nomenclature have been elaborated further for a couple of families where they include other letters a part from A. Those letters are used only to specify subfamilies in the same group for example, SLC35 (A-G) (**13**). Also, SLCA21 has a special nomenclature where the "A" and the "21" have been replaced by "O" (organic transporter) due to rapid evolution giving rise to new isoforms within a given species (**14**).

It's interesting to mention that in the recent years a lot of new genes were found out and new families were added, for instance, 10 years ago the total number of proteins within this group was 285 instead of 385.

1.5 Sodium/Glucose Cotransporter: SLC5 and hSGLT1 (SLC5A1)

The sodium/glucose cotransporter family, SLC5, has 220 members in animal and bacterial cells (**16**). The founder of this family is the sodium glucose cotransporter, SGLT1, which is encoded by the SLC5A1 gene and its function is to transport glucose and other hexoses from the lumen to the enterocytes. Also, SGLT1 protein is linked to diseases like glucose-galactose-malabsorption (GGM.) and therapies like Oral Rehydration Therapy (ORT.). Although SGLT1 is going to be our protein of interest, this family comprehends other glucose cotransporters as well, like the SGLT2, encoded by the SLC5A2 gene, which is responsible for glucose reabsorption in the kidney (**20**). It's important to note that, other members of this family have affinities for a wide range of other substrates than hexoses (**23**).

The SLC5 pool of genes in all kingdoms of life is huge, and out of all of them, 12 are human genes expressed in different tissues, such as brain, small intestine, kidney muscle and thyroid gland. These 12 proteins present a high sequence identity/similarity (21-70%) but gene structure is diverse. Out of the 12 genes, 8 of them have their coding sequence in 14-15 exons (SLC5A1-2, SLC5A4-6, and SLC5A9-11). The coding sequences for SLC5A7 (CHT) and SLC5A3 (SMIT) are contained in 8 and 1 exons respectively (**18**).An unrooted phylogenetic tree of 12 human members of the SLC5 familty of cotransporters of known function is shown in figure 1.5.1.The rabbit form of sodium glucose cotransporter was the first one from SGLT1 to be cloned and identify with the use of, at that time, a new expression cloning method (**17**).

From all SGLT1 orthologs, the human protein (hSGLT1) has been subjected to research for drug design and therapy reasons. hSGLT1 is a 664 amino acid protein with 14 transmembrane spans with both N and C termini facing the extracellular fluid (**19**) as represented in the scheme in figure 1.5.2. Nevertheless, this topology was not so clear in early years, when the transmembrane

prediction was 11 transmembrane segments, and it was not clear if the C terminus was intracellular or extracellular (**24**). Later on, some other predictions suggested 12 or 13 transmembrane domains (**25**). Although in the end, the number of helix was solved with a glycosylation studied which concluded, empirically, the number of transmembrane spans of hSGLT1 is 14 (**26**).

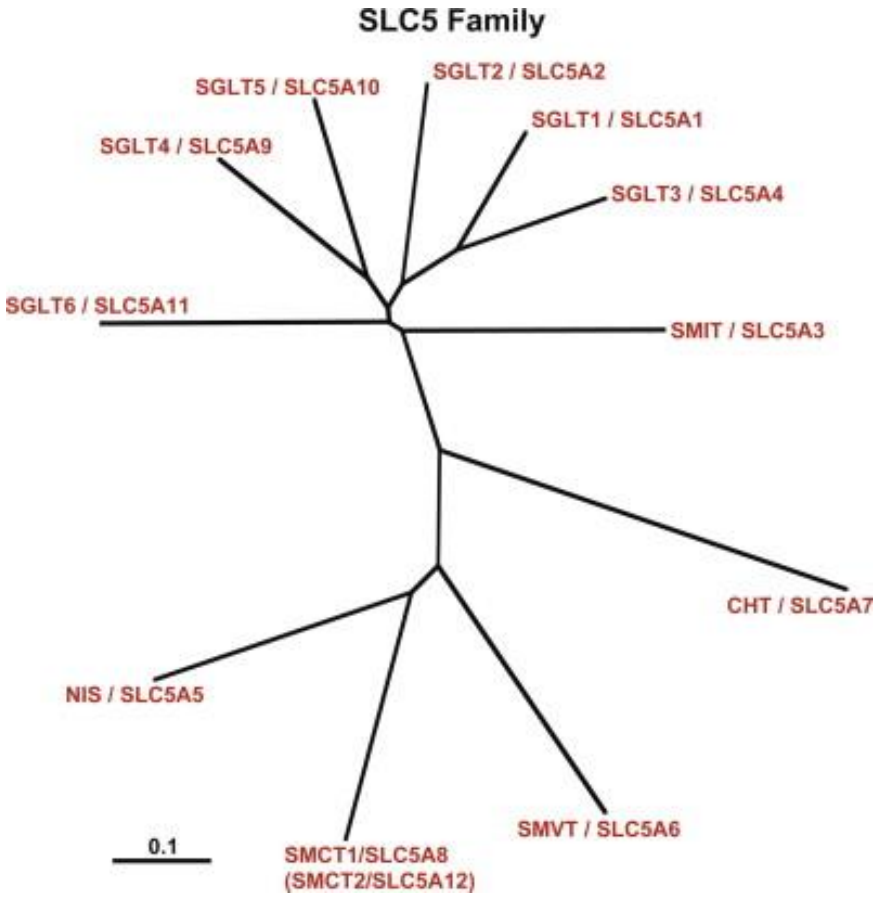


Figure 1.5.1: Unrooted phylogenetic tree of the twelve human members of the SLC5 family of cotransporters (**15**)**.**

The uncertainty in the determination of the actual number of transmembrane spans lies in the fact that between helices 13 and 14 there is a large loop. Studies working with different chimeras of hSGLT1 and hSGLT2 revealed that the substrate recognition region is in a distal protein position, from amino acid 380 to 664 (**27, 28**). Therefore, the large loop between helices 13 and 14 might be linked to the mechanism of protein function in substrate (sugars) recognition, or, the movement of the loop is critical for protein function. Although it is still under debate the role for the loop, several studies suggest that this loop translocates when substrate is added. Upon substrate binding, the loop internalizes into the membrane causing a portion of the protein to face the extracellular side (**29, 30**).

Regarding the sodium binding, hSGLT1 has been extensively studied in *Xenopus laevis* oocytes using electrophysiology (**36**). These studies revealed that the sodium binding site is close to the N terminus instead of C terminus (**37**).

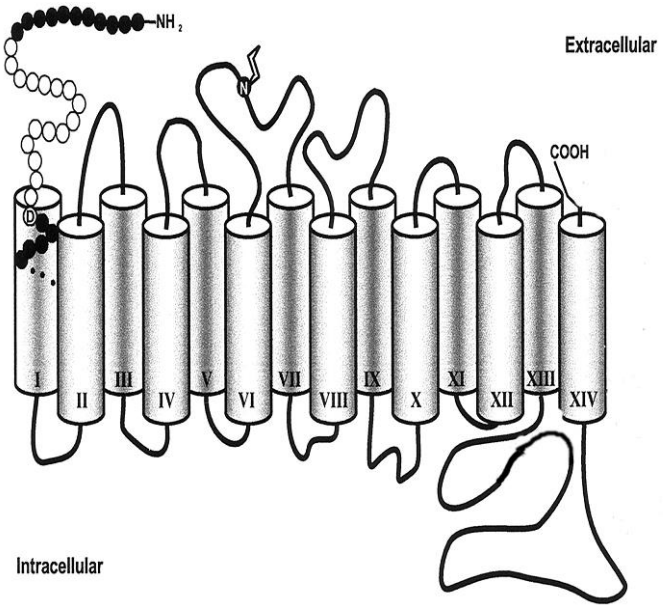


Figure 1.5.2: Tentative secondary structure model of hSGLT1. hSGLT1 consists of 14 α-helical transmembrane domains (shaded columns) with the N-and C-terminal located on the extracellular phase of membrane. The N-glycosylation site (Asn-248; N) is indicated (**32**).

High resolution tridimensional structure for hSGLT1 is still missing. The main reason why the structure for hSGLT1 has not been solved yet is due to lack of an expression system suitable to produce high amounts of pure protein in a suitable recombinant expression system (**31, 32**). So far, only an isoform from *Vibrio parahaemolyticus* SGLT1 (vSGLT1) has been crystallized successfully (**33**). The crystal structure of vSGLT1 has proven itself very useful when studying hSGLT1 since both are homologous proteins and share an identity of sequence of 32% and 75% similarity between the amino acid sequences (**34**). This similarity and sequence identity is pretty high for a membrane protein so it gives a hint on how the protein might be. The main difference between hSGLT1 and vSGLT1 resides in the loop between helices 13 and 14. Several laboratories have modeled the structure of hSGLT1 based on the vSGLT1 model, which allowed gathering and obtaining data regarding the structure and function of the protein (**35, 42**). Some key amino acids have been identified regarding the binding of sodium and glucose. The amino acids involved in sugar recognition are H83, E102, T460, D454 and Q457 (**35, 39, 41, 42**) which is highly conserved in different types of SGLT (**38**). While the amino acids related to sodium binding are A76, I79, S389, S392 and S393 (**35, 138**).

The exact model mechanism of the protein is still not clear but a 6-state ordered kinetic model has been established (**43, 44, 45**). Kinetic studies found out some of the proposed transitional states of

the protein which supported the already established 6-state model (**35**). The 6-state ordered kinetic model describes sodium/glucose cotransport as a series of ligand-induced conformational changes. On the external surface of the membrane, two Na+ ions bind to the transporter (**35, 42**) before sugar. The empty transporter C (states 1 and 6), Na+-loaded transporter CNa2 (states 2 and 5), and the sugar-loaded transporter CNa2S (states 3 and 4) can "cross" the membrane. Membrane voltage affects Na+-binding and translocation of the empty transporter. A representation of the different states is represented in figure 1.5.3.

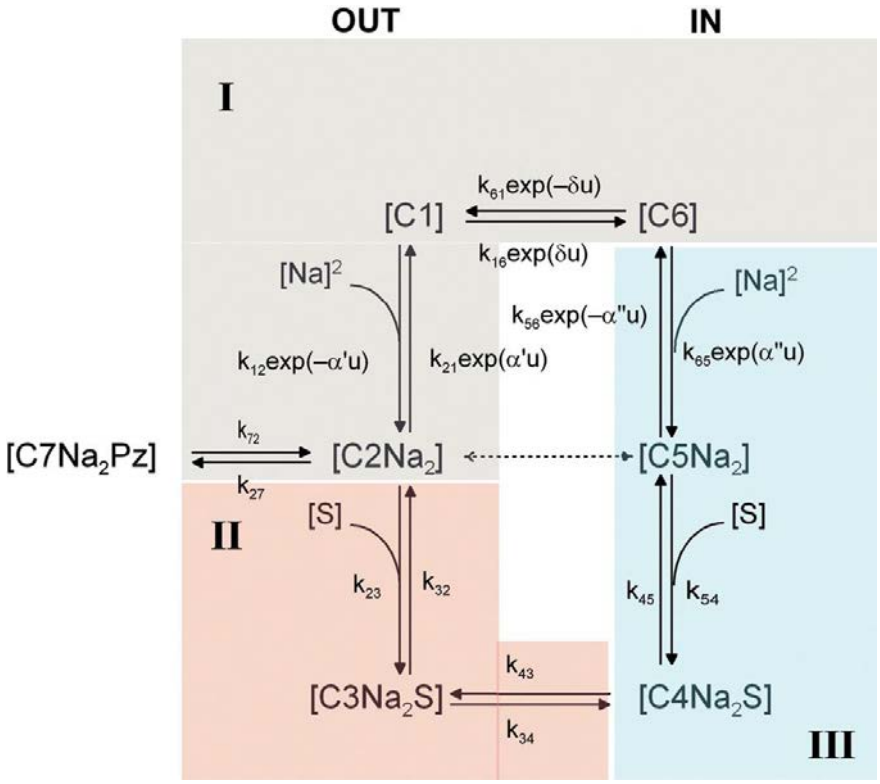


Figure 1.5.3: Six-state ordered kinetic model for Na-glucose cotransporter. Scheme showing the six kinetic states and the rate constants for the transitions between them. In the absence of ligands, the transporter exists in two states (C1 and C6). At the external surface, two Na binds to the transporter to form the complex $C2Na_2$. The sugar-loaded transporter (C3Na2S) undergoes a conformational change ($C3Na_2S$ to $C4Na_2S$) resulting in Na-glucose cotransport. The reaction from state $C2Na_2$ to $C5Na_2$ (dashed line) is the uniport mode of Na transport by SGLT1. Pre-steady-state currents are associated with the partial reactions $C2Na_2$to C1 to C6. Region I (gray) is the pre-steady-state currents. II (red) includes external sugar binding/dissociation ($C2Na_2$ to $C3Na_2S$) and sugar translocation across the cell membrane ($C3Na_2S$ to $C4Na_2S$); and III (blue) are the internal Na_ release ($C5Na_2$ to C6) and sugar release ($C4Na_2S$ to $C5Na_2$) steps. $C2Na_2$ to $C7Na_2Pz$ refers to the blocked state of the protein after adding the inhibitor (**34**).

Besides the already mentioned amino acids related to binding of sodium and hexoses, other residues are important. Some of those residues are involved in the outer-gating and inner-gating. This gating mechanism has been observed already in the vSGLT1 (**40, 46**). For example, three critical residues were described for the outer gate: L87, F101 and F453 (**35**). Those three residues act as a triad for glucose gating. On the other hand, only one residue was described as crucial for

the inner gate: Y290 (**35, 42**). This tyrosine residue can rotate allowing the substrate release. This mechanism was already described in vSGLT and other family members of SGLT (**42, 47**). Also, there are other known important residues of hSGLT1. For example, N248 is the only described residue to be glycosylated (**48**) by N-glycosylation. No other evidence supporting posttranslational modifications like O-glycosylation was found (**49**). Surprisingly, this N-glycosylation is not necessary for the protein to function properly (**48**).

One interesting characteristic of hSGLT1 is that functions as a monomer just like vSGLT1 although the crystal structure of vSGLT1 is actually a dimer (**36, 50**). This conclusion was based on the cross-sectional area of the proteins within the plasma membrane analyzed by freeze-fracture electron microscopy in oocyte membrane. The analysis revealed that the hSGLT1 in those oocytes was a monomer (**33**). Nevertheless, some discrepancy is still under debate because a radiation inactivation analysis of transport in rabbit brush-border membranes suggested that SGLT1 was actually a homotetramer (**51**). Overall, it looks like hSGLT1 function as a monomer but it can dimerize in some condition but without any evidence of the physiological implications of this process.

1.6 *Pichia pastoris* as an expression host for membrane proteins

Pichia pastoris is and efficient system for the expression of heterologous genes. Precisely, some functional membrane proteins have been successfully expressed, such as Na+, K+-ATPase (**52**), multidrug resistance protein1 (**53**), ATP-binding cassette transporter (**54**), NOP-1 (**55**), mammalian intestinal peptide transporter (**56**), human μ-opioid receptor (**57**) and 5HT5a-serotoningeric receptor (**58**).

Several factors have contributed to the rapid acceptance of *P. pastoris* as a regular system for membrane protein heterologous expression, the most general important features include:

1. A promoter derived from the alcohol oxidase I (AOX1) gene of *P. pastoris* that is uniquely suited for the controlled expression of foreign genes. Constitutive promoters like, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (**56**), can also be used instead of non-constitutive like AOX1.
2. Ease to translate the yeast molecular biology methods from the model system *S. cerevisiae* to *P. pastoris.*
3. Strong preference of *P. pastoris* for respiratory growth instead of fermentation, a key physiological trait that greatly facilitates its culturing at high cell densities.
4. A 1993 decision by Phillips Petroleum Company to release the *P. pastoris* expression system to academic research laboratories, the consequences of which has been an explosion in the knowledge base of the system

As a budding yeast, *P. pastoris* is a single-celled microorganism that is almost as easy to manipulate and to culture as *E. coli*. In addition it is a eukaryote, thus capable of many of the post-translational modifications performed by higher eukaryotic cells, such as proteolytic processing, protein folding, disulfide bond formation, and N- and O-glycosylation. Most of the proteins that end up as inactive inclusion bodies in *E. coli* are produced as biologically active molecules in *P. pastoris*. The *P. pastoris* system is also generally regarded as being faster, easier and less expensive to use than expression systems derived from higher eukaryotes, such as insect and mammalian tissue culture cell systems, and if it works, higher expression levels are achieved, the limiting step in structural biology (**62**).

P. pastoris also serves as a useful model system to investigate certain areas of modern cell biology, such as the molecular mechanism involved in import and assembly of peroxisomes, the selective autophagic degradation of peroxisomes and the organization and function of the secretory pathway in eukaryotes.

More detailed information on the *P. pastoris* system can be found in the numerous reviews describing the heterologous protein expression system and the Pichia Expression Kit manual from Invitrogen or in the Invitrogen website (**59, 60, 61**)**.**

1.7 General perspective: Interest of study

Human membrane transporters are related to a huge number of important functions and are linked to some severe diseases when the primary sequence of the protein is mutated which causes the protein to malfunction. Human sodium glucose transporter (hSGLT) is no exception and it is linked to: diabetes, GGM, Alzheimer's and others. (**20**) Therefore, all information regarding the function and mechanism of transport of the protein is of special interest. The fully elucidate the mechanism of transport; it is mandatory to known the structure of the protein which requires, in general, a good diffraction crystal. Nevertheless, only few membrane transporters have been crystalized and they structure and mechanism is well understood, for example: MelB, LacY, vSGLT1 and Glut1 (**21, 22, 36**). From those three transporters, only Glut1 is a glucose human transporter because the rest are prokaryote transporters so, it's clear that there is a lack of information in human glucose membrane transporters regarding them structure and mechanism.

Nowadays some drugs are actually targeting SGLT2 and are under clinical trials which clearly indicate the relevance and importance of those transporters. As mention, it's difficult to express a human recombinant protein in high amounts necessary to perform structural studies like crystallization and others. Normally human membrane proteins present hindrances when are being expressed in a non-eukaryotic expression system, therefore, a good expression system must be found. hSGLT1 was described once to be expressed in *P. pastoris* successfully (**63**) which is clearly a good starting point for our project.

2-Objectives

Secondary transporters are membrane proteins which activity is of paramount importance for the living cell by making possible, for example, the uphill transport of nutrients by coupling it to the dissipation of ionic electrochemical gradients. Some of these secondary transporters, and in particular, some related to the transport of sugars in prokaryotes (LacY and Mel B from *E. Coli* or *S. tiphimurium*; vSGLT from *Vibrium*) have been extensively produced as recombinant proteins and structurally characterized, including the resolution of their molecular structure by X-ray crystallography.

An important human version of these sugar transporters is the glucose transporter hSGLT. It is a well-known membrane protein expressed in the human gut (enterocytes) and it is medically relevant for its function during the re-hydration treatments against cholera. Besides, some of its isoforms are associated to pathologies such as diabetes. The structural study of these human secondary transporters is, however hampered by the difficulties in producing them in amounts high enough for the application of suitable biophysical techniques. Therefore, this has been the main goal of the present doctoral project, the expression and purification of hSGLT1 in an eucharyotic system (*Pichia pastoris*) for its structural and functional characterization. Our starting point has been the only work in the bibliography where it had been reported that the expression of hSGLT1 in Pichia was possible. From there, this general objective was designed to be accomplished via the following specific objectives:

1- To optimize hSGLT1 clonning, transformation and expression using the construct hSGLT1+eGFP.
2- To optimize hSGLT1 solubilization from the membrane: detergent screening.
3- To improve hSGLT1 purification.
4- To reconstitute hSGLT1 in model membranes (liposomes).
5- To characterize hSGLT1 structure by using CD and FT-IR spectroscopy techniques.
6- To characterize the function of protein: Transport assay.

3-Materials and methods

Protocols were adapted from the EasySelect *Pichia* Expression Kit (**64**) unless it's mentioned.

3.1 Materials

3.1.1 General equipment

Autoclave:	Stericlav-75, Raypa Trade (Barcelona, Spain)
Bead Beater:	Bead beater, serial nº 1107900-101, Biospec products
Centrifuges:	Heraeus multifuge X3R, Thermo Scientific
	Scanspeed 1236R, Labo Gene (Barcelona, Spain)
	Discovery M150, Sorvall-Thermo Scientific
	Combi Plus, Sorvall-Thermo Scientific
Chromatography System:	ÄKTA Purifier System, GE Healthcare (Chicago, IL)
Chromatography Columns:	Superdex 200, HR 10/30, Superose 6, 10/300 Gl, GE Healthcare (Chicago, IL)
Dynamic Light Scattering:	Microtac UPA 900, model nº: 92-23510-002-U1075, Microtrac
Electroporation apparatuses:	Mini PROTEAN tetra cell and Power PAC 300 (Hercules, CA)
Freezer:	Isotemp Freezer 00502, Fisher Scientific
Gel's camera:	Fisher Bioblock Scientific, Serial nº 07-10604
Homogenizer:	Electronic type 790, Metabo (Barcelona, Spain)
Incubators:	MaxQ4000, model nº: SHKA4000, Thermo Scientific
Infrared spectrophotometer:	Varian 7000e FT-IR Spectrometer, Varian Medical Systems (California, USA)
pH meter:	Basic 20 pH, model nº 210291, Crison (Barcelona, Spain)
Plate reader:	FLUOstar optima, serial nº 413-3670, BMG Labtech (Ortenberg, Germany)
Rotary Evaporator	Heidolph (Barcelona, Spain)
Scales:	BP 3100 P, Sartorius (Göttingen, Germany)
Semi-dry blotter	CamLab (Cambridge, England)
Sonicator:	Sonic dismembrator, Dynatech (Zaragoza, Spain)
Spectrophotometer:	Cary 50 Bio UV-Vis spectrophotometer, Varian Medical Systems (California, USA)

Thermocycler:	Doppio Thermocycler, Serial nº VWRI732-1210, VWR International bvba (Barcelona, Spain)
Transilluminator:	Dark reader transilluminator, DR-46B, Clare Chemical Research
Vortex:	Vortex mixer, serial nº VB3B011224, Licuos essential instruments
Water baths:	Tectron Bio, J.P Selecta (Barcelona, Spain)

3.1.2 Chemical reagents

Acrylamide/Bis 37.5:1; 40% (w/v) Solution	Bio-Rad
Antibody anti hSGLT1-rabbit AP00343PU-N	Acris Antibodies GmbH
Antibody anti rabbit-HRP P0449	Agilent Technologies
Amicon Ultra-15	Millipore
Agar	Conda pronadisa
Agarose	Sigma-Aldrich
Antiflag M2	Sigma-Aldrich
Ammonium persulphate	Melford
Asolectin from soy bean	Fluka, 1145
Biotin	Sigma-Aldrich
Bio beads	Bio-Rad
Bromophenol blue	Sigma-Aldrich
BSA (bovine serum albumin)	Melford
Cholesterol	Sigma-Aldrich
Chloroform	Merck
CHS	Sigma-Aldrich
Dideoxynucleotides	New England Biolabs
n-Decyl β-D-maltoside (DM)	Anatrace
n-Dodecyl β-D-maltoside (DDM)	Anatrace
DMSO	Sigma-Aldrich
DTT	Sigma-Aldrich
EDTA	Merck
Ethanol	Merck
Ethidium Bromide	Bio-Rad
FosCholine-12	Anatrace
Glass beads	BioSpec
D-[6-3H] Glucose	Sigma-Aldrich
D-Glucose	Melford
Glycerol	Melford
Guanidinium Chloride	Melford
HEPES	Melford
HisTrap HP 1mL	GE Healthcare
HisTrap HP 5mL	GE Healthcare
Isoamyl alcohol	Merck
Isopropanol	Merck
LDAO (Lauryldimethylamine-N-Oxide)	Anatrace
LMNG (2,2-didecylpropane-1,3-bis-β-D-maltopyranoside)	Anatrace
Methanol	Merck

2-β-Mercaptoethanol	Sigma-Aldrich
Methyl-α-D-[14C] glucopyranoside (α-MDG)	Sigma-Aldrich
2-NBDG (Deoxy-2[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose)	Thermo fisher scientific
NdeI	New England Biolabs
NotI	New England Biolabs
OG (n-Octyl-β-D-Glucoside)	Anatrace
Peptone	OXOID
Potassium phosphate dibasic K_2HPO_4	Fluka
Potassium phosphate monobasic KH_2PO_4	Melford
Phlorizin	Sigma-Aldrich
Potassium Chloride (NaCl)	Melford
PMEI	New England Biolabs
SDS	Sigma-Aldrich
Sodium Azide	Serva
Sodium Chloride	Melford
Sorbitol	Melford
T4-DNA ligase	New England Biolabs
TRIS buffer	Melford
TRIS/Glycine/SDS buffer	Bio-Rad
Triton X-100	Sigma-Aldrich
Tryptone	CONDA pronadisa
Tween 20	Melford
Immobilon Western	Millipore
Urea	Melford
Yeast extracts	CONDA pronadisa
YNB	CONDA pronadisa
Zeocin	Sigma-Aldrich

3.1.3 Buffers and media

All buffers and media were prepared using regular distilled water.

3.1.3.1 Media for yeast and bacterial growth

Most of the medium listed below can be found on Easy select *Pichia* Expression Kit (**64**).

Low salt L.B (Luria-Bertani) medium ± Zeocin

1% (w/v) Tryptone
0.5% (w/v) Yeast extract
0.5% (w/v) NaCl
± Zeocin 25µg/mL

Low salt L.B (Luria-Bertani) agar plate ±Zeocin

1% (w/v) Tryptone
0.5% (w/v) Yeast extract

0.5% (w/v) NaCl
1.5% (w/v) Agar
± Zeocin 25µg/mL

Yeast extract peptone dextrose medium (YPD) ± Zeocin

1% (w/v) Yeast extract
2% (w/v) Peptone
2% (w/v) Dextrose (glucose)
± Zeocin 100µg/mL

Yeast extract peptone dextrose agar plate (YPD) ± Zeocin

1% (w/v) Yeast extract
2% (w/v) Peptone
2% (w/v) Dextrose (glucose)
2% (w/v) Agar
± Zeocin 100µg/mL

Stocks for large and medium expression cultures

Biotin 500X (0.02% w/v)
Yeast Nitrogen Base (Y.N.B) 10X (13.4% w/v)
Dextrose 10X (20% w/v)
Methanol 20X (10% w/v)
Glycerol 20X (20% w/v)
1M Potassium phosphate buffer (KPi) 10X
Glycerol 80X (80% w/v)

Minimal dextrose medium (MD)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
2% Dextrose

Minimal dextrose agar plates (MD)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
2% (w/v) Dextrose
1.5% (w/v) Agar

Minimal methanol medium (MM)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
0.5% (w/v) Methanol

Minimal methanol (MM) agar plates

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
0.5% (w/v) Methanol
1.5% (w/v) Agar

Buffered minimal glycerol medium (BMG)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
1% (w/v) Glicerol
100mM KPi

Buffered minimal glycerol (BMG) agar plates

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
1% (w/v) Glicerol
100mM KPi
1.5% (w/v) Agar

Buffered minimal methanol medium (BMM)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
0.5% (w/v) Methanol
100mM KPi

Buffered minimal methanol (BMM) agar plates

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
0.5% (w/v) Methanol
100mM KPi
1.5% (w/v) Agar

Buffered glycerol complex medium (BGY)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
1% (w/v) Glicerol
1% (w/v) yeast extract
2% (w/v) peptone
100mM KPi

Buffered methanol complex medium (BMY)

$4x10^{-5}$% (w/v) Biotin
1.34% (w/v) Y.N.B
0.5% (w/v) Methanol
1% (w/v) yeast extract
2% (w/v) peptone
100mM KPi

Media for electroporation set-up (YPD+HEPES)

1% (w/v) Yeast extract
2% (w/v) Peptone
2% (w/v) Dextrose (glucose)
200mM HEPES pH 8.0

3.1.3.2 Buffers for protein: extraction, solubilization and purification

The buffers presented here can be prepared either in NaCl or KCl.

Breaking Buffer (BB)

50mM Tris-HCl, pH 7.4
1mM EDTA
5% Glycerol
0.1mM PMSF
10mM DTT
Complete protease Inhibitor (10x)

Tris Potassium Chloride Buffer (TKCL):

20mM Tris-HCl pH 7.6
150mM KCl
1mM DTT

Tris Potassium Chloride Glycerol Buffer (TGKCL): SEC

20mM Tris-HCl pH 7.6
150mM KCl
5% Glicerol
1mM DTT

Solubilization buffer (TKCL)

20mM Tris-HCl pH 7.6
150mM KCl
1mM DTT
1.2-0.7% Fos-Choline 12

Ni-NTA Purification Buffer A

20mM Tris-HCl pH 7.6
150mM KCl
1mM DTT
10mM Imidazole (not treated)
0.2-0.15% Fos-Choline 12

Ni-NTA Purification Buffer B

20mM Tris-HCl pH 7.6
150mM KCl
1mM DTT
250mM Imidazole (not treated)
0.2-0.15% Fos-Choline 12

FLAG-tag Equilibration Buffer

20mM Tris-HCl pH 7.6
500mM KCl
1mM DTT
0.2-0.15% Fos-Choline 12

FLAG-tag Elution Buffer

200mM Glicine pH 3.5
150mM KCl
1mM DTT
0.2-0.15% Fos-Choline 12

Thrombin Digestion Buffer

20mM Tris-HCl pH 7.6
0.2-0.15% Fos-Choline 12
150mM KCl
10mM $CaCl_2$

3.1.3.3 Buffers for protein detection and DNA

Transfer Buffer

25mM Tris-HCl
192mM Glycine
10% methanol

Blocking Buffer (TTBS)

50mM Tris-HCl

150mM NaCl
0.05% Tween 20
5% (w/v) BSA

First antibody incubation for hSGLT1

50mM Tris-HCl
150mM NaCl
0.05% Tween 20
1:5000 rabbit anti hSGLT1 (Acris)
5% (w/v) BSA

Second antibody incubation for hSGLT1

50mM Tris-HCl
150mM NaCl
0.05% Tween 20
1:10000 goat anti rabbit HRP
5% (w/v) BSA

Loading SDS buffer dye (5x) for protein

15% Glycerol
1.6% Sodium dodecyl sulfate (SDS)
50mM Tris pH 6.8
0.0002% (w/v) bromophenol value (spoon)

Loading DNA buffer dye (5x)

40% Glycerol
0.25% Bromophenol blue

3.1.3.4 Buffers for transport assays

Uptake Buffer (α-MDG)

PBS buffer
40µM cold glucose
10µM α-MDG
Membrane Filtration Buffer

150 mM Choline Chloride Buffer
10mM HEPES/Tris Buffer pH: 7.5

Goodie Buffer

20mM Tris/HEPES pH: 8.0
300mM KCl

2mM DTT
2mM EDTA
20% Glycerol

3.2 Protocols and methods

3.2.1 Vector design: Cloning

In order to generate a vector that expresses hSGLT1 in *Pichia pastoris*, it is necessary to have the hSGLT1 gene sequence and an expression vector that is suitable for this type of yeast.

The vector we worked with is a modified version of pPCIZ from Invitrogen: pJIN-APM (**64**). This vector has been used widely for the expression of recombinant proteins in yeast. In addition, we obtained a sequence of hSGLT1 from another expression vector that was a kind gift of Prof. Kinne from the University of in Germany (**65**).

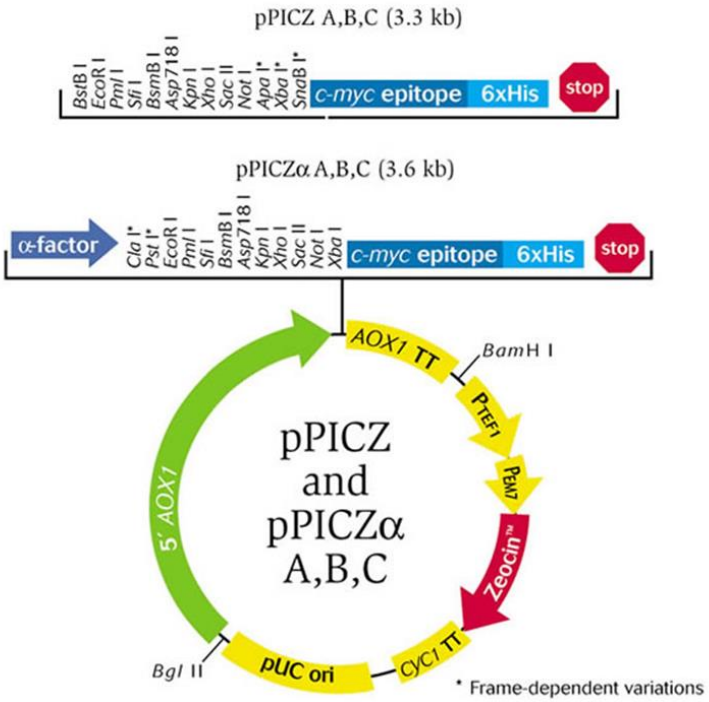


Figure 3.2.1.1: Representation of pPCIZα vector. The vector we used is a modified version of this vector that contains a 8xHistidine tail. The vector also has an NdEI restriction site on the multiple cloning site (MCS) that the original pPCIZα does not have (**64**).

Two primers were design in order to carry out the cloning. Those primers were used to amplify by PCR (Polymerase Chain Reaction) the sequence of hSGLT1 and then clone the PCR product on the pJIN-APM. The primers were designed so after the amplification the amplified DNA could be digested with the restriction enzymes (NdEI and NotI). By following this method, the amplified DNA could be used directly to ligate with the desired vector, as long as the vector has the same restriction site in their MCS (Multi Cloning Site).

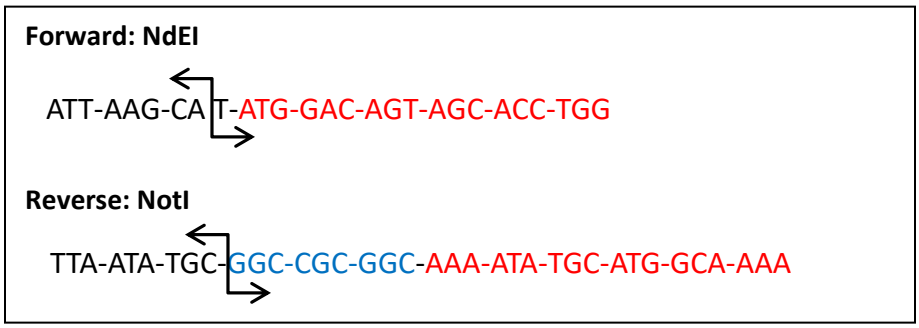


Figure 3.2.1.2: hSGLT1 designed primers. Highlighted in red, the sequence that corresponds with the template hSGLT1 sequence. In blue, additional nucleotides that do not correspond to the sequence of hSGLT1. The arrow represents the cleavage site of the restriction enzyme within the sequence.

The PCR master mix used for the reaction was:

Buffer 10X (NEBuffer4): 5 µL
$MgSO_4$ (25ug/ml): 4 µL
dNTPs (10mM): 5 µL
H2O: 32 µL
Primer Forward (100µM): 1 µL
Primer Revers (100µM): 1 µL
hot Start DNA polymerase (KOD): 1µL
DNA template (600µg/ml): 1 µL

After the master mix was prepared we set up the PCR parameters:

1) Denaturalization: 95°C for 2 minutes.
2) Amplification (25 cycles): The amplification process is done in 3 steps. Those 3 steps will be repeated for 25 times.
    a. Denaturalization: 95°C for 20 seconds.
    b. Annealing: 55°C for 30 seconds.
    c. Extension: 72°C for 1 minute.
3) Final extension: 72°C for 5 minutes.

The PCR product was analyzed by agarose gel electrophoresis 1% prepared in TAE buffer (pH 8.0). After confirming the correct amplification, the PCR product was purified using the PCR-Extraction kit (Invitrogen). The purified DNA was eluted with 30 µl of water so the concentration was the highest as possible. The DNA concentration was checked with a spectrophotometer UV-Vis (Cary 50 Bio UV-Vis spectrophotometer). The eluted DNA (insert) and the vector pJIN-APM were digested with the restriction enzymes NdEI and NotI so the cloning could be perform. The digestion of the insert and the vector was done separately:

Insert digestion reaction:
Insert (85ng/µl): 6 µL
BSA (10mg/ml): 5 µL
NEBuffer3: 5 µL
NdEI (10units/µl): 2 µL
NotI (10units/µl): 2 µL
H2O: 30 µL
Total reaction: 50 µL

Vector digestion reaction:
Vector (375µg/µl): 1 µL
BSA (10mg/ml): 5 µL
NEBuffer3: 5 µL
NdEI (10units/µl): 2 µL
NotI (10units/µl): 2 µL
H2O: 35 µL
Total reaction: 50 µL

The digestion was performed at 37°C for 30 minutes only, due to the low activity of NdEI. After that step, alkaline phosphatase (CIP) was added to the vector digestion medium 6µl and 3µl of NEB3 buffer so the total volume was 60µl and the digestion was carried on for 1 hour. The CIP is used to prevent, in the next ligation step, the vector recircularization.

The digested DNA, for the insert and the vector, was purified with the PCR extraction kit. Since the concentration of the purified DNA after this step is going to be very low, it is necessary to load the digested product on a gel in order to know if the concentration of both samples is similar or not. In our case, the amount of DNA was already adjusted in the digestion so eventually we got the same amount for both digestions after purification. The ligation was set up with different Insert:Vector ratios (3:1 and 6:1) and vector alone in order to ensure obtaining a ligated product. It's important to note that there is no specific method to ligate a vector.

Insert:Vector ratio (3:1)	**Insert:Vector ratio (6:1)**	**Control ligation**
T4 ligase (1u/µL): 1 µL	T4 ligase (1u/µL): 1 µL	T4 ligase (1u/µL): 1 µL
T4 Buffer 10x: 1 µL	T4 Buffer 10x: 1 µL	T4 Buffer 10x: 1 µL
Vector: 1 µL	Vector: 1 µL	Vector: 1 µL
Insert: 3 µL	Insert: 6 µL	Insert: 6 µL
H_2O: 4 µL	H_2O: 1 µL	H_2O: 1 µL
Total reaction: 10 µL	Total reaction: 10 µL	Total reaction: 10 µL

The ligation reactions were carried out overnight at 16°C. The product of ligation was ready to transform into Dh5α (*E.coli*) by heath shock. Protein and cells were plated in low salt L.B. medium with 25µg/mL zeocin. After 24 hours we picked the detected colonies and grew them overnight in L.B. + 25µg/mL of zeocin liquid media. For each colony, a miniprep (QIAgen Kit) was done and with the purified DNA a double digestion with NdEI and NotI was carried out so, if the cloning was correct, our insert would be released. The product of digestion was loaded on an agarose gel electrophoresis 1% prepared with TAE so we could see the release of our insert.

As a final step, we carry out a sequenciation of our miniprep DNA sequence to confirm our ligation.

3.2.2 Transformation in yeast: Electroporation

There are several methods to transform cells like yeast, but, one of the methods with the highest success rate is electroporation, although it might present some disadvantages (**66, 67**).

The first step was to prepare electro competent cells. We grew our empty strain (not transformed) in an YPD plate for 1-2 days. There are several strains to work with (**64**) we used SMD1168h which is a protease deficiency strain.

10 mL of YPD fresh media with a single fresh *P. pastoris* colony were inoculated and were left growing overnight at 30°C. The culture was transferred to 500mL of fresh YPD. The starting OD_{600} of this culture should be around 0.01 and it was left to grow at 30°C until the OD_{600} reached 1.0 (around 10 hours). Cells were harvested by centrifugation at 2.000g at 4ºC for 5 minutes and,

then, the supernatant was discarded and cells were resuspended in 100 mL of fresh YPD medium plus HEPES (pH 8.0, 200mM) in a new sterile flask. After the resuspension, 2.5 mL DTT 1M were added (5mM DTT) and it was mixed gently and for 15 minutes cells were left to recover at 30°C.

Cold sterile water (150 mL) was added into the culture and it was centrifuged at 2.000xg to spin down the cells. Cells were resuspended once in 250 mL of cold sterile water but, at this step, it's critical to mix without vortex and it's advised to either do it by gently shacking or slow pipetting.
As a final step, cells were harvested at 2.000xg at 4°C for 5 minutes and were resuspended in 20mL of cold sterile sorbitol 1M. Cells were then harvested once more and resuspended in 1M sorbitol to a final total volume of 1 or 1.5 mL. Cells were used immediately for electroporation after the last resuspension or frozen at -70º in aliquots of 40 or 80µL. It's recommended to use cells fresh in order to get higher yield of transformants.

The electroporation method can be really tricky and requires some trials before it works out but, the main critical features to worry about are: Presence of salts in the purified vector, the amount of DNA being used and the linearization of the vector to transform. The presence of salts must be minimized since it lowers the yield of transformation and the same happens if the vector to transform it's not linearized.

The vector pJIN-APM was first linearized with a restriction enzyme PmEI. It is necessary to always check if the restriction enzyme, recommended in the manuals (**64**), will cut your gene of interest within the vector. The digestion with PmEI reaction was set up overnight at 37ºC:

Linearization with PmE I:x4
Vector pJIN-APM: 7 µL (6µg)
PmE I (1u/µL): 2 µL
NEBuffer4: 4 µL
H_2O: 27 µL
Total reaction: 40 µL

The digestion reaction was carried out in quadruplicates so we digested initially 24 µg of vector. The whole digestion was loaded in an agarose gel electrophoresis at 1% in TAE so we could check if the vector had been digested. The digested band was cut and purified using the QIAquick PCR purification kit (QUIAGEN). Since the band is going to be very heavy, it might be necessary to divide the band in two or more purification columns. We eluted the final DNA in a total volume of 200 µL in water. In this step all salts should have been eliminated. The DNA was lyophilized overnight and can be kept or used directly next day. The lyophilization is useful so we can work with the highest concentration of DNA as possible.

The last step is the electroporation itself. First, electrocompetent cells (40-80 µL) were mixed in the lyophilized DNA tube until no white DNA pellet was visible. The cells with the DNA were transferred into a 2mm gap electroporation cuvette and left on ice for 5 minutes. Afterwards, cells were pulsed at 1.500 Volts. Immediately after the pulse, 1 mL of 1M sterile sorbitol was added to the cuvette and transferred into a sterile eppendorf (2 mL). Cells were left for 3-4 hours at 30ºC with slow shaking (100 rpm) in order to let the cells recover after the heat shock. Around 50-100

µL of cells were plated in YPD plates with different concentrations of zeocin: 100µg/mL and 500 µg/mL for 2 days.

Following this method, it is very likely that false positives colonies appear on the YPD due to the amount of death cells caused by the voltage pulse. So, it is important to always add a transformation control (no DNA) in order to prevent false positive colonies. Each of the colonies that grew on the plate need to be restreak again in a new YPD plate with zeocin to confirm that is actually a positive colony.

3.2.3 Clone selection

All transformed colonies have integrated the vector on their genome and can technically express our recombinant protein but not all colonies or clones will express it in the same way, so, it is necessary to select which one of those clones expresses our recombinant protein (quantity and quality of expression). There is a correlation between the number of copies of our gen of interest into the genome and the expression levels of protein. Several methods with different approaches can help us out to select a good clone with a good expression level.

3.2.3.1 Serial dilution: Zeocin viability

This method is the easiest and the fastest and was the first one to use (**64**). Every single colony to be tested was inoculated in a 1 mL tube of YPD without zeocin and left growing overnight at 30°C.

Next day, for each colony to be tested, an OD_{600} measure was taken. (Normal values are around 4-6). The optical density was normalized (by dilution) to a value of 1 into small tubes and a serial dilution was carried out from an OD_{600} of 1 to $1x10^{-4}$. From each one of these dilutions, a 5µL drop was put into YPD plates with three different zeocin concentrations: 100 µg/mL, 500 µg/mL and 1 mg/mL. We left the plates until all drops were dry and then we transferred them to an incubator at 30°C for 1-2 days. In 1 day, colonies should start to appear.

For the drop test, it's recommended to work with bigger plates than the usual petri plate's size (60mm to 15mm) so several colonies can be checked at the same time with several dilution factors. An example is shown in figure 3.2.3.1.

3.2.3.2 Colony PCR

This technique is a very straight forward method (**70**). The only issue is that it requires a whole colony to carry out the PCR so it's necessary to make a replica plate of the transformation or streak them in a new YPD + zeocin plate one by one. A PCR mix was prepared with the following composition:

TaqPolymerase (5u/µL): 1 µL Taq Buffer 10X: 5 µL Primer Forward (100µM): 1 µL Primer Reverse (100µM): 1 µL $MgCl_2$ (25mM): 3 µL dNTPs (10mM): 1 µL H_2O: 38 µL Total reaction: 50 µL

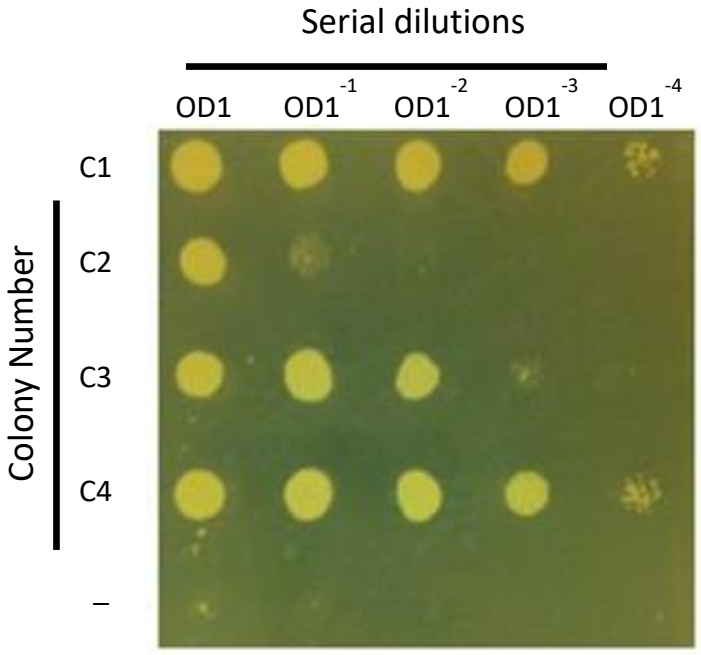


Figure 3.2.3.1: YPD plate with 100µg/mL after 48 hours of growing. Each line represents the grow of a specific colony to be tested with all dilution

The primers used in the reaction were the same used for the cloning. After the mix is prepared, a colony is picked with a pipette tip without touching the agar since it might inhibit the polymerase reaction. After that, the PCR conditions were set up:

1) Denaturalization: 95°C for 5 minutes.
2) Amplification (30 cycles):
    a. Denaturalization: 95°C for 20 seconds
    b. Annealing: 55°C for 40 seconds
    c. Extension: 72°C for 1.15 minutes
3) Final extension: 72°C for 5 minutes.

After the amplification, the PCR reaction product was loaded in an agarose gel electrophoresis at 1% in TAE.

3.2.3.3 Genomic extraction

Since, in some cases, the colony PCR can give us no product of amplification or faint bands, the method can be replaced by a genomic extraction followed by a PCR reaction (**71**).

First, we extracted the genomic DNA for each colony to be tested and then we set up a PCR reaction for those colonies. For the genomic extraction, 10 mL of YPD media were inoculated with positive clones from the electroporation step and left to grow overnight at 30°C. Cells were collected after centrifugation at 3.000xg at 4°C for 5 minutes. The supernatant was discarded and the pellet was resuspended with 400 µL of lysis buffer and 400 µL of acid washed glass beads were added. Cells were lysed with vigorous vortexing for 90 seconds with the glass beads. After, 400µL

of phenol/chloroform/isoamyl alcohol (25:24:1) were added and the cell suspension was vortexed again for another 90 seconds, cells were centrifuged at 16.000xg for 3 minutes.

The resulting aqueous phase (on top) was transferred to a screw cap vial. It is important to avoid pulling material from the inferface at this step. Another extraction was repeated by adding again 400µL of phenol/chloroform/isoamyl alcohol, vortexing for 90 seconds and spinning down. The extracted DNA was precipitated with 1 volume of isopropanol, mixed and left in ice for 15 minutes and afterwards the sample was spun down at 16.000xg for 20 minutes. The supernatant was completely removed and the pellet was washed twice with 1mL of 70% ethanol. It's important to remove all the ethanol left. The final washed pellet was resuspended in 20 µL of TE buffer supplemented with 0.5 µg of RNase A (DNA free) and the mixture was incubated overnight at room temperature in order to remove all the RNA content. Finally, 30 µL of TE buffer pH 8.0 were added and the concentration of DNA was checked with an UV-Vis measure. The extracted genomic DNA samples were loaded on a 1% agarose gel to visualize the DNA. The DNA was stored at -20ºC until further use.

A PCR reaction was set up under the same conditions described in the colony PCR but with 1 µg of genomic DNA as a template.

3.2.3.4 Plate induction

This method is the fastest and gives us a way to quantify how much protein might express a colony (**66**). This method consists in inducing our recombinant protein in a fresh methanol plate (BMM plate) and then checking the fluorescence of the plate using a dark reader transiluminator. Obviously, this method only works when our recombinant protein is fused to eGFP so we can monitor the expression.

In order to prepare BMM plates, it is very important to add the methanol at the end, once the media has cooled down a little, to avoid methanol evaporation. Like in the serial dilution method, bigger plates are required in order to screen as much colonies as possible, so typical size petri plates are not recommended when a large number (25-50) of colonies is going to be screened.
5 mL of YPD media were inoculated with a positive colony of the electroporation and left growing overnight at 30°C. Next day OD_{600} measures were taken for each grown colony and were normalized (dilution). For all colonies, 100µL were set aside to a final $O.D_{600}$ of 1 in BMM. A drop test of 5 µL for each of those colonies was carried out in a freshly prepared methanol plate. Once the drop has dried out in the plate, we left the plates at 30°C for 48 hours. It is very important not to put the plates upside down in the incubator until the drops are dry.

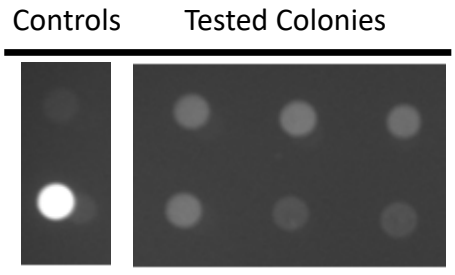


Figure 3.2.3.4: Transiluminated BMM plates after 48 hours of induction. First lane represents the two controls that should be always included in those plates: The first dot represents the SMD1168h colony without being transformed and the second dot is the positive control (transformed with the empty vector). The remaining three lanes are different colonies that were being tested.

At 24 and 48 hours we checked the fluorescence by transiluminating the plates and a picture was taken. The picture was analyzed by the quantity one program so each dot could be quantified.

3.2.3.5 Semi quantitative PCR

As mentioned before, after electroporation some colonies have more than one copy of the gen of interest and those colonies are more likely to express protein so it's important to know the gen dosage of the colonies. The PCR reaction can give us this information. The semi quantitative PCR is different from the genomic screening method because in this case the PCR is set up to work in the exponential phase instead of working at the plateau (**71**). A polymerase chain reaction theoretically amplifies template DNA fragments in an exponential manner and, depending on the amount of DNA template, there will be more or less amplified DNA (amplicon) but, in the plateau phase, the reagents necessary for the PCR are running out so the number of amplicons will not be proportional to the starting DNA template.

Based on this, an agarose gel can be helpful as a tool for evaluating the number of copies (thicker or thinner band on the gel) depending on the starting material. This method was only used for the Wild Type (WT) construct because with the fused recombinant protein to eGFP the plate induction is better.

It is important to define the number of cycles we are going to work with in our PCR so we are not in the plateau phase. A colony was tested using its genomic extracted DNA, described in the 3.2.3.3 section, with the same PCR setup and the same reaction conditions. A small aliquot was taken at: 18, 20, 22 and 24 cycles. The PCR product was loaded on an agarose gel 1% in TAE buffer. After, a new PCR reaction is set up with all the clones to be tested with the same conditions as tested before. Once the PCR was done, we run the products on a new agarose gel 1% in TAE buffer so we could compare the level of expression.

3.2.4 Expression: Screening conditions

All the screening for expression conditions was carried out with the hSGLT1+eGFP construct.
A fresh YPD agar plate (without zeocin) was streaked with our selected colony, for 1-2 days at 30°C, until colonies appeared on the plate. Next day, a colony was inoculated to a 25 mL of Minimal Dextrose Medium (MD) and was left overnight to grow at 30°C. The culture was centrifuged at 3.000xg for 5 minutes, the supernatant was discarded and the pellet was washed once with water to completely remove the media. Cells were spun down one more time at the same speed. Again, the supernatant was discarded and the pellet was resuspended in Minimal Methanol Media (MM) and the culture was transferred back to the incubator at the desired temperature for induction.

Small aliquots (1 mL) were removed from the culture at the desired time and measures were taken at the plate reader or the sample was kept at 4°C for further use. For each well, 10µL of sample were diluted with 190 µL of media. Triplicates were measured for all conditions and a blank with media was used. Fluorescence and OD_{600} measurements were taken for each well. The optical density is useful to normalize the fluorescence signal so values from different cultures can be compared because the initial OD_{600} of all cultures it's not the same. If, after the dilution, we get an optical density measurements higher than 2, another dilution is required.

It is important to add that media and temperature were screened in this way so, depending on the medium and temperature to be screened, different media and temperatures were selected as starting culture conditions and the same for the induction media.

3.2.5 Expression: Big scale

At this point, medium conditions and temperatures were determined and were scaled to bigger volumes and applied for both constructs (**64**)

A fresh YPD plate was streaked with our selected colony and was left growing for 2-3 days at 30°C until small colonies appeared. It is important to note that lower levels of expression would be detected if the plate is older than one week 200 mL of BMG media were inoculated with a fresh colony and left to grow at 30°C for 1 day. Next day, the culture was transferred and divided to a new freshly prepared BMG media of 4L and incubated at 30°C for another day (**72**). Afterwards, we added up to 1% (w/v) of Glycerol so the culture can keep growing for another day at the same temperature. We worked with a high concentration of sterilized glycerol (80%) in order to prevent the dilution of the media after the glycerol addition. Next day, we spun down the culture at 3.000xg for 10 minutes. The resulting pellet was resuspended with 4L of BMM and the culture was transferred once more to an incubator at 24°C. The induction was always done for 20-22 hours, after; the culture was centrifuged once more (**65**). The pellet was cleaned with water so no remaining media was left. The final pellet was weighted and kept at -80°C or, in case of having liquid nitrogen available; a snap freeze was done first and then kept at -80°C. Typically around 60 g of cells were obtained from 4L of culture.

3.2.6 Fluorescence imaging: Protein localization

It is possible to check the localization of our expressing product with the use of a fluorescence microscope or a confocal microscope but only with our GFP fused construct.

1 mL of fresh cells expressed in different medium was spun down at 3.000xg for 5 minutes and washed with water to completely remove the media. The protocol for protein expression is in the one described in 3.2.5. Cells were normalized to an OD_{600} of 1.A drop of 5 µL of cells was put on a slide and the cells were covered with a coverslip.

3.2.7 Cell break down: Glass beads

A mechanical process with the use of glass beads was applied to break *Pichia* cells. The method can be used either for large extractions (bead beater) or for medium/small scale (regular vortex) (**65, 99**).

Cells were slowly unfrozen at 4°C and resuspended in breaking buffer (BB). Up to 80 g of cells can be resuspended in a total volume of 175 mL of BB but around 40-50 g of cells is recommended. The resuspension was performed with a vortex. We added 10mM of DTT and protease inhibitor to prevent proteolysis during and after the extraction. Different protease inhibitors were used: Complete protease inhibitor cocktail (ref) or 0.1mM PMSF plus pepstatin, leupeptin and aprotinin. Afterwards, 1 volume of glass beads of 0.5-0.2 mm is added to 1 volume of BB plus cells in a steel chamber.

In order to avoid heating as much as possible, the bead beater pieces were put previously on a -80°C and used in the cold chamber at 4°C. Up to 20 cycles (1 minute per cycle plus 2 min resting time in between) were carried out for cell disruption (**65**). The bead beater was kept full with ice throughout the process. After the breaking cycles, the supernatant was decanted into a beaker while trying to avoid decanting glass beads. Afterwards, more B.B medium (the same volume as the volume that was decanted) was added to the glass beads in the steel chamber.. If the bead beater was full (350 mL), around 150 mL are recovered by decanting (300 mL total). Later, we centrifuged at 3.000xg for 10 minutes in order to remove the unbroken cells and the big broken particles. Afterwards the supernatant was transferred to an ultracentrifuge tube and the pellet was discarded. The membranes were spun down at 100.000xg for 45 minutes to 1 hour. The membrane enriched pellet was weighted and kept at -80°C.

When working with smaller volumes, the procedure is exactly the same but instead of working with a bead beater a regular vortex is instead used. In this way, it's important to use a glass tube with the glass beads instead of a plastic tube to ensure a good yield.

3.2.8 Detergent screening: Solubilization

All of the procedures presented here, except the western blot analysis, were carried out with the GFP fused construct. The methods presented are different ways of monitoring the transporter solubilization using detergents in order to find out the optimal conditions (**73, 75, 76**).

3.2.8.1 Plate reader

Cell membranes resulting from the bead beater breaking process were homogenized in TKCL or TGKCL at a specific ratio (specifications are given in the corresponding results section) of grams of membranes per volume of buffer. The homogenization was carried out using a regular drill homogenizer. Aliquots of 1mL were prepared with the homogenized membranes and a specific detergent was used for each aliquot. It's important to work above the detergent critical micelle concentration (CMC).

Each aliquot was left to solubilize at the desire temperature for 1-2 hours or, in some cases, overnight. Afterwards, the solubilized material was centrifuged at 50.000 to 100.000xg for 1 hour and, in case no ultracentrifugation is available or the volume is too small, a fast spin down can also be done in a table top centrifuge for 10 minutes at maxim speed (17.000xg). It is important to set aside a small volume before centrifuging the sample since it's going to be useful from analysis as the starting total material. Up to 200 µL were taken for each sample and a fluorescence measurement was taken at the plate reader. The excitation filter and the emission filter used were 485nm and 520nm respectively. For each detergent condition to be tested three measures were taken; one for the supernatant, the total solubilized material and the pellet (not solubilized).

3.2.8.2 High pressure liquid chromatography (HPLC)

Membranes were unfrozen slowly and homogenized in TKCL or TGKCL at a specific ratio of grams of membranes per volume of buffer (specifications are given in the corresponding results section). 1mL Aliquots were separated and the detergent was added. Since Membranes were solubilized for up to 2 hours. Afterwards solubilized membranes were centrifuged at 50.000xg for 20 minutes.

Although each sample might have been solubilized in one detergent, the HPLC column will be equilibrated in a specific buffer that it might be different from the buffer use to solubilize our sample.

For the GFP signal (480/510 excitation and emission filters), the high sensitivity program was used and around 500ng to 1µg is the amount of recommended protein to inject. The other detection system is the Tryptophan fluorescence (280/335 excitation and emission filters) and a medium sensitivity program was used since this method of detection is noisier. The recommended amount of protein for injection in this case is 3 µg. Two different columns were used: Superose 6 10/30 and TSKG3000SW_{XL}. Each of those columns is a size exclusion molecular weight column (SEC) so, depending on the resolution, the profile will be different.

This method will give us much more information than the plate reader because we can check fluorescence intensity but also monodispersity since it's a SEC purification.

3.2.8.3 Western blot analysis

Cell membranes were solubilized with a specific detergent overnight as in 3.2.7. Next day, the samples were centrifuged at 100.000xg for 45 minutes. A small aliquot of each condition was set aside before centrifuging. The supernatant was set aside and the pellet was suspended in the same initial buffer at the same volume.

The pellet, the supernatant and the total sample (not centrifuged) were mixed and solubilized, separately, in loading buffer for SDS gel plus β-mercaptoethanol. 10 µL of each sample were loaded in a SDS electrophoresis gel (10-12%) and we run the gel for around 2 hours until the front (of the dye) runs out of the gel.

Afterwards, the gel was transferred with a semi dry blotter system for 1 hour. An already activated nitrocellulose membrane was used. After that, the membrane was removed from the blotter system and soaked in TTBS buffer to remove the rests of transfer buffer. Membrane blocking was carried out in TTBS with 5% (w/v) BSA for 1 hour. Afterwards, the membrane was incubated with the first antibody (rabbit anti hSGLT1) in TTBS for 1 hour and then the membrane was rinsed 3 times with TTBS for 10 minutes each time. The membrane was then incubated with the secondary antibody (goat anti rabbit), in TTBS for 1 hour and followed by 3 washes of 10 minutes with TTBS.
The resulting membrane was developed using a chemiluminescent HRP substrate. A small exposition is enough (less than 1 minute) to visualize the different bands.

3.2.9 Purification of hSGLT1

The purification of hSGLT1 consists of two or three purification steps depending on the purity of the final product: Two Affinity column purification (IMAC, Flag-tag) and a size exclusion chromatography (SEC)

3.2.9.1 Affinity chromatography: Immobilized metal ion affinity column purification: (IMAC)

The IMAC purification method is based on the interaction of a polyhistidine tail to metals such as nickel (Ni^{+2}). The nickel is bound to nitrilotriacetic acid (Ni-NTA) and to an agarose bead that can be packed as a column. The Ni-NTA resin has a binding capacity of 10 mg/mL for soluble proteins and around 3 mg/mL for membrane proteins. This purification method was carried out in two different ways: Connected to an AKTA system or by gravity flow (manual) but both methods share the same principle.

For the AKTA system purification, a 5 mL Ni-NTA column was used a; HisTrap HP (High Performance) column. This column was connected to the AKTA system and two different buffers: Low imidazole buffer (A) and high imidazole buffer (B) (see section 3.1.3.2 for more details).

The column was initially washed with 2 column volumes (CV) of buffer B (100%) to strip out any proteins remaining bound to the column. Afterwards, the column was equilibrated with a 2 column volume (CV) of buffer A (100%). After the column has been equilibrated properly, the sample is injected via pump. An accessory to the AKTA system is necessary to do that. The sample is injected at a low flux of 0.1-0.3 mL/min and its left running over night since the volume of the sample is large. After sample injection, a washing step was carried out at the same flux with a gradient of buffer B (5%) and buffer A (95%) of 15 CV. This washing step was done to elute unspecific proteins from the column. The washing should result into a stable reading of the OD_{280} signal Afterwards an elution step is done with 2 CV of buffer B (100%). Fractions of 500 µL were collected and analyzed by SDS Polyacrylamide Gel Electrophoresis (PAGE).

The gravity flow method requires, initially, packing a 5mL column of Ni-NTA resin. The packing is done in water so the ethanol (20%), in which the resin is kept, can be removed. Afterwards, the column is equilibrated in 5 CV of solubilization buffer, the sample is loaded and the flow is stopped. The resin with the sample is left in batch mode shacking at 4°C for 3- 4 hours (**71**). It is important that the resin does not sediment or dries out during the batch. The batch mode will improve the binding of the protein with the resin if the interaction is lower than expected. Once the batch is finished, the resin with the sample is transferred back to the column and we open the flow and let the sample flow through the resin. The flow through is inserted back to the column to make sure no resin was lost during the batch to column transfer. Afterwards, an initial washing step (1) of 5 CV of TKCL buffer plus 0.2% F12 was carried out followed by washing step (2) of 10 CV of TKCL buffer plus 0.2% F12 and 10mM of Imidazole. This was followed by an elution with 5 CV of TKCL plus 0.2% F12 and 150mM Imidazole. After elution, a spectrum from 600 to 200 nm was done to check the concentration of protein with the use of the molar extinction coefficient.

3.2.9.2 Affinity chromatography: FLAG-tag purification

The FLAG-tag method is based in the interaction with a specific peptide sequence to a specific antibody bound to an agarose resin.

This purification method was used after the Ni-NTA purification. The sample, which was eluted with high imidazole concentration in the Ni-NTA, is concentrated and diluted to remove imidazole to a final concentration of 20mM. The buffer used for dilution must be the same but without

imidazole (TGKCL + 0.2% F12 +1mM DTT). High imidazole concentrations can impair the FLAG-tag purification so it's better to remove it or lower it down.

The flag-tag resin was poured in a gravity flow column and washed with 2 CV of PBS to remove the ethanol with which the resin is kept. Afterwards, the column was equilibrated with 2 CV FLAG-tag equilibration buffer (see 3.1.3.2). This buffer must have a high concentration of salts to avoid unspecific binding to the column. The sample was introduced to the packed column and left at 4°C in batch mode overnight. Next day, the resin was packed again and all unbound proteins were let flow through from the column. It's important to keep that flow through to evaluate later the purification. More equilibration buffer was added until no protein came out from the flow through. Afterwards, an elution step was done with 5 CV FLAG-tag elution buffer (see 3.1.3.2). Since the elution buffer is very acidic (pH 3.5) and most proteins aggregate or precipitate at this pH, a high concentration of Tris-HCl Buffer (1M pH 8.0) is prepared on the elution tubes which will equilibrate the pH. The amount of 1M Tris-HCl required in the elution tube has to be checked previously which depends on the final pH need it and the final volume of each elution tube. It's possible that 5 CV is not enough for completing the elution, so, it's better to keep checking the protein concentration until no protein is detected.

Once the elution is complete, it is important to regenerate the column or otherwise the elution buffer from the column can damage the resin and impair future purifications. To readjust to neutral pH, 3 CV of 1 M Tris-HCl at pH 8.0 was used. Afterwards, the column was washed out with 10 CV of PBS with sodium azide 0.02% so the resin could be preserved.

3.2.9.3 Size exclusion chromatography (SEC)

This purification step is always the last one and samples used for this purification come either from the Ni-NTA or from the Flag-tag purification. Nowadays, this technique is used mainly connected to an AKTA system. The columns used were a Superdex 200 HR 10/30 anda Superose 6 10/300 although most of the runs were done with the Superdex 200 HR 10/30.

In order to get good separation and resolution with the SEC columns, the sample volume cannot exceed 1% of the column volume Since all columns used were around 25 mL of packed bed volume, the injection sample was very low and must always be concentrated firstly. Also, the injection sample must be centrifuged before injecting at least at 17.000xg for 15 minutes up to 50.000xg for 15 minutes (recommended) to spin down any aggregates. The sample was concentrated using a regular centricon with a 50 kDA cut off membrane. This step helps to remove contaminant proteins of low molecular weight. It's important not to concentrate the sample higher than 5-6 mg/mL or otherwise the protein might aggregate so, in some cases, several injection rounds must be carried out to inject the whole sample.

Elution with the AKTA column was carried out at a flux of 0.4 mL/min. Aliquots of 0.5 mL were collected. For a higher purification protein, the fractions that contain hSGLT1 were mixed, concentrated, centrifuged and were injected agian into the column. This step was only done when the initial sample weren't purified with the Flag-tag column.

3.2.9.4 His tail and GFP tag removal: Thrombin digestion

This protocol was used to remove the whole tag of hSGLT1 fused to GFP with a thrombin digestion. For more details about the construct check chapter 3.2.1.

Two different methods were carried out to remove the tag with the thrombin because, in some cases; it's described that it's difficult to perform the thrombin digestion (**77, 78**). The digestion was done: In Ni-NTA column or after Ni-NTA.

3.2.9.4.1 Thrombin cleavage in the Ni-NTA column

The solubilized protein is loaded into the Ni NTA column as described before. Once the protein is bound to the column, thrombin and $CaCl_2$ to 10mM are added. The packed resin is removed and left under stirring over night at 4°C (batch-mode). The amount of thrombin added depends on the amount of hSGLT bound to the column. A fluorescence measure was taken with the solubilized membranes before starting the purification as described in (**74**). This measure gives us an idea of the total amount of recombinant protein so we can calculate how much thrombin should be added. The thrombin added was equivalent to the amount needed to digest all the protein in half an hour at room temperature. Longer timings can be done up to 3 days of digestion since the yield can be very low.

Next day, the resin is transferred to a column so it can be packed again and we let the flow through pass. If the digestion was a success, the recombinant protein should elute from the column.

3.2.9.4.2 Thrombin cleavage after Ni-NTA purification

The hSGLT1 can be cleaved with thrombin once it has been eluted from the Ni-NTA column. Since imidazole can interfere, it is necessary to remove it. To remove the imidazole, a centricon with a 50 KDa cut off membrane was used. The final goal is to reduce the imidazole concentration of our purified protein down to 10mM with a final volume of 1 to 1.5mL. Afterwards, a spectrum at 280 nm was measured in order to determine protein concentration and to check for possible aggregation of protein due to the concentration. Thrombin was added and the digestion was carried out overnight at 4°C under gentle shacking. The amount of thrombin added was determined as described in 3.2.9.4.1.

Next day, the same Ni-NTA column was equilibrated with thrombin digestion buffer. Afterwards, the digested protein was added to the column with equilibration buffer (3 CV) and it's left in batch mode at 4°C for half an hour. Next step was to open the column and let all the digested protein flow through the column because it cannot bind to the column due to the loss of the histidine tail. Afterwards, an elution step is performed with TKCL + 250mM of Imidazole with Fos-12 at 0.2% so the non-digested proteins are eluted.

3.2.10 Proteoliposomes preparation and dynamic light scattering (DLS)

This method requires two main things: An initial preparation of empty liposomes and purified protein which afterwards will be analyzed by DLS.

3.2.10.1 Liposomes preparation

Asolectin from soybean, Phosphatidylcholine (PC) and Phosphatidylserine (PS) were used as lipids.

Lipids (supplemented with cholesterol (10:1 w/w ratio) when indicated) were weighted and solubilized in chloroform (100%) in a round bottom flask. In some cases, lipids are difficult to solubilize in chloroform alone due to lipid hydration so a solution of chloroform (80%) methanol (20%) was used. Afterwards, the solvent was removed by the use of a rotatory evaporator and a lipid film should be visible at the bottom of the flask. If the solvent was completely eliminated, the film shouldn't have any trapped methanol or chloroform bubbles and the flask should have an acrid smell, otherwise, rotoevaporation should be repeated.

Buffer (TKCL) was added to the resulting lipid film at a final concentration of 10 mg lipid/mL up to 50mg lipid/mL. These liposomes can be stored up to one month but it's recommended to prepare them fresh (**79**).

3.2.10.2 Proteoliposomes preparation

Initially, liposomes were solubilized with a detergent (F12 or TX-100). The amount of detergent added depends of: At which set of solubilization we are going to work with, the lipid concentration and the type of lipid (**80**). In our case, we have been working at total solubilization (mixed micelles detergent/lipid). As an example, for a final concentration of PC at 1mg/mL, the TX-100 concentration used is 0.5% (w/v).

Since no bibliography exists on the use of F12 to solubilize liposomes, it was characterized here for the first time (see results section 4.9.1). These values are different for every detergent and experimental condition and must be calculated previously unless they have been described before. The solubilization with TX-100 was carried out for an hour at 4°C. It is recommended that an UV-VIS spectrum is done to ensure that the sample is not turbid at the end of the solubilization. Afterwards, the protein was added to the solubilized liposomes. The amount of protein added might be higher or lower depending on what are those proteoliposomes going to be used for, although the standard amount of protein used was a ratio of 1:100 (w/w) of protein to lipid. Immediately after, bio-beads were added to the final mixture in order to remove the detergent. The bio-beads were previously weighed and cleaned once with methanol 100%, then, the methanol was removed by decanting and 2 more additional wash steps were done with water. In this process, some bio-beads are going to be discarded due to flotation (air bubbles trapped inside the bio-bead). The beads were weighted wet after these wash steps.

Different bio-beads amounts were tested since it can affect the rate of detergent removal and the adsorption of lipids and protein to the bio-beads (see the results 4.9.1) (**80, 81**). Nevertheless, the additions were done always in three steps of 1 hour each at 4°C except the last one which was done overnight. Next day the sample should be slightly turbid. As an additional final step, liposomes can be centrifuged at 100.000xg for 45 minutes. The supernatant is discarded and brownish pellet should be visible and has to be resuspended with the desired buffer.

3.2.10.3 Dynamic light scattering (DLS)

Dynamic Light Scattering (DLS) is used to study the size distribution (polydispersion) of a sample. An initial blank is done with just water or either buffer. After the blank, the sample is measured but with a minimum volume required of 2 mL. Each measure taken was an average of accumulated scans (5 minutes of reading). The machine has a range of detection between 10 nm to 5 µm. An optimal size of proteoliposomes should be around 100-200 nm and a monodisperse distribution.

3.2.11 Fluorescence measurements

Fluorescent measurements were carried out for hSGLT1 in proteoliposomes and in micelles.

All fluorescent emission measurements were done at an excitation wavelength of 280 with an emission maximum at 330 nm at 0.5nm/s. The tryptophan, tyrosine and the phenylalanine absorb at 280nm and emit light at 300-350 nm. Nevertheless tryptophan is the main contributor. The temperature of the sample was controlled by a peltier and was set up at 25°C unless it's stated otherwise.

All kinetic measurements were done with agitation and the sample volume for all fluorescent spectrums was 1 mL.

3.2.12 Voltage clamp in planar lipid membranes

The experimental protocol employed here is used to study the electrophysiology of the protein channel inserted into a planar lipid membrane; this method is based on (**82**).

Within this methodology, a Teflon (polytetrafluoroethylene) chamber formed by two compartments, with 2 mL capacity each, is used for planar membrane reconstitution. Chamber compartments are separated by a thin Teflon film 10 µm thick with a handmade 100 µm diameter orifice. The membrane is formed by adding lipids in both sides of the chamber but, previously; the Teflon was pretreated with a drop of hexadecane causing it to be hydrophobic. Lipids used in this technique are very rigid and impermeable like: DPHPC (1,2-diphytanoyl-sn-glycero-3-phosphocholine) (other similar detergents can be used too). Due to the Teflon hole and the biophysics characteristics of DPHPC lipids, the resulting membrane is a planar bilayer. See figure 3.2.12 for more details on the set-up system.

The experimental chamber is connected to a signal amplifier through specific electrodes fabricated in the laboratory, see figure 3.2.12. Together, the signal amplifier and a recording system form a circuit suitable for the recording using the voltage-clamp technique. In the voltage-clamp technique, the membrane potential is held constant while the current flowing through the membrane is measured.

The membrane cell is an equivalent to a parallel resistor and a capacitor when an external voltage is applied in the chamber. When specific external voltage equation is applied to the chamber the resulting signal must be equivalent to a resistance of the membrane (capacity). If the membrane was formed correctly, a quadratic signal should be observed of 100 to 80 mV which reflects the capacity of the membrane. Higher membrane capacity or lower might affect protein

incorporation. This step is critical because it is mandatory to check that the membrane is formed correctly before adding protein. This membrane capacity should be observed for both compartments, in order to do so, the liquid level in each side of the chamber is diminished. If membrane was formed correctly, a quadratic signal should be observed for both sides.

Afterwards, protein is added. Normally few µL of protein are added until protein insertion is observed. Typically not more than 10 µL of protein at (3-4mg/mL) were used. The detergent-micelle protein is added in one side of the chamber. The detergent is immediately diluted to values below CMC which will cause the protein either ton aggregate or to insert in the planar bilayer. As mentioned before, the formed membrane is very rigid and, therefore, will not allow ions to flow from one side to the chamber. So in a basal state no current should be measured when an external voltage is applied.

Nevertheless, if the protein is incorporated in the membrane some current can be measured when a high external voltage is applied (150 mV). A signal of around 3-5 pA should be observed if protein was well incorporated. This is because protein insertion causes the membrane to be less rigid and some ion can cross it causing a low but measurable current.

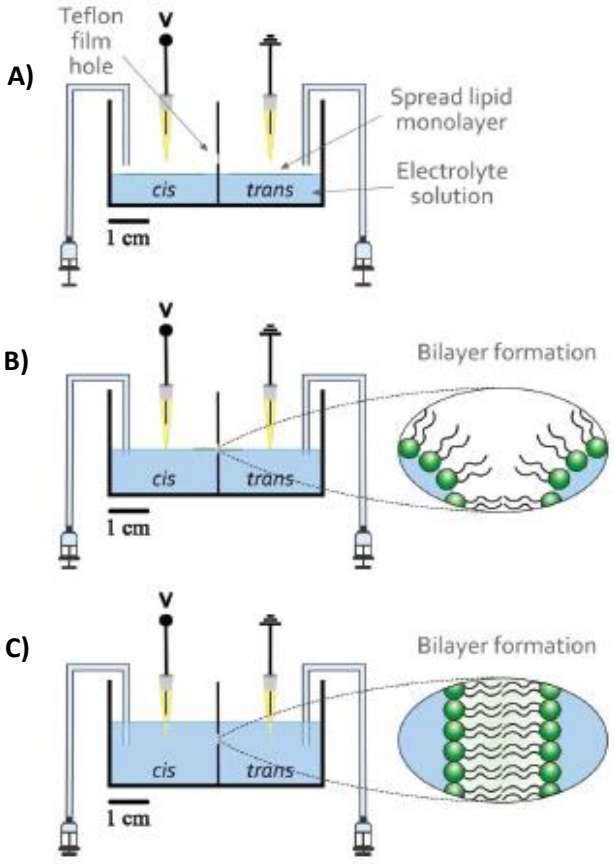


Figure 3.2.12: (A) Small drops (around 15 mL) of the selected lipid immersed in pentane (5 mg/mL) are spread on top of the electrolyte solutions that filled each compartment of the chamber to form lipid monolayer. (B) Solution level in each compartment is raised gently to allow for bilayer formation. (C) Once solution levels are completely covering the hole, a stable bilayer is formed. For all the experiments performed, chamber compartments are labeled as *cis* (compartment where the protein will be added) and *trans* (side set to ground).

The system is set up in a way which low currents can be measured. The whole system is inside a non-vibration chamber or otherwise the environmental would be higher than the recorded signal.

Once protein is incorporated, measures at ± 150 mV, 100 mV and 50 mV were recorded for each condition to test. Initially, an addition of sodium chloride 150mM of NaCl was done and the currents were measured. Afterwards, D-glucose was added to 5mM and the currents were measured. Finally, phlorizin (hSGLT1 inhibitor) was added to a final concentration of 100 µM. After adding inhibitor, a new membrane is prepared and a new insertion has to be made.

3.2.13 Transport Assays

The transport assay was done at three different stages: For cells, for isolated membranes and for proteoliposomes, so, the difference between each protocol is mainly the starting material.

The transport reaction was done by mixing 10 µL of liposomes, membranes or cells and 40 µL of transport buffer, as described in the section 3.1.3.3, in a plastic tube. A triplicate was done for each sample to check. The tubes were transferred to an incubator at 37^{o}C for 20 minutes. Then, the samples are mixed with 4 mL of membrane filtering buffer, described in 3.1.3.3, and then filtered in a vacuum system with a 0.3 µm membrane filter. Afterwards, another filtration is done with 1mL of membrane filtering buffer. After the last filtration step, the membrane is removed from the vacuum system and transferred to scintillation vial. It's not recommended to start the transport at the same time for all of the reaction tubes and, instead, leaving and interval of 1 to 2 minutes from one sample to another it's mandatory so there's enough time to process and filter the samples. Membranes are dissolved with 1 mL of ethyl acetate for 30 minutes, although it's an optional step, it'll ease the readings in the scintillator counter. As a final step, 6mL of scintillation liquid is added to enhance the signal and readings are taken.

As said before, the starting material can be cells, proteoliposomes or membranes. In the case of working with cells, a small scale (50 mL) culture was grown and induced as described in 3.2.5. After 24 hours of induction, the cells were kept on ice and the optical density at 600nm was checked. Cells at different optical density were used; at 0.1 and 1.0. Higher optical densities will clog the membrane filter.

For proteoliposomes, an initial preparation of 5 mL was setup over night at 1.2mg/mL of lipids. After removing the beads, the proteoliposomes were ultra-centrifuged and the small pellet was resuspended with 100 µL of goodie buffer, described in 3.1.3.3, using a syringe. Those proteoliposomes will be used directly for the transport assay unless the pellet is white. A white pellet, it's probably due to protein aggregation and it's not worth using it, but, if the pellet is brownish, it can be used.

In the case of membranes, cells are broken down and after some centrifugations, as described in 3.2.7; the membranes are weight and kept at -80^{o}C. Membranes are suspended in vesicle transport buffer (described in 3.1.3.3) so the final optical density at 600nm for each sample to test is the same. The final optical density at 600nm used for those assays was 1-2.

4-Results and discussion

4.1 Vector designs: Cloning

Two expression vectors for hSGLT1 were designed, one with hSGLT1 fused to eGFP and a 8X His-Tag and another one for hSGLT1 fused to a 8X His-Tag, both in frame with the C-terminus of hSGLT1. Those two vectors were obtained by the same method as described in the material and methods chapter 3.2.1.

4.1.1 PJIN_hSGLT1 fused to eGFPhSGLT1 expression vector for fluorescence monitoring

The initial step to obtain this vector was the amplification by PCR of the hSGLT1 sequence with the primers described in the material and methods and then ligated in our vector of interest. The target vector for the ligation is a modification of pPCIZα which contains an eGFP expression sequence.

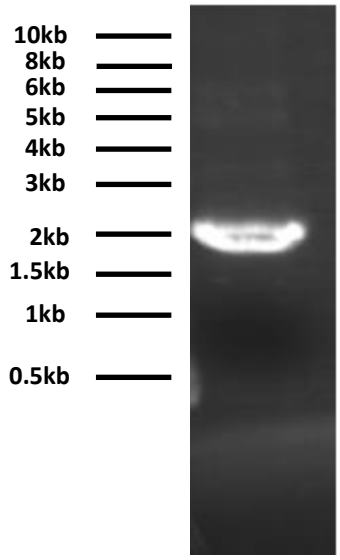


Figure 4.1.1.1: Agarose gel electrophoresis at 1% in TAE Buffer of the PCR product after amplification with a DNA ladder

hSGLT1 gene cloned in a pPCIZα vector (**65**) was used as DNA template for the PCR reaction. The hSGLT1 coded is modified by the addition of a FALG-tag at D_{574} (D_{574}AEEEN to D_{574}YKDDDDK).

The pPCIZα is a 3,3kb vector (**64**) and the hSGLT1 insert sequence is 1.9kb. The product of the PCR reaction is shown in figure 4.1.1.1 in an agarose gel electrophoresis. A very bright and intense band appears at 2kb, as expected; also, some non-specific bands with very low intensity are visible at higher molecular weights.

PCR product and vector were digested with NdEI and NotI restriction enzymes, bands excised and purified by standard methods (see Materials and Methods section 2.2.1). Purified bands were ligated and transformed in DH5α *E. coli* competent cells by heat shock protocol. Cells were plated into selective zeocin plates and Plasmid DNA was isolated from 4 colonies. Isolated DNA was digested with the NdEI and NotI to assess the presence and size of the insert. In figure 4.1.1.2, the product of digestion of different miniprep colonies is represented in an agarose gel electrophoresis with the corresponding controls linearized with NdeI.

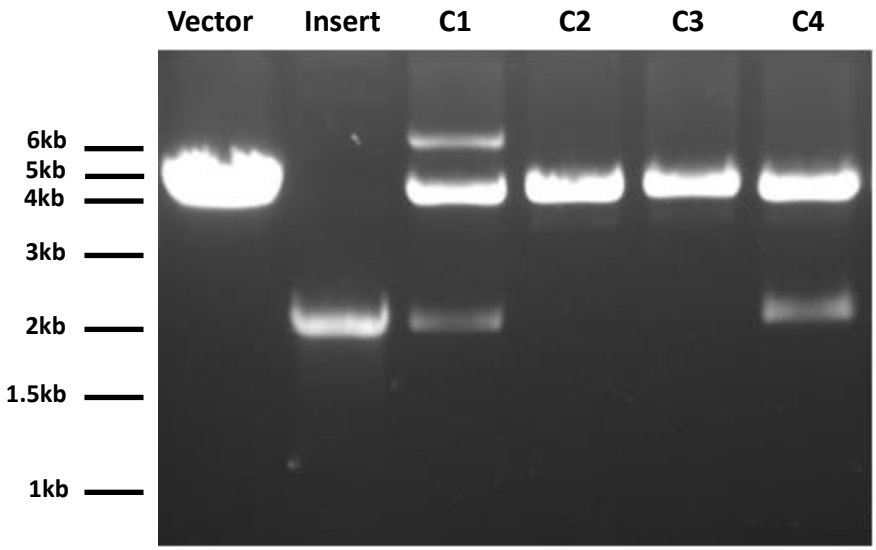


Figure 4.1.1.2: Agarose gel electrophoresis at 1% in TAE Buffer of different colonies that were positive after transforming them in LB low salt with 25µg/mL of zeocin. L: ladder; Vector: full length without digestion; Insert: linearized sequence of hSGLT1; C1-C4: miniprep digested DNA with NdeI and NotI.

Both the insert and the vector appear at around 2kb and at 3-4kb, respectively, in lanes C1 and C4. Plasmids from lanes C2 and C3 were discarded. In C1, an insert is released but another band appears at higher molecular range around 6kb. That band could be non-linearized vector due to a poor digestion or a resulting product from a bad ligation. Because of that, the last lane, C4, was chosen as the right one and selected for DNA sequencing, which turned out to be correct.

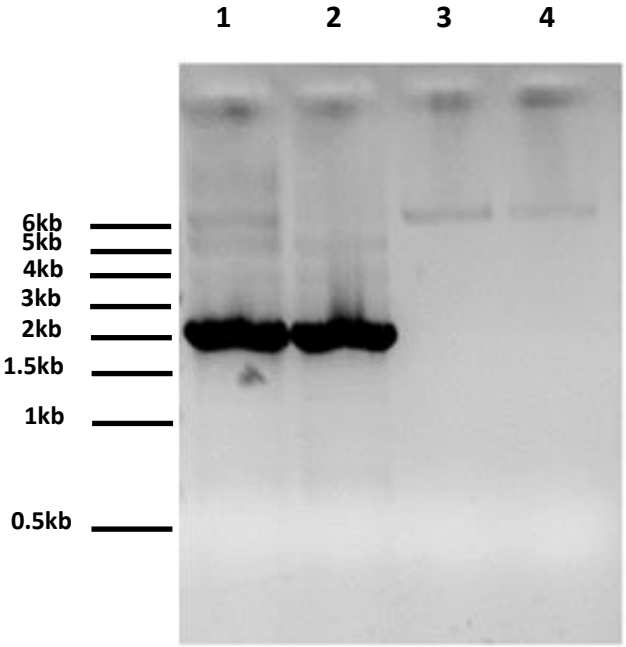


Figure 4.1.2.1: A 1% agarose gel electrophoresis in TAE Buffer of the PCR product after amplification with a DNA ladder

4.1.2 PJIN_hSGLT1 expression vector for large scale expression purposes

The process to obtain this construct presented here is the same as the one presented previously in 4.1.1 but, in this case, the vector to perform the ligation is slightly different because contains no eGFP expression sequence.

The results of the PCR reaction for the cloning are shown in the figure 4.1.2.1.

Lanes 1 and 2 represents the product of PCR after the PCR reaction of different reactions. A thick band appears in both lanes at around 2kb which is exactly the size of the hSGLT1 insert sequence. Some other bands appear at higher size, and like before, those bands could be unspecific amplicons from the PCR or the DNA template from the empty vector. Lanes 3 and 4 were an inner control of the PCR reaction. The same amount of DNA template used for the reaction is loaded in the agarose gel. A small band appears at around 6kb and corresponds to the non-linearized size of the vector.

This PCR product was used later on for the ligation and transformation. Like before, plasmid DNA isolation was done for each positive colony, which was subsequently digested with NotI and NdeI in order to validate the presence of the insert (fig. 4.1.2.2 and fig. 4.1.2.3)

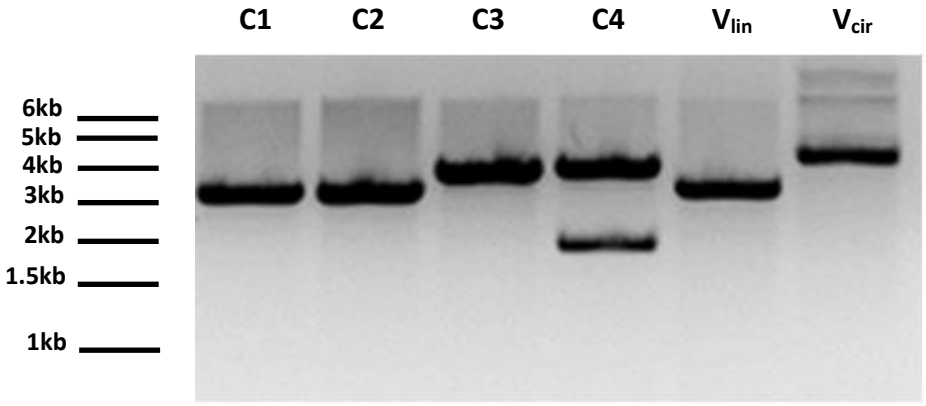


Figure 4.1.2.2: A 1% agarose gel electrophoresis in TAE Buffer of different colonies that were positive after transforming them in LB low salt with 25 µg/mL of zeocin. DNA from all lanes was digested with NdeI and NotI except lane V_{cir}. Lanes C1-C4 corresponds to digested-DNA from different transformants. Lane V_{lir} is the empty linearized vector and lane V_{cir} is the empty non-linearized vector.

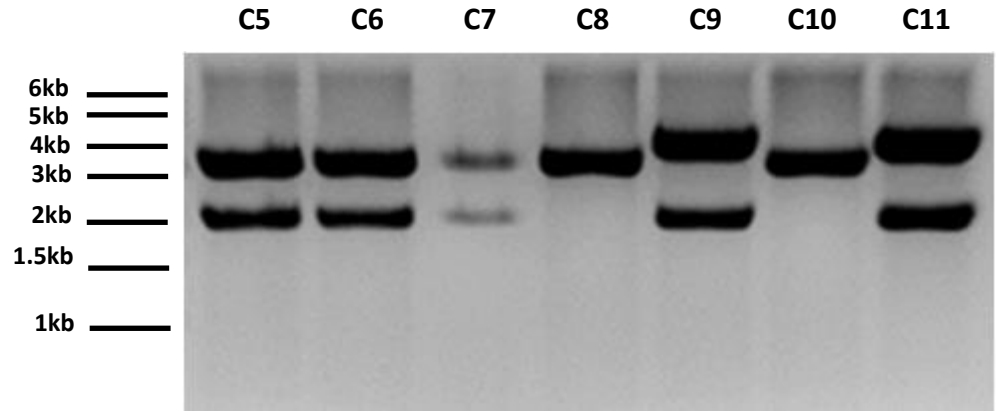


Figure 4.1.2.3: A 1% agarose gel electrophoresis in TAE Buffer of different colonies that were positive after transforming them in LB low salt with 25 µg/mL of zeocin. DNA from all lanes was digested with NdeI and NotI. Lanes C1-C7 corresponds to digested-DNA from different transformants.

In figure 4.1.2.2 and 4.1.2.3, we can confirm the presence of several positive clones because the size of the insert (C4, C5, C6, C7, C9, and C11), but some of them had a suspicious larger size for the vector (C4, C9, and C11). Only clones C5, C6, and C7 satisfied both insert and vector size criteria. C5 and C6 were subjected to DNA sequencing, which turned out to be correct.

4.2 Transformation: Electroporation

Electroporation was carried out for both constructs. The initial step, as it was described in materials and methods, was the digestion with PmeI. Since electroporation requires a lot of DNA, it's very likely that not all the vector is fully digested. Non-linearized DNA electroporation is discouraged because the genome integration efficiency is very poor (**83, 84**).

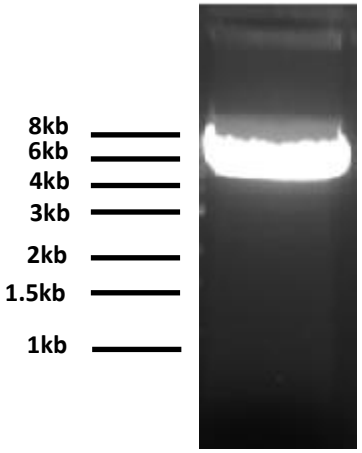


Figure 4.2.1: Agarose gel electrophoresis at 1% in TAE Buffer of the pJIN+eGFP (hSGLT1) vector after digestion with PmeI at 37^{o}C over night.

The PmeI digestion product can be used directly for electroporation or the whole digestion product can be loaded in an agarose gel and the linearized product further purified. A gel picture of a PmeI-digested pJIN+hSGLT1-eGFP product in a 1% agarose-TAE gel is shown in the figure 4.2.1. This image is a representation of the amount of DNA used in the digestion and, although the gel should have run more in order to differentiate the circular vector from the noncircular, we tried to avoid UVA exposure as much as possible to prevent pyrimidine dimers formation (**85**). The big thicker band in the figure 4.2.1 is the linearized vector (5.3kb) and, since the amount of DNA is very high, it is difficult to assess the exact size of this band. An additional thinner band appears at higher molecular weight (figure 4.2.1), matching the circular DNA form from the vector (non-digested with PmeI).

When the gel ran longer the linear DNA band separated clearly from the circular DNA band, and the linear DNA band was excised from the gel. As it can be seen, the percentage of non-digested DNA was nearly negligible compared to the digested DNA. Consequently, when pJIN+hSGLT1 vector was digested with PmeI, the whole digested DNA was used for the electroporation without further purification. Electroporated cells were platted in YPD at two increasing concentrations of zeocin: 100 µg/mL and 500 µg/mL. Antibiotic stringency is a way to select higher electroporation efficiency, which may lead to multiple genomic integration events of the insert of interest. For every transformation, a control was used with the same strain (SMD1168H) which is planned to transform but without DNA. An example of an electroporation result is shown in figure 4.2.2. The

negative control plate shows no colony growth whereas there is growth in the transformation plates with the pJIN+hSGLT1-eGFP vector.

The amount of colonies present in YPD+500 µg/mL is lower compared to the plate with 100 µg/mL zeocin. All the colonies seen in figure 4.2.2 were later on analyzed for protein expression. Transformation is a process which involves a lot of cell death because of the removal of yeast cell wall and the electroporation process itself, since a high voltage in the method (1.8 kV) is applied. It is known that this high cell death results in a layer of debris that prevents the antibiotic action, allowing antibiotic-sensitive cells to grow as well (**86**) This might be a problem because, even in negative controls, some colonies (false positives) can grow. This problem is also present with the tested colonies so it requires checking for false positives. This issue will be addressed in the next section, clone selection.

Figure 4.2.2: YPD plates with different concentrations of zeocin. Plate on the left is YPD +zeocin (100 µg/mL) for SMD1168H cells without DNA (negative control). Plate in the middle is a YPD +zeocin (500 µg/mL) with SMD1168h cells transformed with pJIN+hSGLT1-eGFP vector. The plate on the right is the same as the one in the middle but zeocin with YPD +zeocin (100 µg/mL).

4.3 Clone selection

The clone selection process is one of the key and limiting steps when expressing any protein in *P. pastoris* (**87**) since not all transformants behave in the same manner in relation to expression levels, trafficking and others. It is suggested in the literature that around 50 to 100 colonies should be screened in order to find the best clone (**64**). There are actually several methods to detect protein expression but, most of them, are indirect methods. The approaches used in this section can vary depending on the construct of expression we worked with. The presence of eGFP in the construct facilitated the screening.

The approach consisted in using several techniques with the aim of narrowing down the search. Some of these techniques were chosen because they were faster and easier to handle and they

helped us to discard at an early stage some of the colonies which, in some cases, were false positives.

4.3.1 Serial dilution: Viability in zeocin

This method was used for both constructs: pJIN+hSGLT1-eGFP and pJIN+hSGLT1 Following the protocol explained in material and methods (section 3.2.3.1). The results presented below are for the pJIN+hSGLT1-eGFP construct and colonies were screened in YPD + zeocin 100 µg/mL and 500 µg/mL. Although around 100 colonies were screened, not all colonies are shown in the figure 4.3.1.

Each plate has a negative internal control of the empty strain SMD1168h which did not grow in any of the plates. Nearly all the colonies tested grew very well in YPD + zeocin 100 µg/mL, except some false positives, but not in 500 µg/mL zeocin. The colonies that did not grow in 500µg /mL of zeocin were automatically discarded because as it was mentioned before, colonies with higher antibiotic resistant are good candidates to have higher expression levels due to multicopy DNA integration. It is always important to compare the growth in 500µg/mL of zeocin to the growth in 100µg/mL because, if the colonies grow well in both plates, it's probably a sign of multicopy DNA integration. As it can be seen in the figure 4.3.1., most of those colonies grow nearly the same in high or low zeocin concentration. In this step and for further methods, we discarded already more than half of the colonies and we selected the ones that grew in YPD+500 µg/mL.

4.3.2 Colony PCR

The colony PCR technique, described in 3.2.3.2, consists of carrying out a fast PCR as an initial screening previous to the application of any other method. The fast PCR rationale is the higher number of insert copies in the genome, the higher DNA amplification. However, this method will not allow comparing different colonies, so it's only recommended as an initial step to verify that the colonies have integrated the plasmid, so false positive colonies can be discarded. Figure 4.3.2.1 is a representation of a colony PCR from several colonies from the electroporation plate. Two controls are needed for the colony PCR technique. In the first control (Con.1), the pJIN+hSGLT1-eGFP vector was used as a template DNA instead of a colony and, in these conditions, a band corresponding to our insert should be expected at around 2kb.

As shown in the figure, a band appears near 2kb but other bands appear at lower molecular weights, which could be due to unspecific amplifications. This control is like a positive control for the PCR reaction, however, since this control was purified DNA, another control is needed where a colony is used instead; this is Con.2 (control 2). For this control, an *E. coli* colony from a regular transformation (heat shock) with the pJIN+hSGLT1-eGFP was used. The amplification in that control should be the same as Con.1 since the *E. coli* colonies contains also the pJIN vector, but, as shown in the figure 4.3.2.1, the amplification was not the same. Two bands appear in Con.2: One of these bands is near 3kb and the other one at 0.5kb, which is the same unspecific band seen in

Con.1. The band around 3kb should appear at 2kb instead if it corresponded to hSGLT1. So, somehow, the amplificon was longer. These differences prove that Con.2 is not useful as a colony PCR control although, in theory, should work. When transforming *E. coli*, the exogenous DNA vector does not integrate in the genome like in *P. pastoris* or other yeast (**88, 89**) and those differences could explain why the amplification was not the same from one to the other.

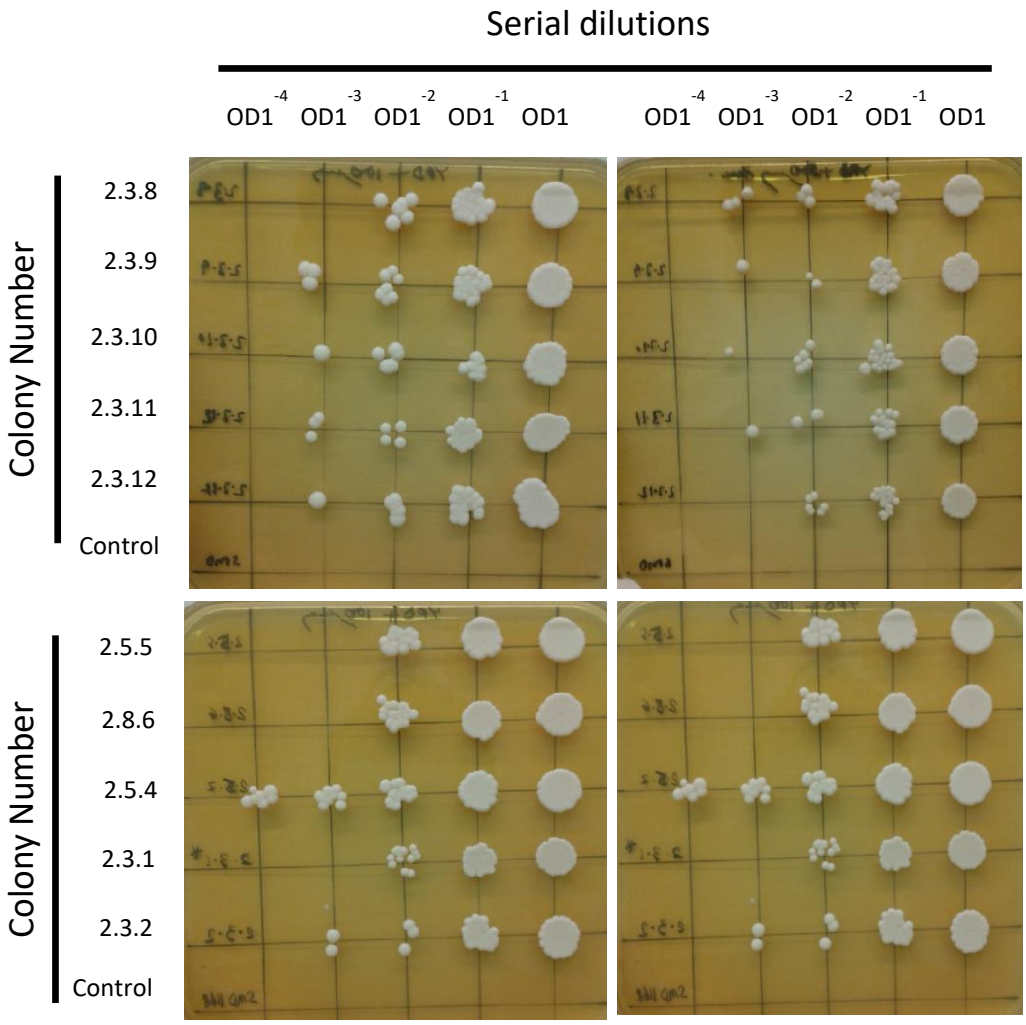


Figure 4.3.1: YPD plates with different concentrations of zeocin after 2-3 days of growth. Plates on the left are YPD +100 µg/mL of zeocin of several colonies to test; Plate on the rights is the same as the one in the left but in YPD +500 µg/mL of zeocin. Each number represents a colony from the electroporation plate. A dilution factor is 10X for each lane, starting from the right to the left.

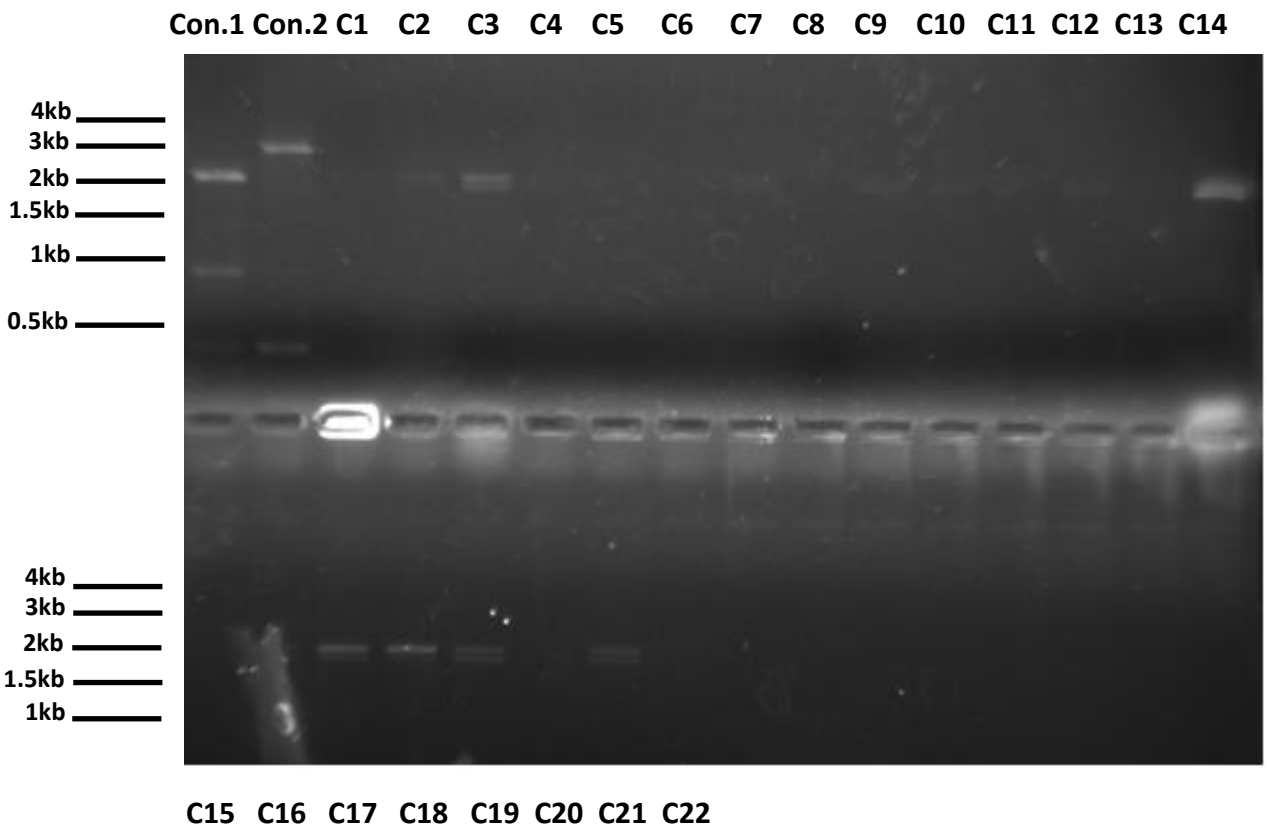


Figure 4.3.2.1: Colony PCR from several colonies from the electroporation plate. L: ladder;Con.1 (control1): DNA template used was the pJIN+eGFP (hSGLT1) vector without a colony; Con.2 (control2): *E.coli* colony from a regular transformation (heat shock) with pJIN+eGFP (hSGLT1); C1-C22: several colonies from the electroporation plate.

The rest of the lanes shown in figure 4.3.2.1 are from different electroporated colonies. A band around 2 kb can be seen for some of the colonies (C) like for example C4, C14, C17 and C18. The band at 2 kb appears as a double band in all the cases, but, in some of those like C17 and C18, the upper band is more intense and, in other cases, both bands have the same intensity.From these two bands, the upper band at 2k.b appears at the same size as the Con.1 which is exactly the size of the hSGLT1 insert. The lower band at 2k.b could be just a different topology of the same amplicon or, also, an unfinished amplification of the insert.

It can be concluded that the amplification worked correctly regardless of Con.2 and that the amplification in all tested colonies gives us two bands instead of one. These results only prove that the transformation worked and that the colonies integrated the insert but cannot discriminate which colonies integrated more or less copies of it.

4.3.3 Plate induction

This technique is exclusively for the pJIN+hSGLT1-eGFP construct and, by far, is the more informative method because it gives a direct idea of the expression levels instead of gene dosage (number of integrated copies of the vector). Although the number of copies integrated in the transformation process is very important, it is known that the expression levels depend also on other factors which are not clearly understood and can vary a lot (**90, 91**).

The plate induction method allows comparing different levels of expression from around 50 to 100 different colonies in one or two methanol 15 mm square plates which will speed up the colony screening (**68, 64**). The key step for this method is the normalization of the amount of cells plated or, otherwise, comparison will be pointless. Also, the methanol plates were left for 48 hours because the fluorescence at 24 hours was nearly the same for all colonies. For more details on the method see section 3.2.3.4. Below, in figure 4.3.3.1, a methanol plate after 48 hours of induction is shown.

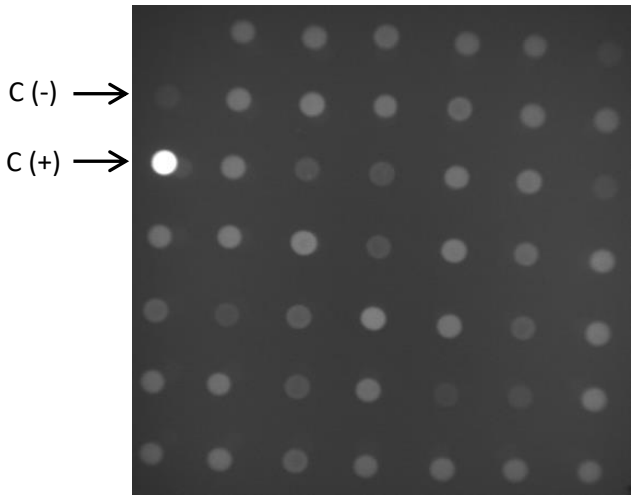


Figure 4.3.3.1: Minimal Methanol plates (MM) after 48 hours at 30ºC. Each dot represents a different colony from the electroporation plate except the controls as pointed out in the picture. C (-) is a SMD11668h colony without vector. C (+) represents a colony transformed with an empty vector which expresses free eGFP.

As it can be seen in the figure, several colonies are more fluorescent including C (-) and C (+). The fluorescence seen in C (-) is due to the intrinsic fluorescence of the colony itself. The high fluorescence seen in C (+) is mainly due to free GFP. This control is very important because it proves that methanol is actually activating the expression of the AOX gene. The rest of the dots including the controls were later image analyzed by densitometry.

The densitometry values were later on represented after subtracting the signal from the background C (-). The positive control is not shown because the signal is too high compared to the rest of the colonies and the value was out of scale. Below, in figure 4.3.3.2, a densitometry of a methanol induction plate is shown. As one can observe in figure 4.3.3.2, some colonies have no expression at all, such as 1.1.2 and 2.6.2, with high expression levels.

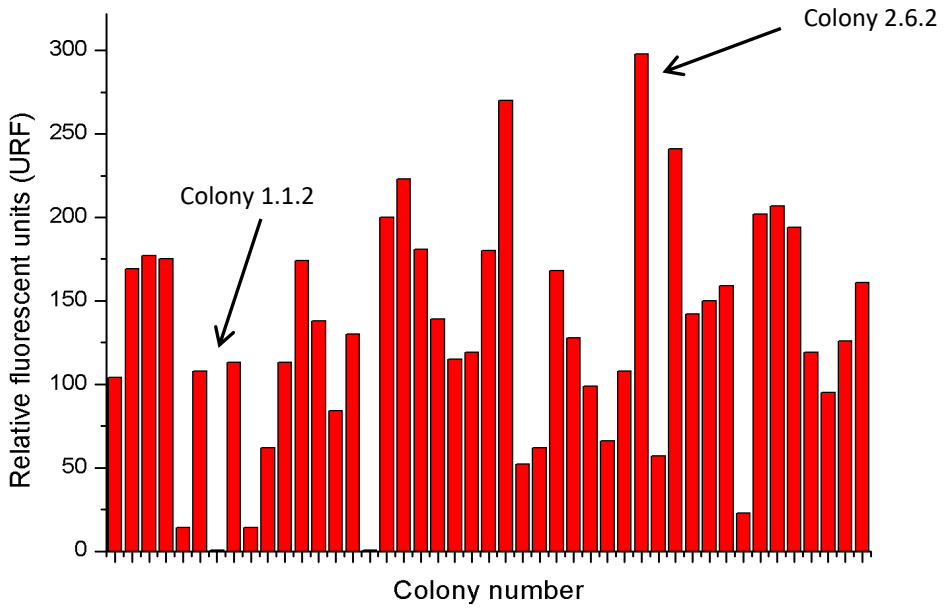


Figure 4.3.3.2: Representative densitometry values of some of the colonies from figure 4.3.3.1 after removing the signal from the negative control. Each colony is represented with an arbitrary number. Arrow points out some specific tested colonies.

It's interesting to remark that, with this technique, we have no idea of how much free GFP is contributing to the fluorescence signal, so, even when the fluorescence values are high, we will not know if the chosen colony is the right one for protein expression.

The colonies with a high level of expression were used to repeat this plate induction method and were represented in the same way as in figure 4.3.3.2. The result is shown in figure 4.3.3.3.

4.3.4 Genomic screening: Semi quantitative end point PCR

In this method, the DNA template used for the PCR was a purified DNA instead of using a whole colony directly as a template. The main advantage of doing a genomic DNA extraction is that the amount of DNA used as a template for one colony to another can be normalized which, in the colony PCR, is impossible. This allows comparing the amplification from different colonies because the intensity of the amplified band is correlated to the amount of transformed vector copies, only if the amplification process is in the growth phase. For this reason the number of cycles of PCR must be determined to avoid the saturation phase, otherwise the PCR signal will be equal for all reactions (**71**).

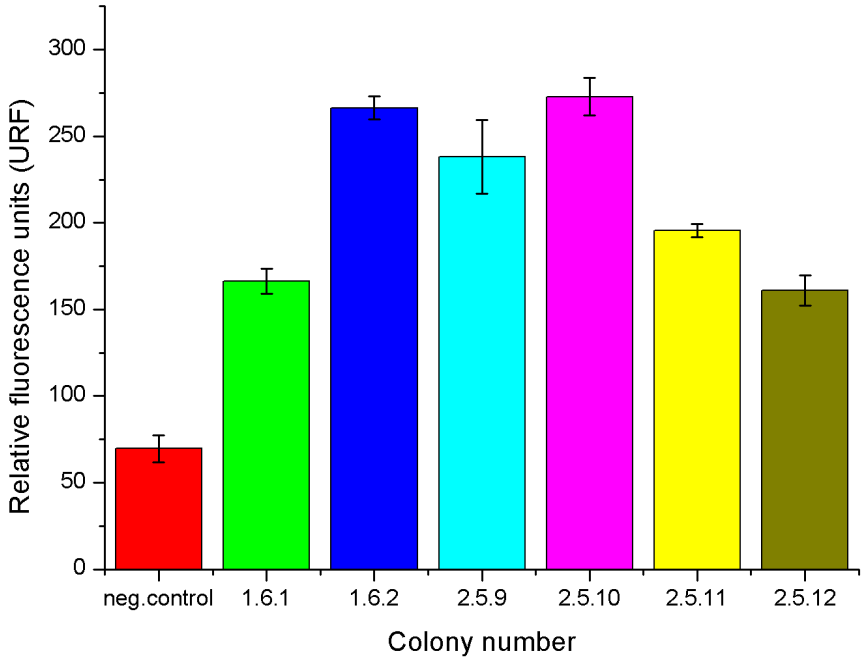


Figure 4.3.3.3: Representative densitometry values of few colonies. Colonies represented here were already selected for having higher fluorescence levels in the methanol induction plate. Error bars are represented for each colony.

An initial PCR was set up exactly as described for the colony PCR and a small aliquot was taken out at 18, 20, 22 and 24 cycles, for more details see section 2.2.3.5. The template DNA used for the PCR came from a genomic extraction of a transformed colony which turned out to be positive in the colony PCR. An agarose gel was run with all the aliquots from several cycles, as it can be seen in figure 4.3.4.1.

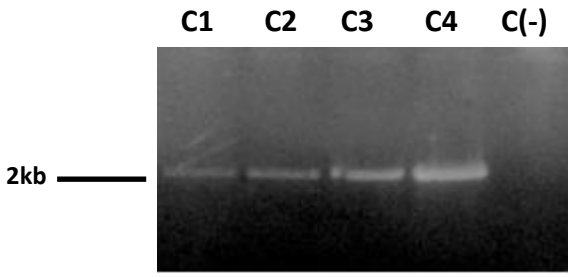


Figure 4.3.4.1: Agarose gel of hSGLT1 PCR amplification at different cycles. L: marker; C1: 18 cycles; C2: 20 cycles; C3: 22 cycles; C4: 24 cycles; C (-): negative control.

The amplification product of hSGLT1 at 2kb increases in intensity at longer cycles, as expected, and the negative control done with non-transformed hSGLT1 gave no amplification product. The number of cycles tested was not enough to appreciate the leveling off stage; nevertheless, the analysis of the bands in the gel ensures that after 22 cycles the reaction is still in the exponential

stage which is the main objective of this PCR. So, 22 cycles were used instead of 30 cycles for the semiquantitatve PCR.

Colonies which tested positive in the colony PCR, grew well in YPD + 500 µg/mL and gave a good expression in the plate induction where analyzed by the semiquantitative PCR technique. Results are shown in the figure 4.3.4.2.

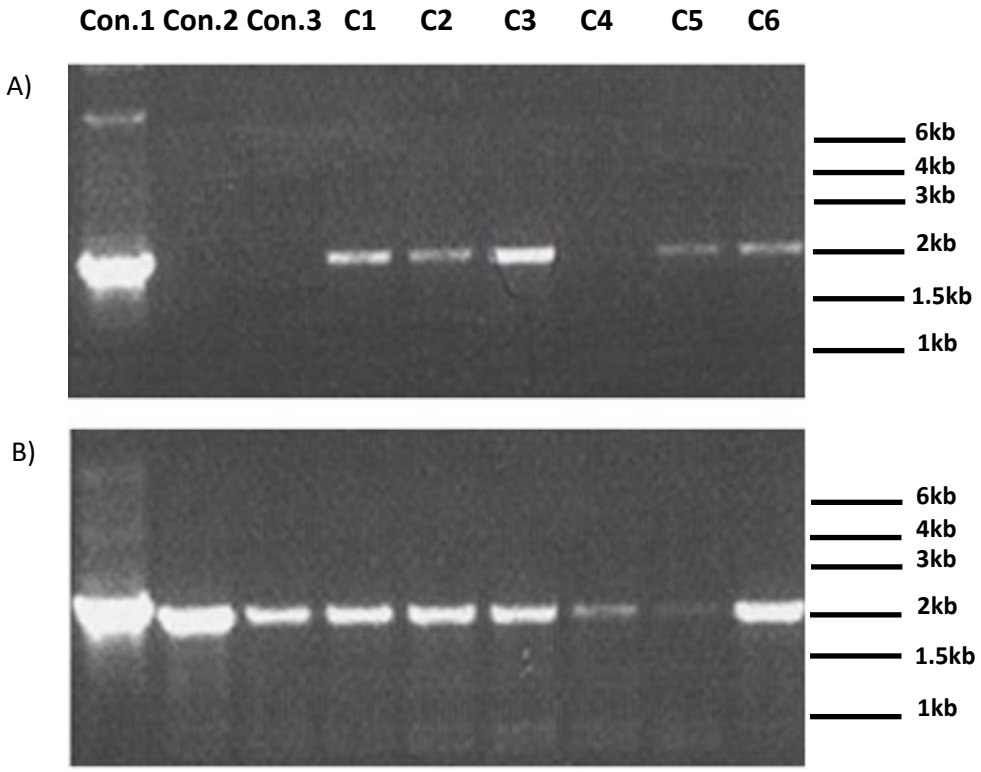


Figure 4.3.4.2: PCR of a semiquantitative PCR: A) Amplification (PCR) with the cloning primers B) Amplification with the AOX primers. Con.1: the template DNA used was the pJIN_hSGLT1+eGFP construct; Con.2: genomic extraction of a non-transformed SMD1168h colony (negative control); Con.3: genomic extraction of a transformed SMD1168h with the pJIN empty vector; C1 to C6 are different amplifications from several colonies from the electroporation plate.

Two amplifications were carried out for this PCR: With the cloning primers, figure 4.3.4.2(A), and with the AOX primers, figure 4.3.4.2(B). Three controls were used in both cases: Con.1 was amplified with the whole pJIN_hSGLT1+eGFP vector as a DNA template. For Con.2, a genomic extracted DNA from a non-transformed SMD1168h was used as a DNA template which was going to be used as a negative control. For the last control, Con.3, the DNA template used came from a genomic extraction of a transformed SMD1168h with the pJIN empty vector.

In the figure 4.3.4.2(A), amplification which corresponds to hSGLT1 is seen as a single band at 2kb in Con.1 but not in Con.2 and Con.3. The amplification with those primers is expected to be seen in

control 1 because the vector itself (pJIN_hSGLT1+eGFP) was generated with the same primers. However, in Con.2 and Con.3 those primers cannot amplify the insert because, in both cases, the genomic DNA from SMD1168h and the empty vector have no targets for those primers. If bands were to appear in Con.2 and Con.3, that would be due to primer unspecificity, which proved to be non-existent. Regarding to the tested colonies, all of them, except colony 4, presented a single band corresponding to hSGLT1 at 2kb. The intensity of the band at 2kb is not the same for all the screened colonies, a fact which is clearly showing different levels of vector insertion.

In the figure 4.3.4.2(B), amplification is seen in all lanes with a band at 2 kb although other faint bands appear at lower molecular weight near 1 kb due to primer unspecificity. The target gene for this PCR is the Alcohol oxidase gene (AOX). The AOX gene must be present in all genomic extractions of *P. pastoris* as explained in the introduction so, even the negative control, Con.2, has a target for the AOX primers. Regarding the tested colonies, all of them present a band at 2kb and, all of them, except C5 and C4, have approximately the same intensity.

Several conclusions can be drawn from figure 4.3.4.2. First, the AOX primers cannot be used for the semiquantitative PCR because of the lack of a reliable negative control. Also, the intensity of the amplified bands from colonies on figure 4.3.4.2(A) and 4.3.4.2 (B) are very different. Precisely, in figure B, all the PCR products present high intensity due probably to faster amplification, so, checking the right number of cycles for the semiquantitative PCR (figure 4.3.4.1) depends on each primer and should be used afterwards with the same primer. It is very likely that in figure 4.3.4.2(B) the PCR is near the saturation phase (instead of the exponential phase like in figure 4.3.4.2(A) with the cloning primers). Second, the technique was very successful for the cloning primers because clear different intensities can be appreciated in figure 4.3.4.2(A) for different colonies.

Finally, the results obtained with this technique are compatible and paired with some the other techniques reported above, such as plate induction, because the best candidates resulting from the application of both techniques tend to be the same colonies.

4.3.5 Western blot

This method was used mainly for the pJIN_hSGLT1 construct. Although this technique can be also used for the GFP construct, it is not recommended because the plate induction method is faster and easier. It is important to note the reproducibility issues of western blot methodology, thus it is difficult to do quantitative assessments when comparing two different experiments. There is one advantage of the western blot technique compared to the induction methanol method: the SDS-page can separate the free eGFP from the hSGLT1+eGFP, which is not possible in the induction methanol method.

In figure 4.3.5.1, an immunoblot is shown for several pre-screened colonies. The amount of protein loaded in the gel was normalized for all cases. A small culture was prepared as described in

the section 3.2.5 and cells were broke down with glass beads as described in 3.2.7. Membranes were washed once with 4M of urea to remove peripheral proteins (**65**) and an SDS-PAGE was run with those samples.

In figure 4.3.5.1, a strong band at around 50- 60 kDa appears in all cases except for the negative control. This band corresponds to the apparent size of hSGLT1 seen in an SDS-PAGE. (**92**, **65**) Nevertheless, other bands and smears are also observed. A faint band appears at around 25 kDA which could be either an impurity or a proteolytic fragment of hSGLT1. Technically, if the antibody was specific no noise from other bands should be seen, but that shouldn't be the case, because the antibody is polyclonal. During the extraction, some proteolysis can occur. If this was the case, proteolysis would have occurred with a different grade of intensity from one sample to another. For example, bands at lower MW in lane 4 are much more intense than in lane 7. Also, a big smear is present in all samples near the stacking and in the stacking gel, except the negative control which is nearly nonexistent. Since the negative control has nearly no signal, the smear must be due to the presence of hSGLT1 although not in its monomeric form. Several possible explanations can be given. It is known that hSGLT1 may aggregate in an SDS-PAGE (**93**) so it's possible to expect some protein entering poorly in the running gel or not at all. Also, it's known that samples enriched with a high concentration of lipids can affect the protein solubilization, and impair proteins to run correctly in the gel (**94**). Another reasonable option is that a proportion of the total hSGLT1 protein is in an oligomeric state before or after the extraction. All of the previous reasons could as well help explaining why some hSGLT1 does not appear at the right apparent molecular size.

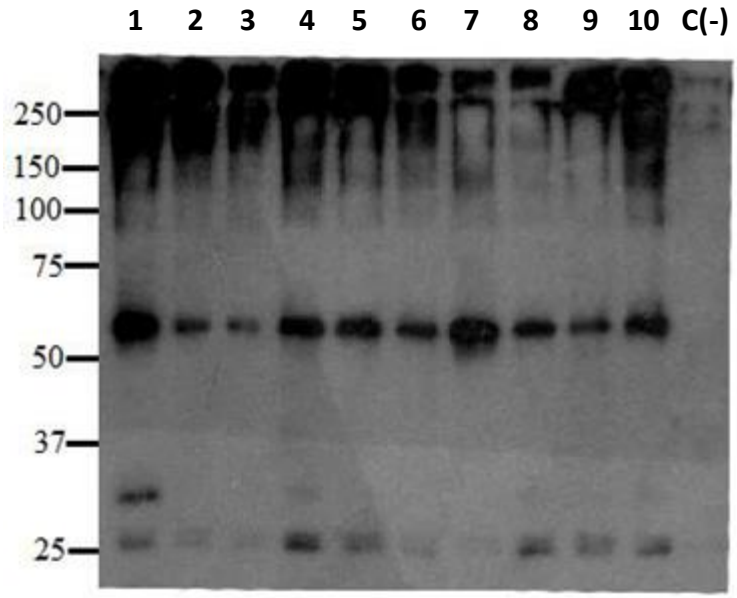


Figure 4.3.5.1: Western blot analysis by anti-hSGLT1 antibody of the screened positive colonies for the pJIN_hSGLT1 vector Lanes 1-10 are membrane fractions extracted from the candidate clones after induction with methanol containing media and washed with 4 M Urea to remove peripheral proteins. C (-) is a negative control that corresponds to the stripped membrane fraction obtained from a strain transformed with an empty vector.

In summary, there are three parameters for selecting one colony from the western blot results: i) the amount and intensity of bands possibly corresponding to proteolysis; ii) the intensity of the hSGLT1 band in the monomeric state (50-60kDA), and iii) the presence of a smear at higher molecular weights.

4.4 Expression optimization: Screening conditions

As explained in the introduction, *P. pastoris* is able to grow in a wide range of conditions (**95**). The main aim of this project is to obtain a highly pure protein in suitable amounts for structural and functional studies, thus it is mandatory to screen the best conditions for protein expression. For this purpose the pJIN_hSGLT1+eGFP construct was used. For more details on how the screening conditions were carried out see section 3.2.4 in the material and methods section.

4.4.1 Media and temperature screenings

Different media can be used for protein expression in *P. pastoris* but the most commonly used are Minimal Methanol Medium (MM), Buffered Minimal Methanol Medium (BMM) and Buffered Methanol Complex Medium (BMY) (**64, 65, 95**). These media were used for induction since all of them contain methanol, and the initial growing was done with the same respective medium but with glycerol instead of methanol at 30°C. Normally, *P. pastoris* has an optimal growth at 28-30ºC but, for protein expression, lowering the temperature has proven a way to increase the total expression levels (**96**). Two temperatures were chosen: 30°C and 24°C.

Figure 4.4.1.1 shows the results from expressing the protein in different media and at different temperature using the pJIN_hSGLT1+eGFP construct. Expression at 24°C is higher than at 30°C for all tested media. Buffered media presents higher growth at 24°C compared to 30°C. So, 24°C were selected for protein expression instead of 30^{o}C. Other temperatures such as 18ºC were tried, but no significant differences were observed (data not shown).

For the minimal medium at 24°C, maximum of expression occurs at around 18-20 hours and it decays afterwards. Also, there's a clear difference between 30°C and 24°C for the minimal medium. On the other hand, the buffered media has a maximum of expression at 25 hours, approximately, and the expression decays steadily compared to the non-buffered minimal media.
Since the goal is to get as much as protein as possible, the buffered media has been considered to be more suitable for expression although minimal media can also be used for faster protein expression. *P. pastoris* metabolism at high OD is known to acidify the media and can affect cell survival (**64, 113**). This might explain the fluorescence decay in the non-buffered media. Also, the same reasoning might explain why the maximum of expression in the buffered media is higher than in the non-buffered media.

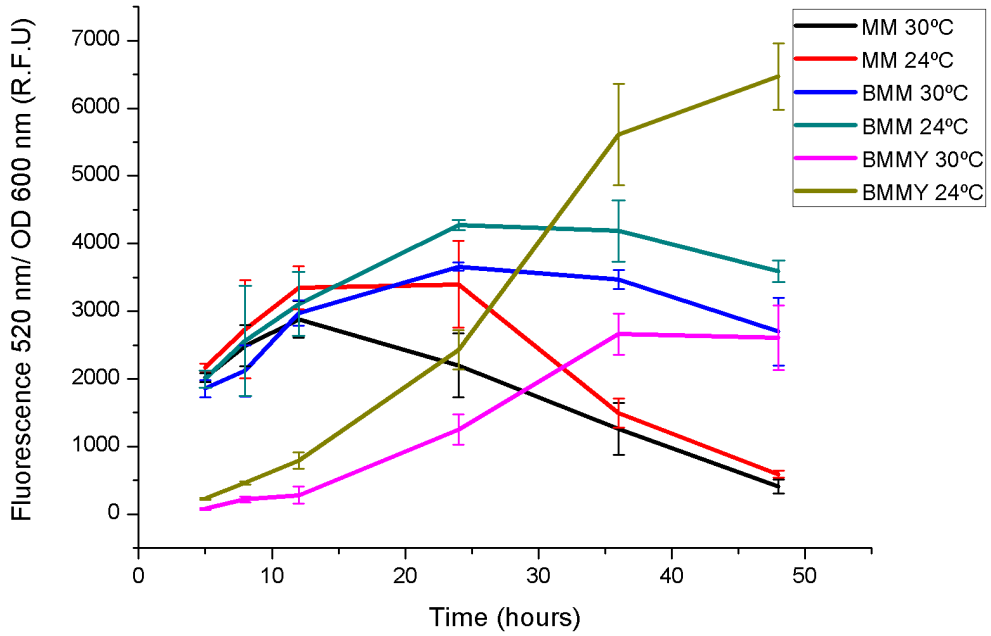


Figure 4.4.1.1: Expression of hSGLT1 in different media and at different temperatures. The fluorescence values were normalized by the optical density (OD at 600nm).

As for the complex media (BMY), protein expression takes longer to happen, with a maximum of expression at 35 hours at 30°C. No saturation point was seen at 24°C, but at 40 hours, the expression values started to stabilize. The maximum level of expression was achieved in complex media at 24°C after 40 hours. Nevertheless, BMY medium was not selected for protein expression due to several reasons: as described in the literature the expression of recombinant proteins in this media implies later problems when the protein is purified and problems related to proteolysis take place. One of the main characteristics of this media is that allows the yeast to keep growing after the induction, while with other media the OD at 600nm is stable after induction and that might explain why the protein expression takes longer time to increase (**64**). It is recommended not to have protein expression while cells grow. For instance, when the constitutive promoter, GAP, (which allows for protein expression during cell growth) has been used (**66**) expression problems have been reported. It's also important to remark that, after 24 hours, more methanol (0.5%) was added to all cultures to get more expression, but surprisingly only with the complex media the fluorescence increased.

4.4.2 Optical density (OD)

As mentioned before, *P. pastoris* cultures can grow up to very high optical density values (**97, 98**). Since protein expression depends on the value of the OD at which it is induced, the objective in this section is to study the level of protein expression as a function of the OD values. A selected colony with the pJIN_hSGLT1+eGFP construct inserted was used. Three OD values were chosen for screening: 5, 10 and 20. Before induction, the initial culture was split and diluted to the desired final OD. These experiments were done using MM media. In figures 4.4.2.1 and 4.4.2.2, OD values are represented as a function of time at 24°C and 30°C.

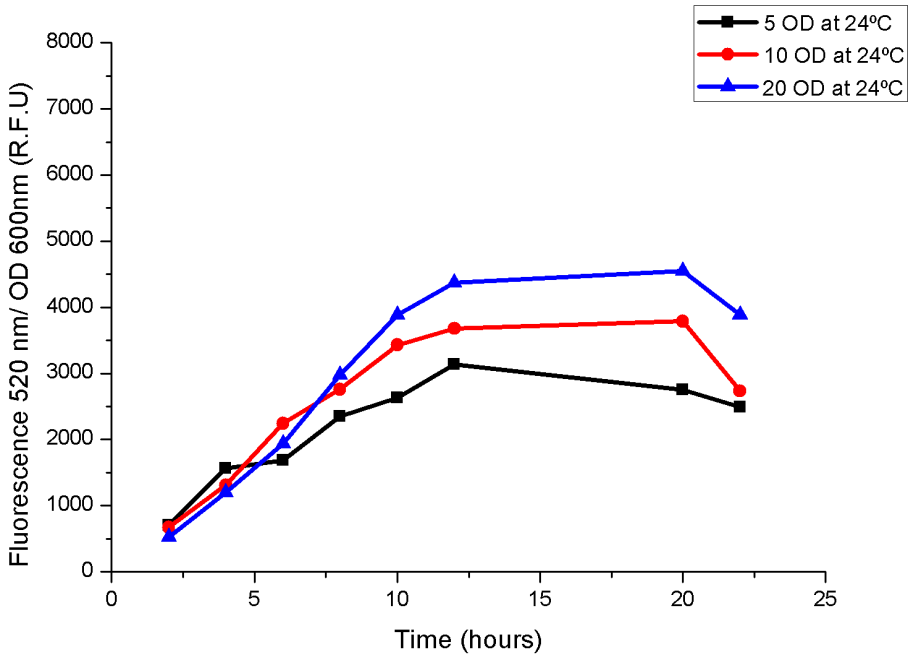


Figure 4.4.2.1: Graphic of the expression of hSGLT1 at different OD at 600 nm in function of time at 24°C in MM. The fluorescence values were normalized by the optical density.

In the figures, 4.4.2.1 and 4.4.2.2, the data of the OD in function of time at 24°C and 30°C is represented.

In the figures all fluorescence values were normalized by the optical density as a way of normalizing by cell mass. Maximal expression is reached between 10 and 20 hours for both temperatures in M.M. This is in agreement with the results in section 4.4.1 and the expression is also higher at 24°C. The expression at 24°C is O.D. dependent. The maximum fluorescence value is reached at an O.D of 20 and it is lower at 10 and 5, and slightly lower at 30°C.

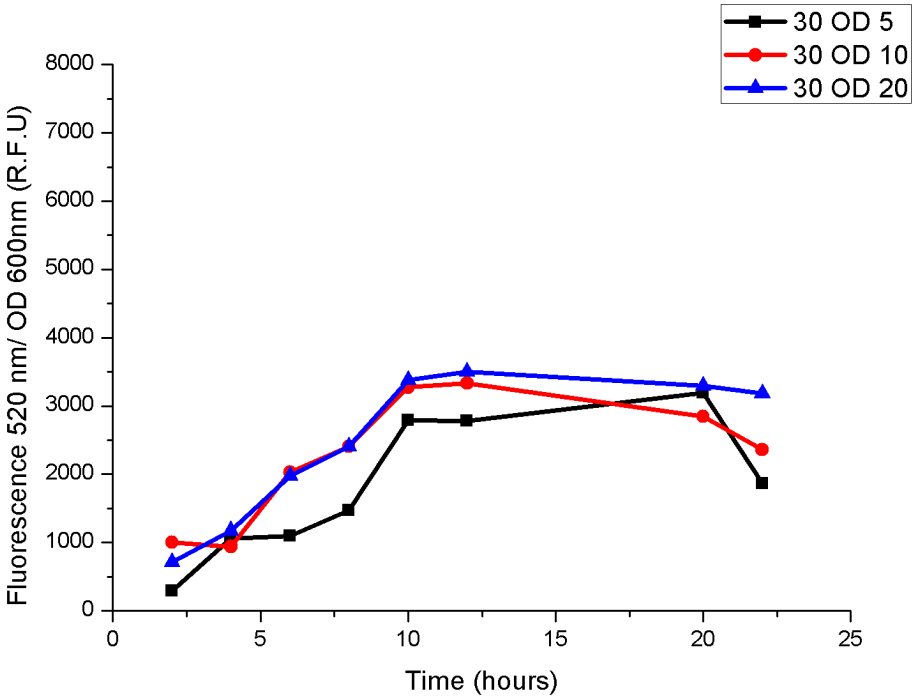


Figure 4.4.2.2: Graphic of the expression of hSGLT1 at different OD at 600 nm in function of time at 24°C in MM. The fluorescence values were normalized by the optical density.

4.4.3 Subcellular localization

The techniques applied in the previous sections served to give an idea of the total protein expression but this is still an indirect measure of expression. There are several significant questions to be addressed regarding the expression of the protein, such as the degree of proteolysis of the recombinant protein during expression, the amount of free eGFP, or the localization of the expressed protein (Endoplasmic reticulum (ER), Golgi's apparatus (GA), plasma membrane or other compartments).

One of the best ways to check protein localization is by microscopy, either with a fluorescence microscope or a confocal microscope. Although both could be used, *P. pastoris* is too small to visualize their cell compartments using a fluorescence microscopy, so, the confocal microscope was used instead due to its higher resolution (**73**). Also, the confocal microscope allows checking the fluorescence at different sections of the sample preparation. Of the three media tested in section 4.4.2, two were used for protein localization studies; MM and BMM. A negative control which expresses only soluble eGFP (empty pPCIZ vector expressed in MM media) was used. Complex medium was discarded because it interferes with the fluorescence measurements.

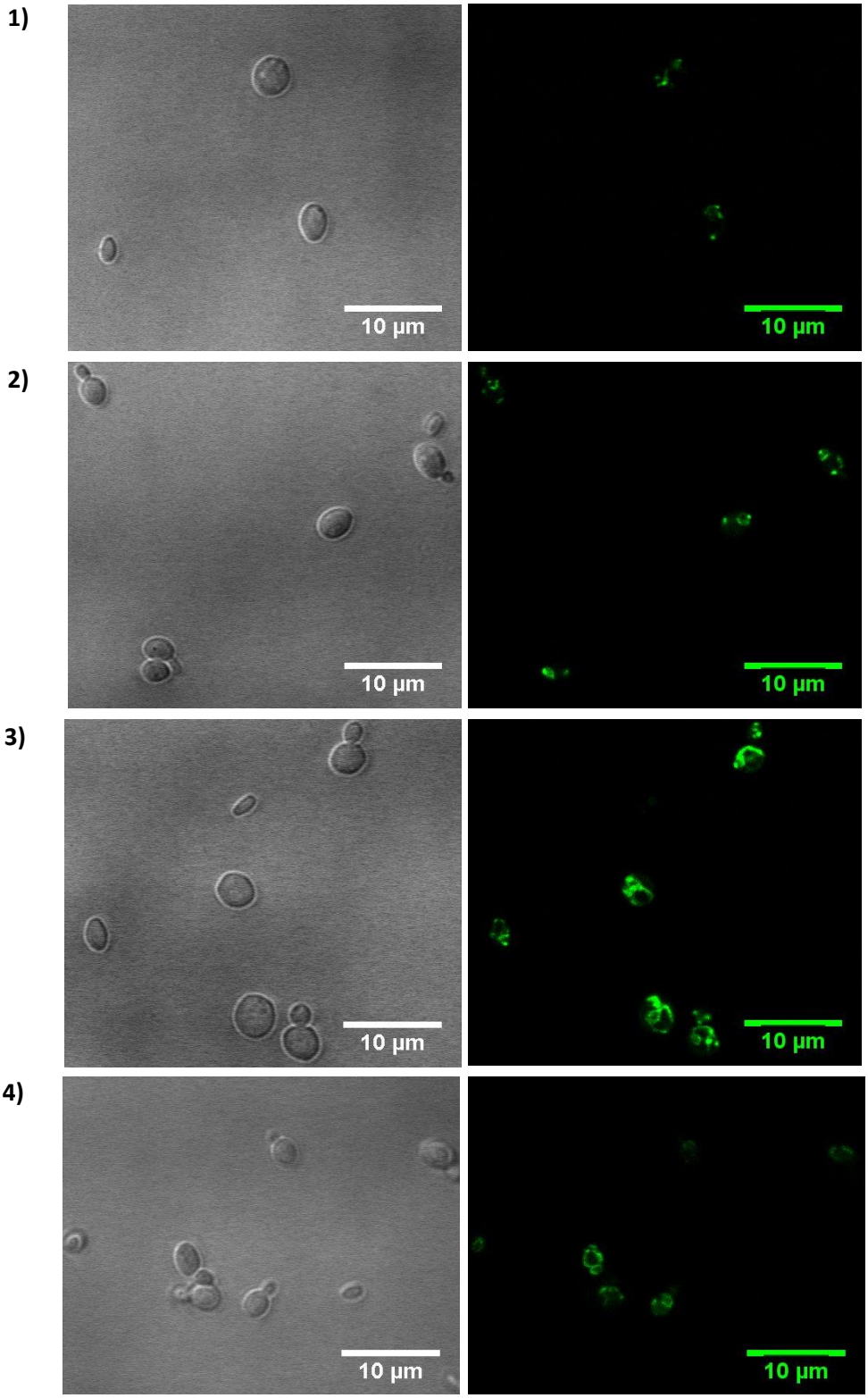


Figure 4.4.3.1: Representative field of the expression at different points of the screened pJIN_hSGLT1+ eGFP colony in B.M.M media. 1) Cells after 4 hours of induction. 2) Cells after 8 hours of induction. 3) Cells after 12 hours of induction and 4) Cells after 24 hours of induction. Bright field is on the left lane and fluorescence image is on the right.

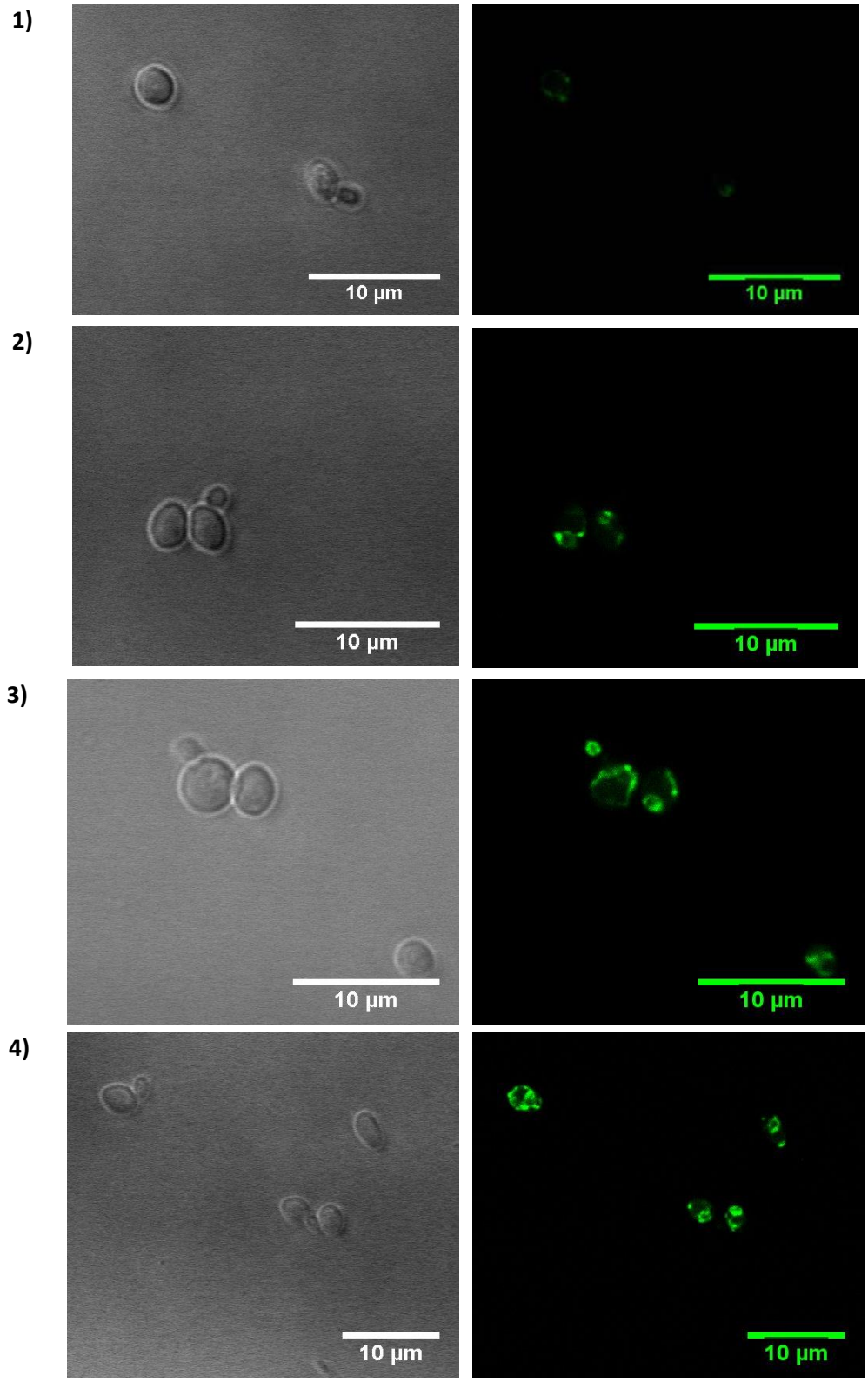


Figure 4.4.3.2: Representative field of the expression at different points of the screened pJIN_hSGLT1+ eGFP colony in.M.M media. 1) Cells after 4 hours of induction. 2) Cells after 8 hours of induction. 3) Cells after 12 hours of induction and 4) Cells after 24 hours of induction. Bright field is on the left lane and fluorescence image is on the right.

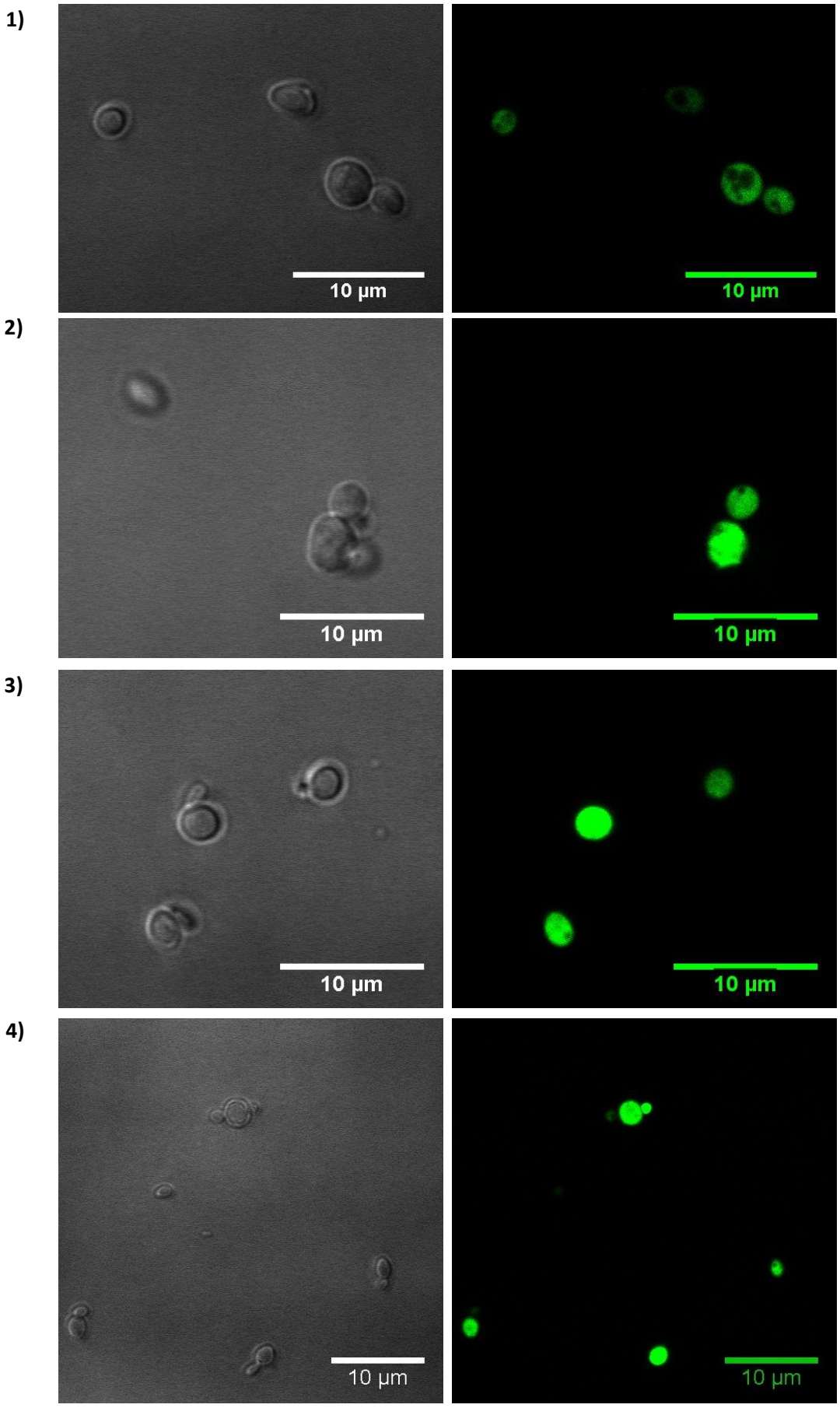


Figure 4.4.3.3: Representative field of the expression at different points of the empty pJIN_eGFP vector in M.M media. 1) Cells after 4 hours of induction. 2) Cells after 8 hours of induction. 3) Cells after 12 hours of induction and 4) Cells after 24 hours of induction. Bright field is on the left lane and fluorescence image is on the right.

For confocal microscopy measurements a small aliquot (1 mL) was taken after culture induction at 4, 8, 12, 24 and 48 hours. Cells were suspended to a final O.D of 1-2 and were flash frozen for further use at the confocal microscope. A representative field was imaged for each time point as shown in figures 4.4.3.1, 4.4.3.2 and 4.4.3.3. For each selected field two pictures were taken, the bright field image and the fluorescent image. Green cells are detected on both 4.4.3.1 and 4.4.3.2 figures. The free GFP can be clearly detected in the control sample (figure 4.4.3.3). This result shows that the selected clone expresses the protein. At 4 hours the expression is very low in both mediums (but not in the negative control expressing free eGFP (soluble). After 4 hours, the fluorescent intensity and distribution of the negative control are quite low. These results are in agreement with the results seen in section 4.4.2 because, at 4 hours, the expression of hSGLT1+eGFP should be very low. At 8 hours, the expression starts to increase for both tested media because more and brighter fluorescent dots appear in the cells (in the negative control, the fluorescence is spread throughout the whole cell at 8 hours and afterwards).

Twelve hours after induction, the fluorescence for both tested media appears as a contour line which can be cell membrane, which would indicate that the protein is expressed in the membrane. In the images, at 8 hours and 4 hours, when the protein expression was starting to take place, the fluorescents dots detected seem as well to be localized at the membrane (or membranes). At 24 hours, the image is similar to that measured at 12 hours. These results suggest that hSGLT1+eGFP is actually correctly expressed at the cell membrane and/or at internal membranes.

4.5. Protein extraction and solubilization: Screening

Once the expressing clone and the conditions for expression are selected, it's necessary to find the best way to extract the recombinant protein from the cell. This includes cell fractionating and solubilization of the desired protein.

4.5.1. Cell fractionation in the bead beater: Conditions

Several methods are described for lysing yeast and/or *P. pastoris* (**100, 74**). The most common breaking method is the mechanical method (glass beads in a bead beater) (**65, 99**) and in the present study it was carried out as described in section 3.2.7. A yield of around 40-50% of recombinant protein is extracted following this method (see figure below 4.5.1.1 for more details).

The typical extraction protocol with the glass bead beater is as follows: 10 cycles of 1 minute for cell break and 1 minute pause between cycles (**75**). The number of cycles during the cell break is a very important factor. The larger the number of cycles and the longer they are will eventually increase the probability of proteolysis and degradation of protein. Length of cycles in the present work was 1 minute of beating (cell breakage) and 2 minutes resting (instead of 1 minute) to avoid overheating the chamber. In figure 3.5.1.1, the number of cycles was tested. At 10 cycles the

recovery was lower (25%) than at 20 and 30 cycles (40-50%). The recovery was nearly the same (45%) between 20 and 30 cycles.

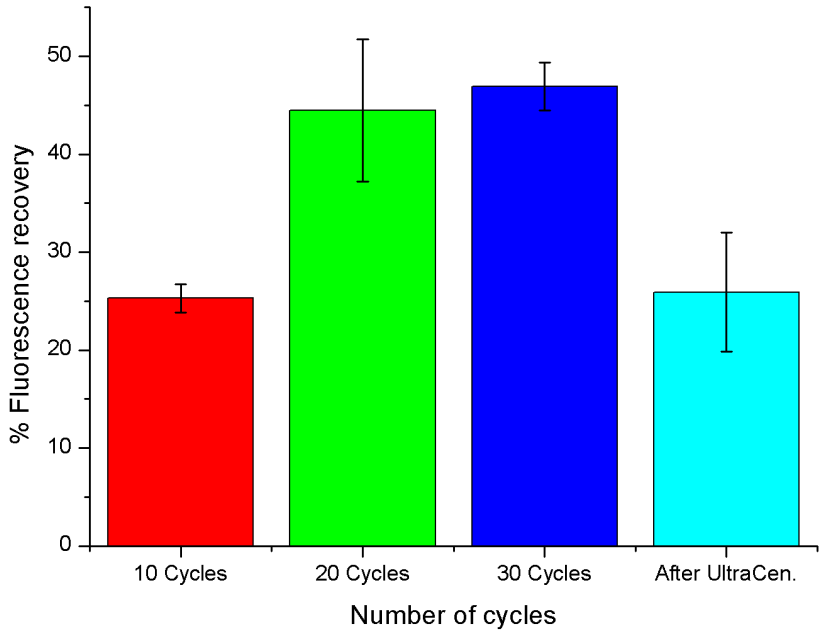


Figure 4.5.1.1: Recoveries expressed in percentage of fluorescence after the extraction with the bead beater. First steps were done at 3.000xg to remove debris while the ultracentrifugation was done at 100.000xg as a final step to recover the membranes.

This indicates that cycles which are longer than 20 aren't necessary because no more protein is extracted in the process. In order to get rid of the free eGFP signal, the cell lysate was ultracentrifugated and the supernatant discarded. Only around 25% of fluorescence was recovered by ultracentrifugation, which is a clear indication of the amount of free eGFP present in the lysate.

4.5.2 Protein solubilization: Screening

In order to purify them membrane proteins have to be solubilized with detergents in order to extract them from the membranes (**101**). In this section, different methods were used to screen the detergents, described as suitable for membrane protein purification. The construct used was hSGLT1+eGFP. The first method consisted in checking the total fluorescence in a plate reader after solubilizing the membranes with a battery of different detergents.

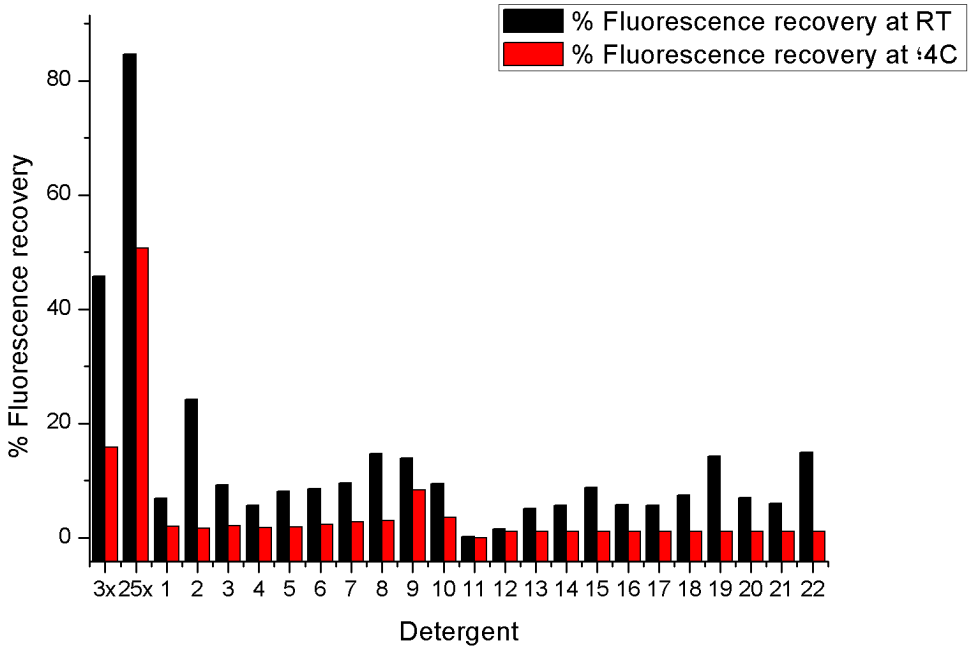


Figure 4.5.2.1: Detergent screening represented in percentage of recovered fluorescence for each tested conditions. Each number (1-22) represents a specific detergent at 3 times CMC (critical micelle concentration). First two lanes were done in F12 at 3 times CMC (3x) and 25 times CMC (25x). Two temperatures were tested: 4°C (red) and room temperature, RT (black) for all detergents. The time of solubilzation was

The fluorescence recovery was calculated by dividing the fluorescence of the pellet after centrifugation by the total fluorescence before centrifugation. Results are shown in figures 4.5.2.1 and 4.5.2.2. As shown in the figure the highest degree of solubilization is achieved when using the detergent F12. In comparison, the rest of detergents screened (see table 4.5.2) have a very low efficiency. The suitability of Fos-choline 12 (F12) has been previously described (**65,102**).

The concentration of F12 used in the literature is 25 times the Critical Micelle Concentration (CMC) (**65**). In general, the main condition to solubilize a membrane protein is to work above the CMC so, for the detergent screening we have ensured that for any given detergent the concentrations was at least 3 times the CMC. This result indicates that no other detergents work better than F12.

Surprisingly, other Fos-choline detergents (F10 and F11) solubilized hSGLT1 very poorly. Other important factors in the membrane protein solubilization process are the solubilization time and concentration (protein or detergent concentration?).hSGLT1 solubilization using FC12 was studied as a function of time using the same fluorescence method.

Table 4.5.2: List of detergents used for protein screening in figure 4.5.2.1.

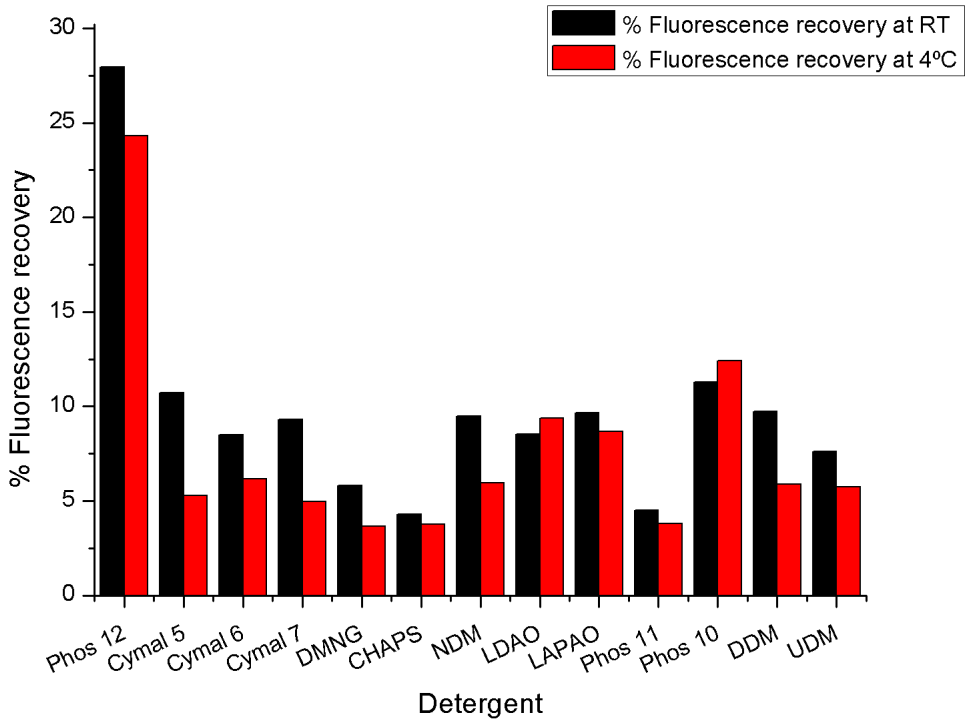


Figure 4.5.2.2: Detergent screening represented in percentage of recovered fluorescence for each tested conditions. All detergent concentrations were 3 times above CMC for room temperature (RT), in black; and 4°C in red.

As it can be observed in figure 4.5.2.3, the solubilization is very fast because the first point at time 0 (immediately after adding the detergent) gives already around 45% fluorescence recovery going up to around 60% after 2 hours. For other membrane proteins the solubilization maximum is normally achieved steadily after 1-2 hours. (**103,104,105**)

Since this time course was carried out at RT (25°), it is likely that some degradation or proteolysis could take place during this time, releasing some eGFP and increasing the fluorescence. This might explain why the solubilization after 2-3 hours starts to increase. In order to try to get some insight into this potential problem, a solubilization with F12 and other detergents were carried out at 4°C to prevent degradation or proteolysis problems. Another parameter tested was detergent concentration. Results are represented in figure 4.5.2.4.

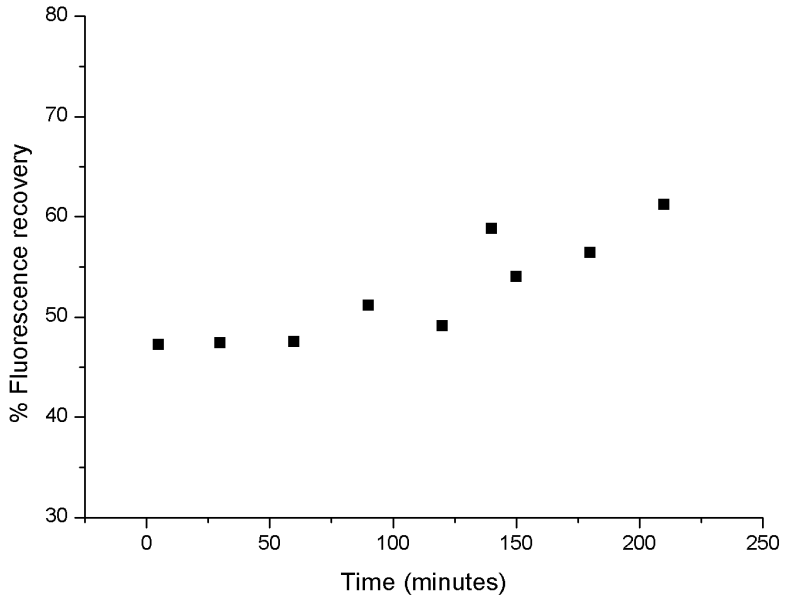


Figure 4.5.2.3: Results of a solubilization time course at RT of hSGLT1+eGFP enriched membranes. The fluorescence was measured after spinning down the solubilized membranes and comparing the supernatant and the pellet.

The maximum of solubilization is achieved at 50 times the CMC. Nevertheless the solubilization at 25 and 15 times the CMC is just slightly lower.

As mentioned above, an important aspect of the fluorescence methodology is that it measures total fluorescence and, although no free eGFP should be present in the membranes, some degradation or proteolysis can take place during the solubilization and interfere with the measure. One of the best ways to avoid this problem is checking the fluorescence in a Fluorescence Size Exclusion Chromatography (FSEC). FSEC allows separating the free eGFP (32 kDA) from the hSGLT1+eGFP (107 kDA).

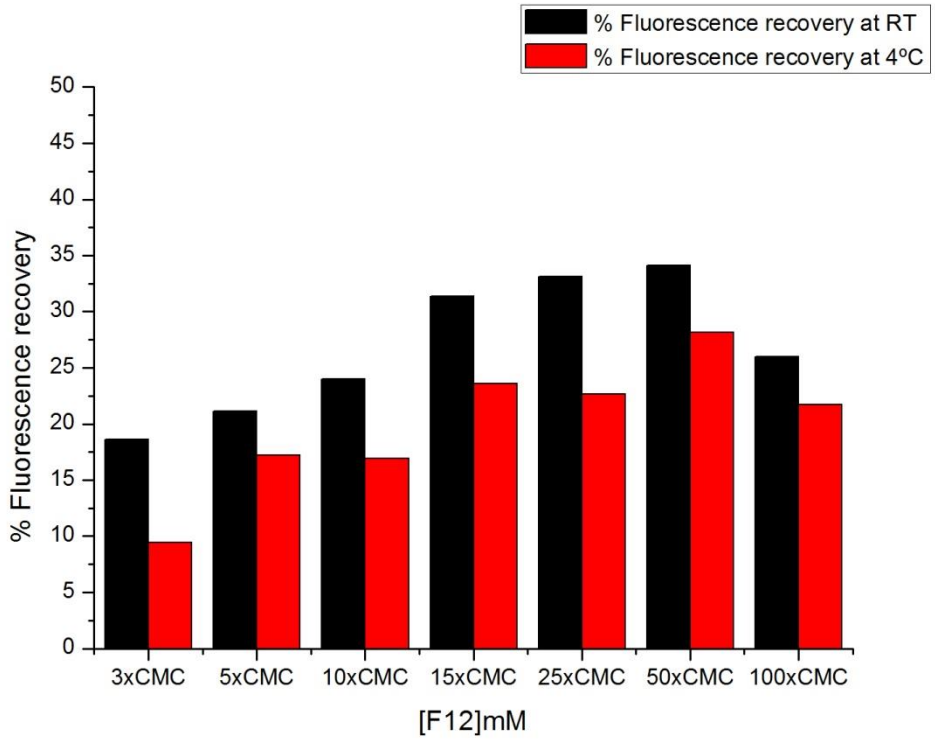


Figure 4.5.2.4: Results of the solubilization in different concentrations of F12 at 4˚C (in red) and 25˚C (in blue) overnight. The concentration values of F12 are represented in units of CMC of the detergent. The CMC of the F12 is 0.047% (w/w) or 1.5mM in water (Anatrace).

Initially, a Superose 6 column with a very broad fractionation range (Mr 5.000kDa to 5.000.000 kDa) and 25mL total volume was used. Membranes were solubilized for 2 hours with different detergents and after spinning them down at 50.000g for 20 minutes, the supernatant was loaded and run in a HPLC system connected to the column. A representative chromatogram is shown below in figure 4.5.2.5. From the three major peaks seen in the chromatogram the first peak at 15 minutes corresponds to the void volume of the column (7.5mL approximately) and is present in all samples, the second peak at 30 minutes is compatible with the size of hSGLT+eGFP (107 kDa) and the last peak at 36 minutes is compatible with the size of the free eGFP. Only in F12 a peak at 30 minutes can be appreciated which means only F12 can actually solubilize hSGLT1.

The relationship between elution time and molecular size is established in relation to a calibration chromatogram carried out with the column using reference proteins.

Surprisingly, the peak of free eGFP is present in all the solubilization trials. These results suggest that during solubilization some proteolysis takes place. In a way, this data is more accurate than the plate reader measurement because, here, the fluorescence signal that comes from free eGFP and from hSGLT1+eGFP can be differentiated. Also, this data explains why, in some cases, the solubilization values may vary from one sample to another.

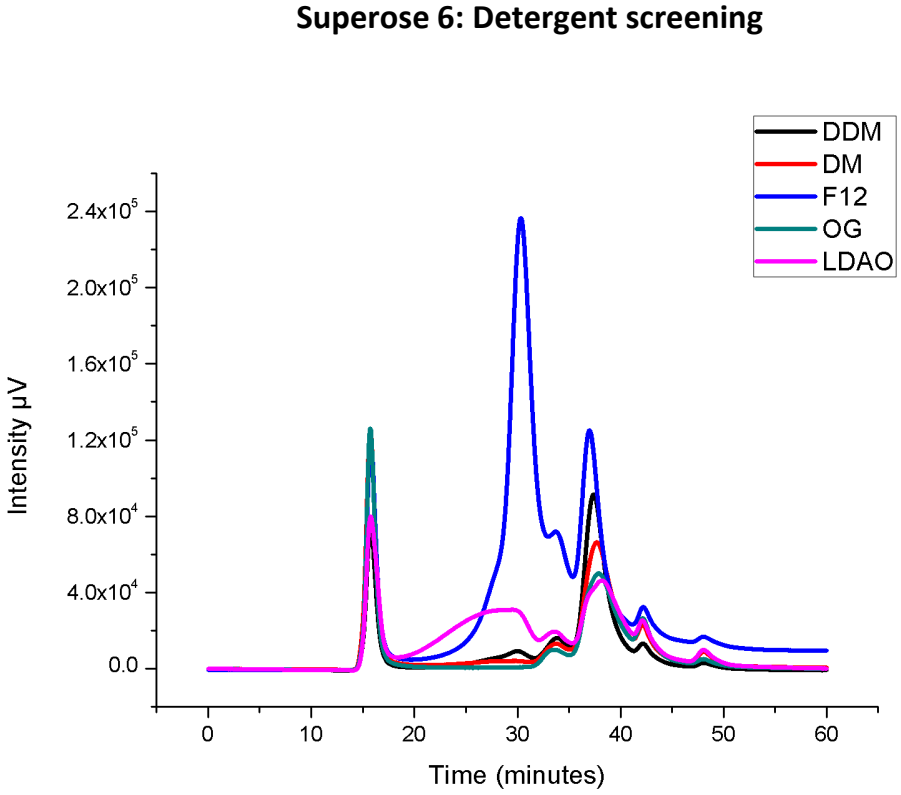


Figure 4.5.2.5: SEC chromatogram of solubilized membranes with different detergents for 2 hours. All runs were carried out in TKCL with F12 at around 3 times CMC at a flux of 0.5 mL/min. Intensity values are represented in µV units.

The Superose 6 column has poor resolution when it comes to separate proteins of similar size. Since the peak of hSGLT1+eGFP in figure 4.5.2.5 is quite wide and has a shoulder that reaches to the peak of free eGFP, a TSKG3000 column with a lower separation range (Mr 10.000 kDA to 500.000 kDA, 10 mL) was used The results are shown in figure 4.5.2.6. The column was previously calibrated in order to identify correctly the peaks. As expected, only F12 solubilized hSGLT1 properly. The peak of free eGFP is also present in this run and appears at 20 minutes as an individual peak which is what we saw in the Superose 6 runs before. Samples solubilized with detergents other than F12 gave mainly only a void volume peak (minute 10) and a free eGFP peak.

The main difference between the TSKG3000 and the Superose 6 chromatograms is that the peak corresponding to hSGLT1+eGFP, which appeared at 30 minutes in the Superose 6 column, is now less monodisperse and resolved in two peaks. One of them appears really close to the void volume, around 11 minutes, while the void appears at 10 minutes, and the other one at 14.5 minutes. The peak at 14.5 minutes corresponds approximately to the elution time of the monomer of hSGLT1+eGFP (107kDa) and the peak at 11 minutes is consistent with an hSGLT1+eGFP a tetramer. Between those two peaks, oligomer and monomer, another smaller peak at 12.4 minutes can be appreciated. It could be that this peak corresponds to another oligomeric state of

hSGLT1 such a dimer. The glucose transporter from *vibrium*, vSGLT, is actually expressed as a dimer whereas there is some dispute in the literature on whether hSGLT1 is actually expressed as a dimer or not (**36,106**, **107,108**).

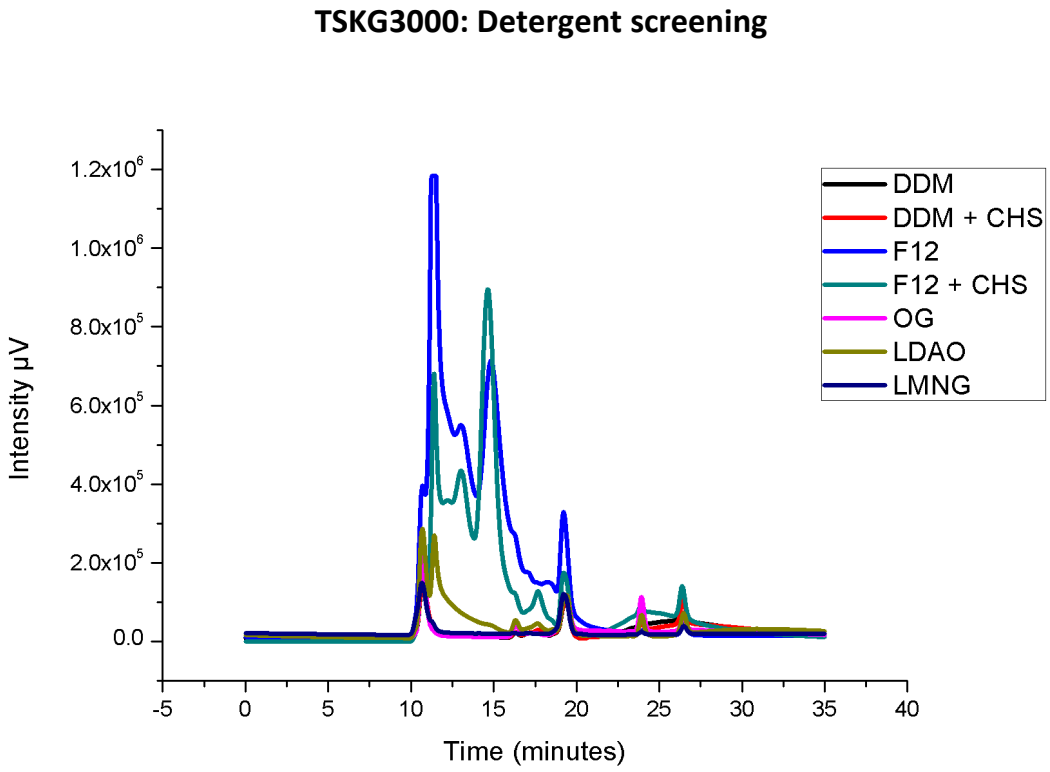


Figure 4.5.2.6: FSEC chromatogram of solubilized membranes with different detergents for 2 hours. Despite the detergent in which the sample was solubilized, all runs were done in TKCL with F12 at around 3 times CMC at a flux of 0.5 mL/min. Intensity values are represented in µV units.

FSEC measurements were extended to the study of the solubilization process as a function of F12 concentration. The results are shown in the figure below, 4.5.2.7.

At 0.1 and 0.2 % of F12 hSGLT1 is not solubilized because only 2 small peaks appear; the void volume (10 minutes) and the free eGFP (20 minutes). At F12 0.47% the solubilization of hSGLT1 takes place, as the appearance of the monomer and oligomer peaks indicates. At higher F12 concentrations the chromatograms do not change much although some details can be appreciated. There's more monomer signal at 0.7 % and 1.2% of F12 than at 0.47% but the oligomer signal is the same for all of them. No differences can be appreciated between the chromatograms at F12 0.7% and 1.2%. At higher concentrations of F12 (2.35%), the signal from the monomer is smaller but the signal from the free eGFP is higher with nearly the same intensity as the monomer. A higher concentration of F12 increased the free eGFP signal. This result implies

that the fluorescence increase measured with the plate reader in the time course experiment (figure 4.5.2.3) was due to free eGFP and not to an increased in solubilized transporter.

The final conclusion from this section is that the best F12 concentration to solubilize hSGLT1 is 0.7% (25 times the CMC). Higher F12 concentrations did not extract more protein but rather increased the proteolytic activity causing the release of eGFP.

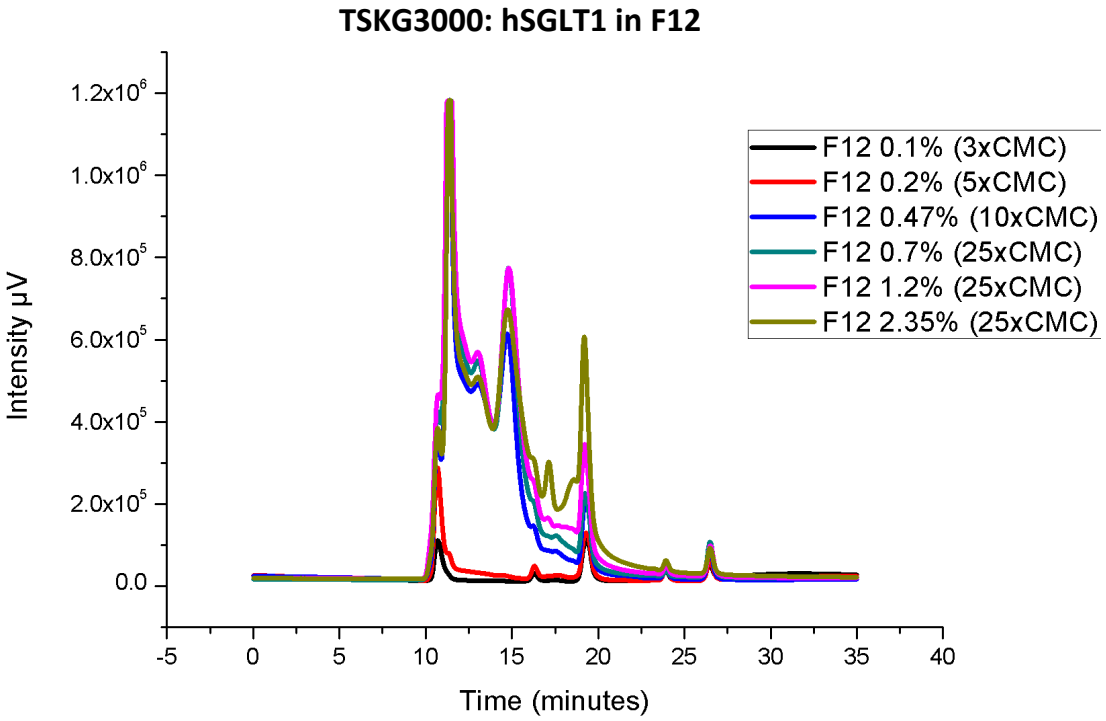


Figure 4.5.2.7: FSEC chromatogram of solubilized membranes for 2 hours with different F12 concentrations. All runs were done in TKCL with F12 at around 3 times CMC at a flux of 0.5 mL/min. Intensity values are represented in µV units. Concentrations values are expressed either in percentage or CMC.

Another factor affecting protein solubilization is the presence of some specific membrane components, such as cholesterol. It is known that cholesterol can improve protein solubilization of eukaryotic membrane proteins (**111**). Since hSGLT1 is a eukaryotic membrane protein, in vivo, it's expected to be embedded in a lipid bilayer containing cholesterol. In order to check the effect of cholesterol on the solubilization ot hSGLT1, membranes were treated with F12 at 0.7% with different concentrations of cholesterol hemisuccinate (CHEMS). CHEMS is used instead of regular cholesterol because it's been describe to actually stabilize more the solubilized complex of lipid and detergent and it's easier to incorporate in lipid detergent micelles (**109, 110**).

The chromatogram results of the cholesterol screening are shown in figure 4.5.2.8. No changes are appreciated in the chromatogram at different CHEMS concentrations. This indicates that CHEMS is

not helping to extract the protein from the membrane. Regular cholesterol instead of CHEMS might still enhance protein solubilization but it is not completely soluble in the presence of F12, thus it was discarded as an option for screening.

The protocol established for previous screenings was the same for this protein solubilization and the same concentrations of F12 used in figure 4.5.2.4 were used for this screening also. As shown in the figure 4.5.2.7, at 0.1 and 0.2 % of F12 the solubilization of hSGLT1 is practically zero because only 2 small peaks appears; the void volume (10 minutes) and the free eGFP (20 minutes) and no monomer or oligomer appears (14.5 minutes or 11 minutes). This results matches with the data seen in the plate reader because, at those low detergent concentrations, the fluorescence signal was low due to mainly free eGFP. At 0.47% of F12 the solubilization of hSGLT1 takes place because the monomer and oligomer peaks suddenly appear. This is an indicator that, at that detergent concentration, hSGLT1 is extracted from the membrane. At higher F12 concentrations the chromatograms don't change much although some details can be appreciated. There's more monomer signal at 0.7 % and 1.2% of F12 than at 0.47% but the oligomer signal is for all the same. It must be add that the oligomer signal is saturated so no conclusions can be done. No differences can be appreciated between the chromatogram at 0.7% and the 1.2% of F12. At higher concentrations of F12 (2.35%), the signal from the monomer is smaller than before but the signal from the free eGFP is way higher with nearly the same intensity as the monomer. Higher concentration of F12 increased the free eGFP sginal which is clearly a main problem for the plate reader screenings. This might explain the inconsistency of the plate reader measures for the time course and the F12 concentration screening

The final conclusion from this section is that the best F12 concentration to solubilize hSGLT1 is actually 0.7% or 25 times CMC, because; higher F12 concentrations didn't extract more protein but rather increased the proteolytic activity causing the release of more eGFP.

Some other critical variables can affect the solubilization like the presence of some additives, for example; cholesterol. It's known that cholesterol can improve protein solubilization of eukaryotic membrane proteins (**73**) so; it would be interesting to study the effects on hSGLT1. Since hSGLT1 is a eukaryotic membrane protein, in vivo, it's expected to be surrounded within the plasma membrane with cholesterol. While working in vitro, the cholesterol alone cannot solubilize the protein by itself so it's always a combination of detergent and cholesterol, so; the best detergent should be selected. Therefore, the screening was done with only F12 at 0.7% with different concentration of cholesterol hemisuccinate (CHEMS). CHEMS is used instead of regular cholesterol because it's been describe to actually stabilize more the solubilized complex of lipid and detergent and it's easier to get incorporate in lipid detergent micelles (**111**). It's important to remember that CHEMS is not soluble in water and it will be actually the detergent which solubilizes CHEMS. Then, firstly it should be checked how soluble is CHEMS in our detergent of use (F12) but, since it has been used in the past with F12, (**73**) this shouldn't be an issue.

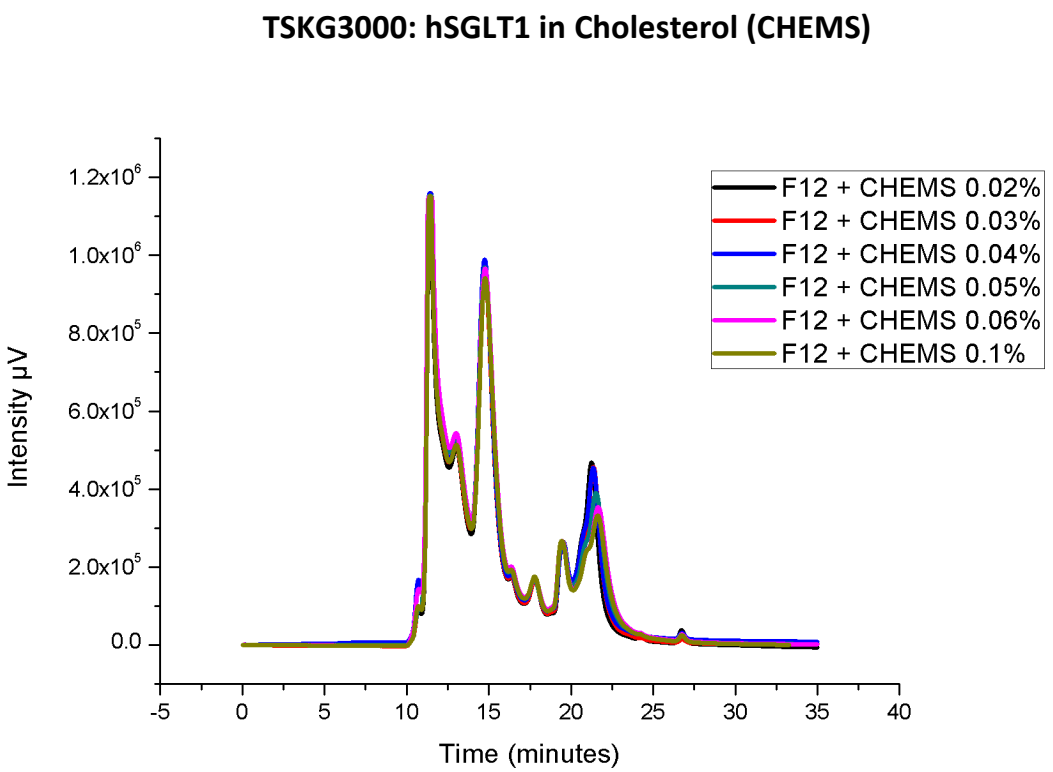


Figure 4.5.2.8: FSEC chromatogram of solubilized membranes for 2 hours with F12 at 0.7% plus different CHEMS concentrations. All runs were done in TKCL with F12 at around 3 times CMC at a flux of 0.5 mL/min. Intensity values are represented in µV units. Concentrations values are expressed either in percentage or CMC.

The chromatogram results of the cholesterol screening are shown in the figure 4.5.2.8. No changes at all are appreciated in the chromatogram between different CHEMS concentrations. This indicates that CHEMS isn't helping to extract the protein from the membrane so it's not a good additive for solubilization. Regular cholesterol instead of CHEMS might still enhance protein solubilization but it's not completely soluble in the presence of F12 and, since CHEMS didn't improve the solubilization, it was discarded as an option to screen.

Cholesterol is one of the most common additives to screen for protein solubilization but there are others, such as dithiothreitol (DTT) and glycerol that, apart from affecting solubilization, affect protein stability.

Glycerol has been reported to affect positively the stability of membrane proteins (**73**) so it's worth trying to see if, for example, less proteolysis occurs when solubilizing the protein in the presence of glycerol.

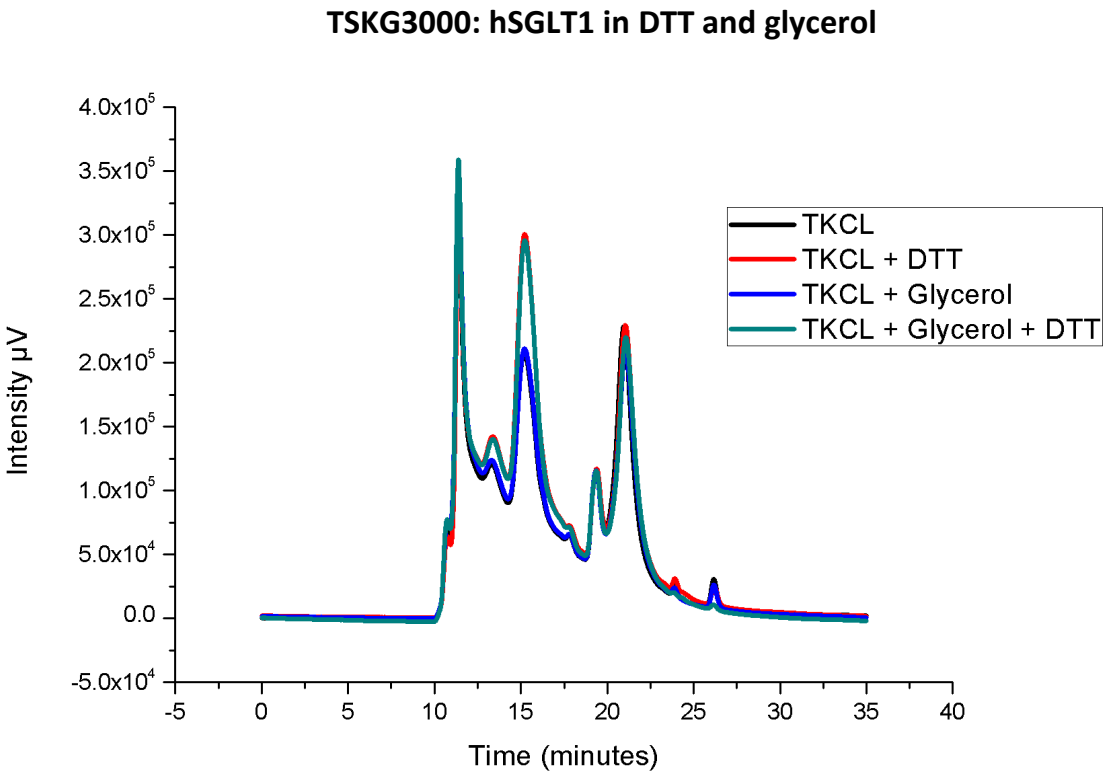


Figure 4.5.2.9: FSEC chromatogram of solubilized membranes for 2 hours with F12 at 0.7 with 1mM DTT and glycerol (5%). All runs were done in TKCL with F12 at around 3 times CMC at a flux of 0.5 mL/min. Intensity values are represented in µV units. Concentrations values are expressed either in percentage or CMC.

The results of the screening checking the effect of glycerol and DTT on the solubilization process are shown in figure 4.5.2.9.

No changes are appreciated in the chromatogram due to the presence glycerol. However, DTT has some effects. The signal from the monomer is more intense in the presence of DTT but the rest of the chromatogram is the same. These results suggest that somehow DTT enhances the solubilization of hSGLT1 and might be an indication of the fact that in the membranes, hSGLT1 is could be in an oligomeric state. Although cholesterol, DTT and glycerol are common additives to screen for protein solubilization, there are other parameters that might affect protein solubilization, for example, the amount of membranes (weight) per total detergent added (volume of solubilization).

In order to study this, two different ratios of membranes were used: 0.08 (w/v) and 0.15 (w/v) and were solubilized each one at 1.2% and 0.7% of F12. A higher ratio equals more membranes per the same amount of detergent. Results are shown in the above figure 4.5.2.10.

As expected, for the same ratio of membrane/detergent, more protein is extracted with higher F12 concentration. Nevertheless, the point of this experiment is to compare different membrane/detergent ratio with the same percentage of detergent for the solubilization. If the membrane/detergent ratio has no effect, with the same amount of detergent, the same protein should be extracted. On one hand, at 0.7% of F12 the chromatogram is very similar for both conditions: low ratio and high ratio, but, at a high ratio, less monomer (peak at 15 minute) is appreciated compared with the low ratio. Also, a peak at 21 minute is higher in the high ratio compared to the low ratio. This peak has more or less the same intensity as the monomer. No realistic conclusion can be said of what this peak actually represents a part from being degraded eGFP. One the other hand, at 1.2% of F12, again, the overall chromatogram is very similar for both conditions but a huge difference can be appreciated in the intensity of the monomer. At lower ratio, the amount of oligomer and monomer is way higher than at higher ratio. This result points in the same direction seen before at 0.7% of F12.

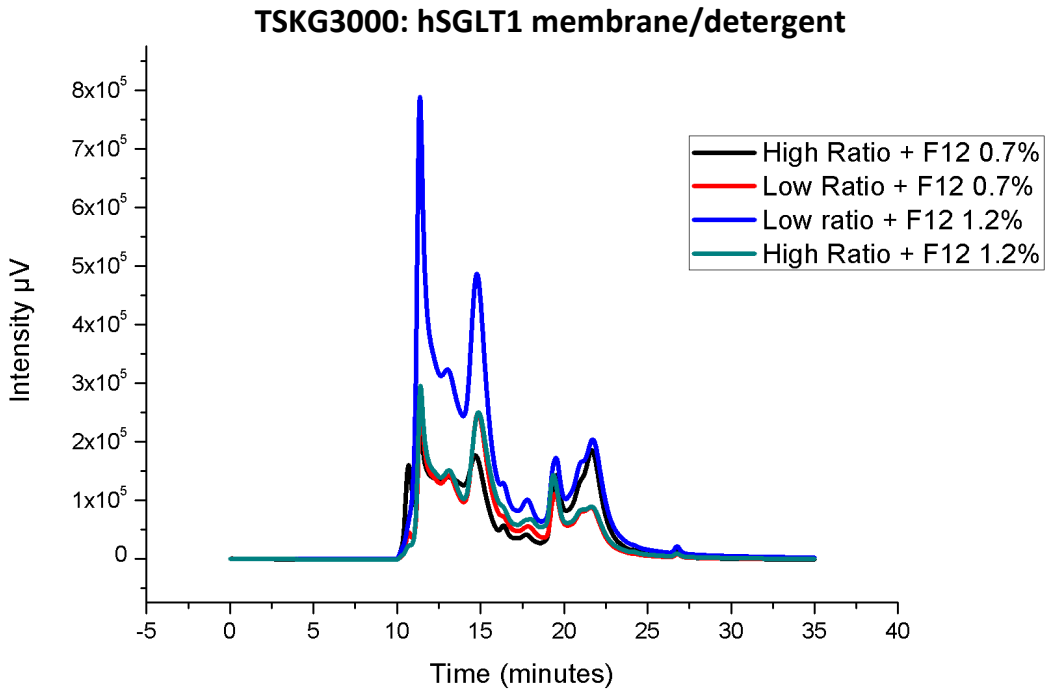


Figure 4.5.2.10: FSEC chromatogram of solubilized membranes for 2 hours with F12 at 0.7 and 1.2% at different membrane/detergent ratio: Low ratio 0.08 (g/mL) and High ratio 0.15 (g/mL) All runs were done in TKCL with F12 at around 3 times CMC at a flux of 0.5mL /min. Intensity values are represented in µV units. Concentrations values are expressed either in percentage or CMC.

In summary, these results suggest that the amount of membrane used to solubilize is even more important than the actual concentration of detergent to work with because, as it was seen, same concentration of detergent extracted much more protein with less membranes. Therefore, is important to maintain this low ratio of 0.08 g/mL for future solubilizations. Also, it's interesting to

add that these results explain why, from one membrane batch to another, different solubilization levels were obtained from the plate reader measurements.

4.6 Protein purification

4.6.1 Purification of hSGLT1+eGFP

Selected yeast for protein expression was grown as explained in section 3.2.5. Cells were lysed with glass beads and membranes were collected and solubilized with F12. After spinning down the solubilized membranes for 2 hours at 100.000xg, the purification is ready to start. The purification is divided in two steps: Ni-NTA immobilized metal affinity chromatography (IMAC) and a size exclusion chromatography (SEC). For details on both techniques see sections 3.2.9.1 and 3.2.9.3.

4.6.1.1 Affinity chromatography (IMAC) and size exclusion chromatography (SEC)

The affinity chromatography can be carried out in a column eluted by, by gravity, or by fast pressure liquid chromatography (FPLC, ÄKTA) system (see materials and methods for details). Figure 4.6.1.1.1, shows an SDS-PAGE gel, of the fractions which resulted from the Ni-NTA column (elution by gravity).

A big smear is appreciated in the Flow through (FT) and in the Starting material (S_m) and no differences are appreciated between those two samples. This can be explained by the fact that not only the recombinant protein with the His-tag, but also other proteins with a high content of histidines will bind to the Ni-NTA column. The smearing could be due to a long batch incubation which might causes protein degradation. In the lanes corresponding to the column-washing steps (W_1 and W_2), some bands can be identified: at 75 kDa and 25 kDa. These three bands appear also in the elution profiles and, plus another band near 100 kDa and, in the last elution samples, another one at 37 kDa (E8-E9).

The size of hSGLT1 is 75 kDa but, like other membrane proteins, has a different apparent molecular weight in SDS-PAGE (**65, 36, 92**). Since hSGLT1 is fused to eGFP, the band should appear technically around 100 kDa if it was migrating properly. Therefore, it's hard to know which band corresponds actually to hSGLT1+eGFP because three bands could match with its size: 75kDA, 85 kDA and 100 kDA.

The best way to solve this problem is to check the fluorescence of the elution samples in gel-fluorescence. Instead of staining the SDS PAGE gel with comassie blue or another dye, the gel is transiluminated to as a way to detect the presence of eGFP-labeled proteins. Results are shown in the figure 4.6.1.2.2 for the eluted fractions from the Ni-NTA column.

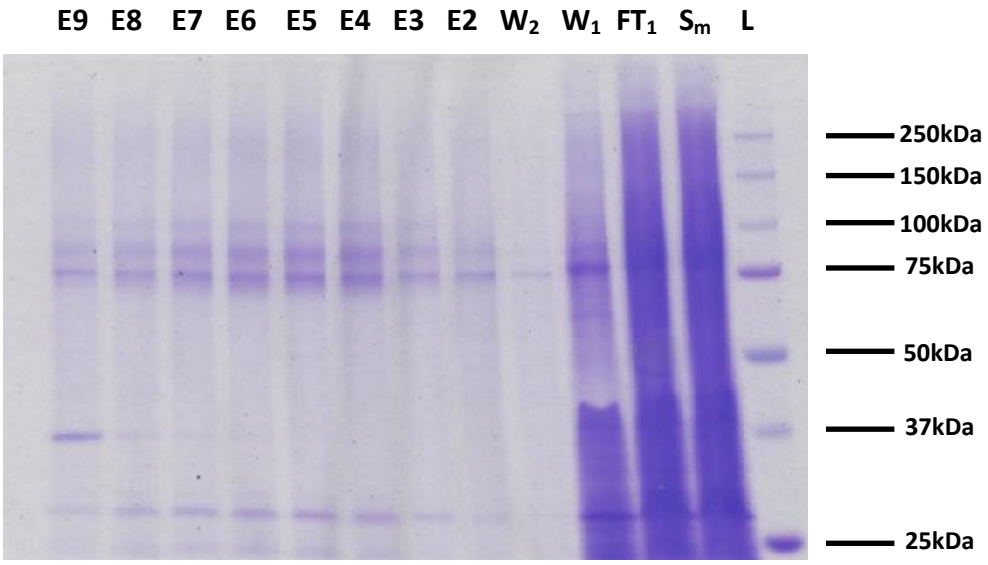


Figure 4.6.1.1.1: SDS-PAGE at 12% stained with comassie blue. Collected samples were loaded in the gel. Wash 1 and 2 are represented as W_1 (10mM Imi.) and W_2 (20mM Imi.), Flow through as FT_1, S_m as starting material and L as the ladder. Elution was done with 100mM of imidazole and collected in different tubes, starting from right to left (2-9). Same volume was loaded for all the samples except the FT and S_m. All buffers were at 150mM KCl and 20mM Tris-HCl at pH 7.6.

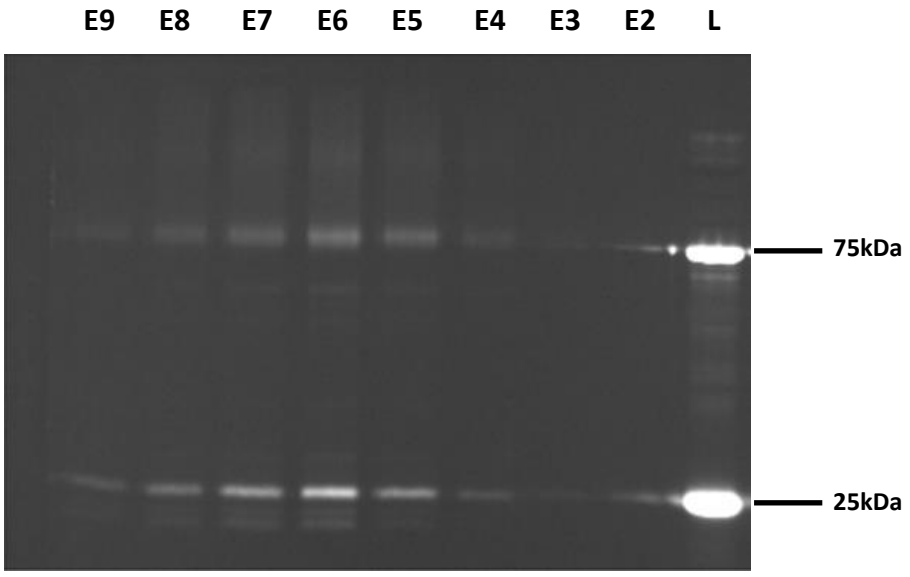


Figure 4.6.1.1.2: Gel fluorescence of and SDS-PAGE at 10%. Collected samples from the elution were loaded in the gel. Elution was done with 100mM of imidazole and collected in different tubes, starting from right to left (2-9). Same volume was loaded for all the samples. All buffers were at 150mM KCl and 20mM Tris-HCl at pH 7.6.

The gel fluorescence results reveal that only the bands at 75 kDa and 25 kDA are fluorescent The 75 kDa band corresponds to hSGLT+eGFP and the 25 kDA band free eGFP. Other fluorescence bands are detected but they are very faint. Above 75 kDa a small band with a smear is appreciated which indicates some protein degradation. Also, below 25 kDa other fluorescence bands are seen which reveals proteolysis. All the non-fluorescent bands detected in the SDS gel must be contaminations but, since they are difficult to remove in the 20 mM imidazole washing steps. It is possible that they are interacting with hSGLT1. Washing with higher concentration of imidazole implied losing hSGLT1-eGFP recombinant protein.

After Ni-NTA chromatography the eluted fractions were pooled together and injected in a SEC Superdex 200 column after centrifuging the sample for 1 hour at 50.000xg. Results are shown in figure 4.6.1.2.3.

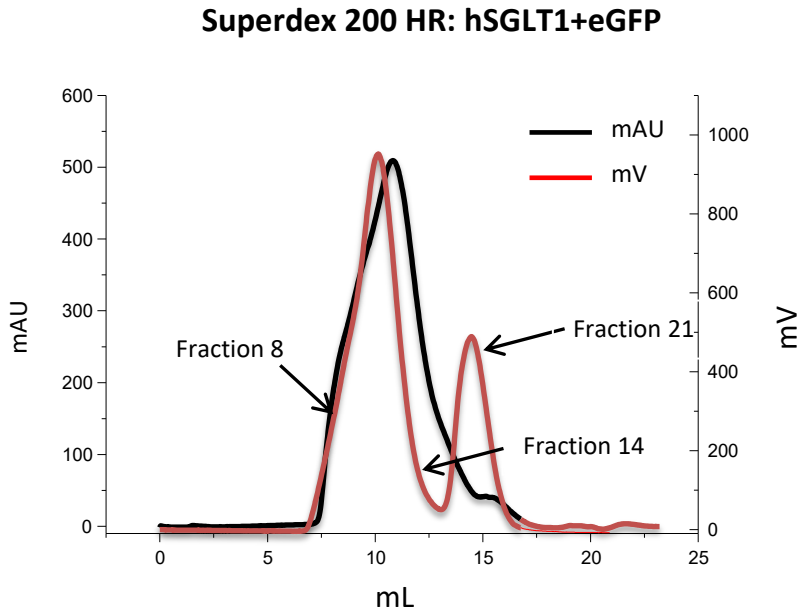


Figure 4.6.1.1.3: SEC chromatogram of hSGLT+eGFP after Ni-NTA purification. The size exclusion was done using a Superdex 200 HR column of 25mL. Two chromatograms are represented related to their respective y axis; left y axis is representing mAU in black (absorbance units) and the right y axis is representing mV (voltage) in red. Black arrows represent the fraction number of the elution.

The eluted fractions were pooled together and were injected, after centrifugation at 50.000xg for 20 minutes into a Superdex 200 column. Results are shown in figure 4.6.1.1.3.

The fluorescence signal (read out of the fluorescence detector in mV), in red, reveals two different peaks, one is free eGFP, at 15 mL, and the other is hSGLT+eGFP, at around 9 mL. The eGFP peak is very symmetric and clearly separated from the hSGLT+eGFP peak. The hSGLT+eGFP peak has a small shoulder on the left (around 8 mL) which indicates some heterogeneity of the sample. The

absorbance signal (mAU) should be very similar to the fluorescence signal of hSGLT peak of the sample if it had no contamination. However, the results shows that the His-tag purified sample presents some impurities which make the mAU chromatogram wider and less homogenous and causes to differ from the fluorescence peak. The eluted protein was fractionated (500 µl per fraction) and fractions of the main peaks were loaded and analyzed in an SDS-PAGE. The gel fluorescence is shown in figure 4.6.1.1.4, and with the result of the comassie blue staining in figure 4.6.1.1.5.

All fractions from the first peak (at around 11 mL) were loaded: fractions 8 to 14 in the gel, which also includes the shoulder fractions at 8 mL. Fractions of the free eGFP peak at 15 mL are fractions 20 and 21 in the gel, a bright and wide band at around 75kDa can be appreciated in figure 4.6.1.1.4, which is in good correspondence with the size of hSGLT1+eGFP and the band detected in figure 4.6.1.1.2. In the lanes corresponding to fractions 8 to 14, a smear-band at higher molecular weight, just below the stacking gel, is also present. This smear fades away as we move towards fraction 14. At the same time, the hSGLT+eGFP band intensity, at 75kDA, increases as we move from fraction 8 to fraction 14, presenting a maximum in intensity at fraction 11. On the other hand, fractions 20 and 21 gave a thin bright band near 25kDA which matches with the size of eGFP (in agreement with the result reported in figure 4.6.1.1.2).

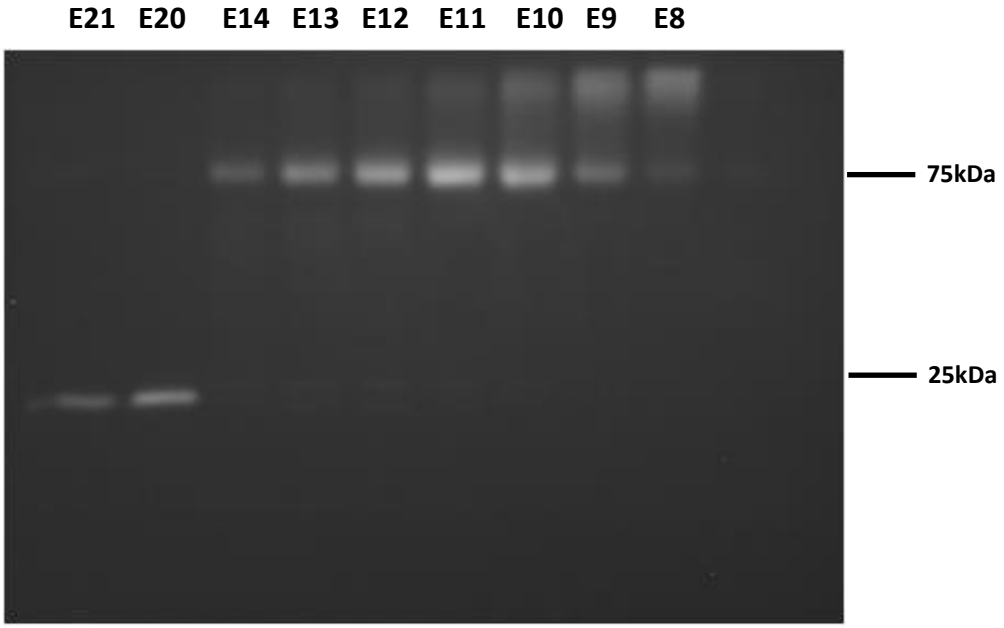


Figure 4.6.1.1.4 Gel fluorescence of and SDS-PAGE at 12%. Eluted samples from the SEC were loaded in the gel. Eluted fractions (E8-21) are loaded from lower elution volumes (right) to higher elution volumes (left). Same volume was loaded for all the samples. All samples are buffered with 20 mM Tris-HCl at pH 7.6 and 150 mM KCl.

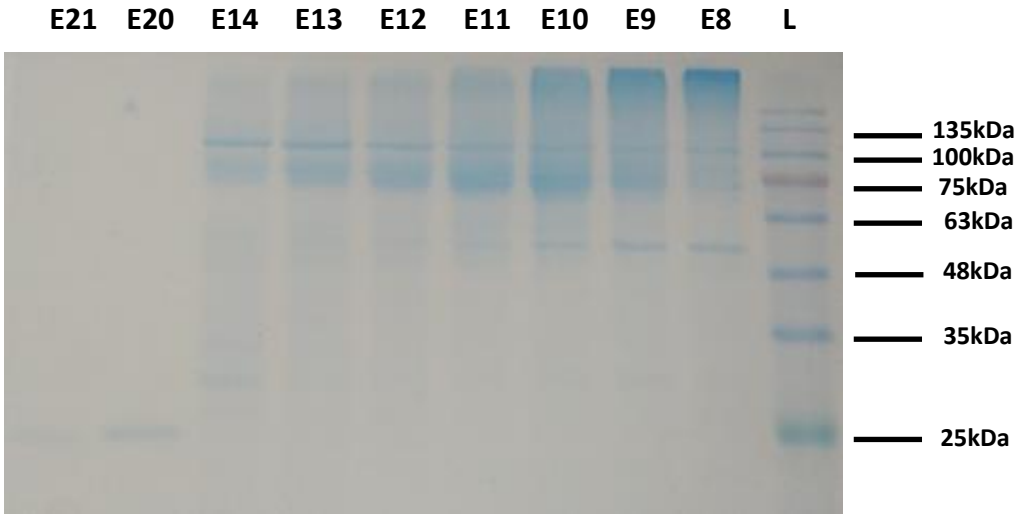


Figure 4.6.1.1.5 Stained SDS-PAGE at 12% with comassie blue. Eluted samples from the SEC were loaded in the gel. Eluted fractions (E8-21) are loaded from lower elution volumes (right) to higher elution volumes (left). Same volume was loaded for all the samples. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl.

Since the smeared band is fluorescent it has to contain eGFP and somehow be a product of the heterologous expression.

Everything so far is clearly suggesting that the higher molecular weight band-smear is some type of oligomeric state of hSGLT+eGFP. It's important to remark that the band cannot be a massive protein aggregate or otherwise it would have not entered the column and a void volume peak would have been seen in the SEC profile. Neither seems to be a simple artifact of the SDS-PAGE because before running the gel, in the SEC chromatogram, a shoulder was already seen in the first peak, already indicative of some kind of heterogeneity due to a higher molecular weight species.
Besides the bands just commented, some other bands were present in figure 4.6.1.1.5. The 50kDa and the 100kDa bands could be interactor proteins or simple impurities.

In any event, in order to get the pure hSGLT1+eGFP a repurification step was added where the eluted protein from fractions 8 to 13 were pooled together and injected again in the same SEC column. The resulting chromatogram is shown in figure 4.6.1.1.6, with the corresponding SDS-PAGE comassie blue gel in figure 4.6.1.1.7.

It can be appreciated in the SEC profile that the same chromatogram is obtained for the fluorescence (in red) and for the absorbance (in black) measurements. This shows that the heterogeneity has decreased. The observed shift of one chromatogram with respect to the other is due to a delay on the detector response. The eluted protein was fractionated and some of the collected samples were loaded in a SDS-PAGE stained with comassie as shown in figure 4.6.1.1.7.
The first lane in the figure 4.6.1.1.7 is corresponds to the starting material, before injection (E8-13 from figure 4.6.1.1.5). The 75kDa band is present in all the eluted fractions with a minimum

intensity in elution 6 and a maximum in elutions 8 to 10 which corresponds to the peak maximum. Between the 75kDa band and the possible interactor protein at 100kDA, another band is present at around 85kDa. This band was also detected in the comassie after Ni-NTA (figure 4.6.1.1.1). Depending on the percentage of acrylamide used in the gel this band is blended with the 75kDA band.

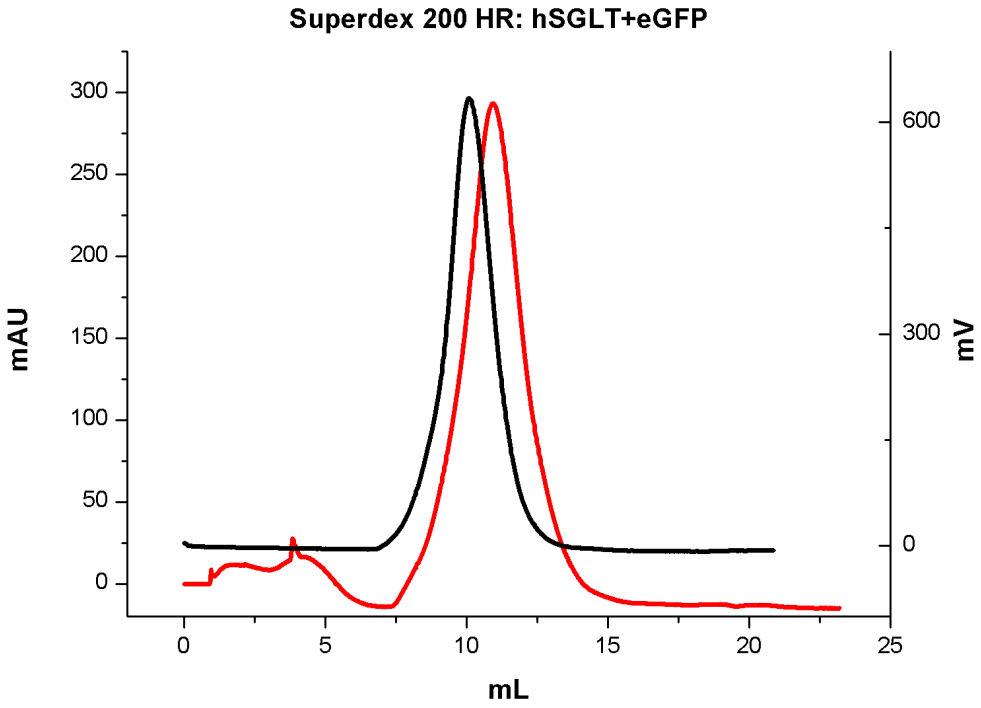


Figure 4.6.1.1.6: Repurified SEC chromatogram of hSGLT+eGFP after a SEC and Ni-NTA purification. The size exclusion was done using a Superdex 200 HR column of 25mL. Two chromatograms are represented related to their respective y axis; left y axis is representing mAU in black (absorbance units) and the right y axis is representing mV (voltage) in red.

It can be appreciated that all those possible interactor proteins follow no pattern of elution because the 100kDa band appears in all eluted fractions and the same goes for the 85kDA band. On the other hand almost no smeared band is detected in the initial fractionated elutions, which means that what could be the protein in some sort of oligomeric state has been eliminated in the repurification

As for the purification yield, after the first SEC chromatogram, around 0.75-1 mg of protein/L of culture is obtained. Despite the impurities in the sample, estimation can be done by measuring the fluorescence signal from the sample (**73**). The estimation was done without considering the possible oligomeric state of hSGLT1 because only the monomer is used for structural studies.

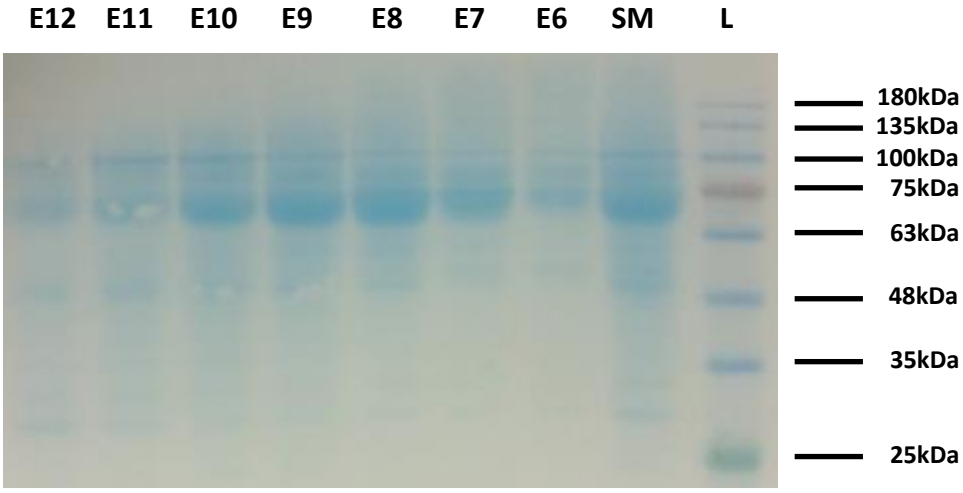


Figure 4.6.1.1.7: Stained SDS-PAGE at 10% with comassie blue. Eluted samples from the repurification SEC were loaded in the gel. Eluted fractions (E6-12) are loaded from lower elution volumes (right) to higher elution volumes (left). First lane (SM) is the concentrated sample before injecting to the column. Same volume was loaded for all the samples except the SM. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl.

4.6.1.2 Thrombin digestion: Improving protein purification

Due to the presence of impurities in the sample after affinity (Ni-NTA) purification and even after size exclusion (SEC), a new strategy to improve the purity of the purified protein was pursued. The approach was to cleave the eGFP from the recombinant protein sequence by using thrombin. The vector was designed so a thrombin cleavage site is between hSGLT1 sequence and eGFP. Initially, the vector was designed with a thrombin digestion site because eGFP it's known to be a problem for structural studies (**114, 115**) or even, in some cases (**116**), for functional studies, although, as already mentioned before, the fused construct presents other advantages. Removing eGFP with thrombin from the main hSGLT1 sequence causes also to remove the histidine tag, which, immediately, makes the protein unavailable to bind to a Ni-NTA column. This approach can be used for improving the purification (**117**).

After binding the protein into a Ni-NTA and washing properly with low imidazole concentration, a thrombin digestion was carried out in the same Ni-NTA resin with the protein bound to it (**118**). This thrombin digestion was done on the same column and left on batch mode overnight (**119**). Although the optimal activity of thrombin is at 37ºC, the digestion was done at 4ºC to avoid protein degradation or proteolysis (see material and method section, for more information 3.2.9.4). The digested recombinant protein shouldn't bind the Ni column, so by restarting the gravity flow and repacking the resin, the protein should elute while other contaminants will still remain bound to the column.

In figure 4.6.1.2, a gel fluorescence image is showing the results after thrombin digestion. Lanes 1 and 2 are controls of purified eGFP obtained from *P. pastoris* and purified hSGLT+eGFP respectively. As it can be appreciated, the eGFP control gives a thick band near 25kDA where the eGFP should appear. As for the hSGLT1+eGFP control, a main band at 75kDA is spotted where hSGLT1+eGFP should appear. Also, other small bands below the 75kDA are present which could be due to proteolysis. Since this purified control wasn't injected to a SEC column, a big smeared band is also present just below the stacking. Lane 4 to 6 is the result after the digestion. Different amounts of total protein were added in each of those lanes because very few amounts of protein were actually eluted: Lane 4 is 1µg, lane 5 is 5 µg and lane 6 is 10 µg of total protein. If the eGFP was cleaved properly no band should appear at 75kDa and, instead, a band at the size of the eGFP should be detected.

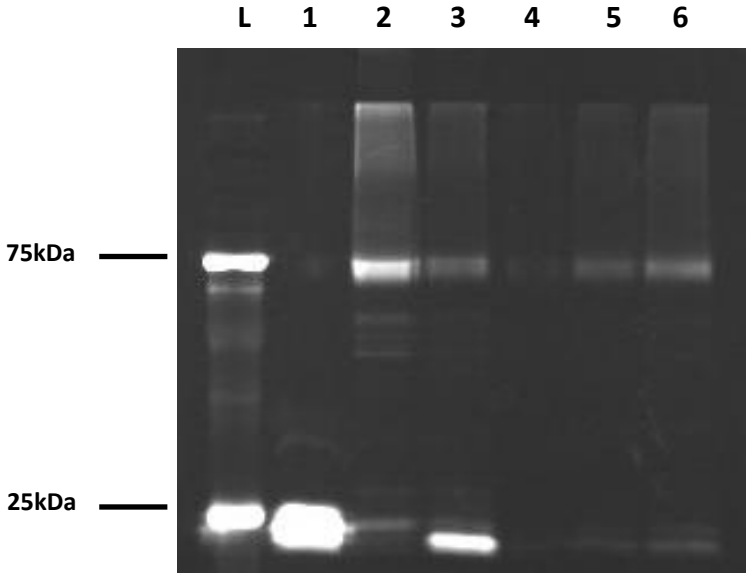


Figure 4.6.1.2: Gel fluorescence of and SDS-PAGE at 12% after digestion with thrombin in Ni-NTA column. L; Ladder, 1; control of eGFP, 2; purified control of hSGLT+eGFP, 3; Imidazol elution, 4-6; elution with thrombin with different total protein added: 1,5,10 µg. Same volume was loaded for all the samples except the controls. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl.

However, In lanes 4 to 6 of figure 4.6.1.2 a bright band appears at 75kDa with a visible band at 25 kDa in lanes 5 and 6. The intensity of those bands increases proportionally to the total concentration of protein. It's clear that free eGFP is present and hSGLT+eGFP is also present after digesting with thrombin which means that the protein, or most of the protein was not cleaved. After letting the protein flow through the colum, an elution with imidazole was done to elute all the non digested hSGLT1+eGFP, the free eGFP which is also bound to the column and possible contaminants The results after this elution are represented in lane 3. A lot of free eGFP and also hSGLT1+eGFP was eluted with imidazole. The amount of protein eluted after the thrombin digestion is more than 10 times lower than the amount obtained after the elution with imidazole,

so, this clearly indicates that the protein was not digested. To confirm that no protein has been digested with thrombin, the same gel was stained with comassie blue. No differences were appreciated. (data not shown) after staining with comassie. Some other additional information can be extracted from this experiment. If the protein was not digested, why some protein is being eluted after the digestion with the thrombin without adding imidazole. As it was suggested before, the protein does not seem to bind very very tightly to the column because increasing the imidazole concentration for washing caused the protein to elute. Therefore, the elution of some protein after the digestion might be explained because of this.There are several possible reasons why the protein is not being eluted after cleaving with thrombin. First, it could be that being the protein within the detergent micelle makes the target sequence unaccesible to thrombin (**118**). Second, it is possible that the detergent is actually affecting the activity of thrombin or that the conditions for the digestion are not optimal. Same buffered detergent was used to test thrombin digestion with no effect on the activity (data not shown).

Other attempts were carried out to make sure that the digestion conditions were or not a problem, as for example; the digestion temperature, the digestion time, or carrying out the digestion after Ni-NTA to make sure the digestion in the resin is not an issue (data not shown). None of this attempts did improve the digestion. In this sense, it is important to note that it is know that for some membrane proteins, cleaving the eGFP from the main sequence it's not allways acomplished (**118**)**.**

4.6.2 Purification of WT hSGLT1

Because of the trouble in getting the hSGLT1+eGFP protein cleaved and since the final goal is to have the hSGLT1 transporter pure, it was decided to undertake the transporter expression and purification using the hSGLT1 construct (wihtout GFP). This approach permitted to take advantage of all the expression and pruification conditions that were optimized using the hSGLT1+GFP construct and avoiding dedicating more effort and time to problem (the GFP cleavage) that was going to be difficult to solve. The clone for the pJIN_hSGLT1 construct was used for large scale expression.

The same purification protocol stablished for the hSGLT+eGFP is used for the purification of the hSGLT1, so the purification is divided in two different purification steps: Affinity chromatography, Ni-NTA (IMAC) and a size exclusion chromatography (SEC). For details on both techniques see sections 3.2.9.1 and 3.2.9.3.

4.6.2.1 Affinity chromatography (IMAC) and size exclusion chromatography (SEC)

In figure 4.6.2.1.1, An SDS-PAGE gel shows the result of analyzing the aliquots eluted form the Ni-NTA column (by gravity).

First of all, no protein is detected in the washing out lane (W), which indicates that the amount of protein eluted when washing with imidazole is very low.

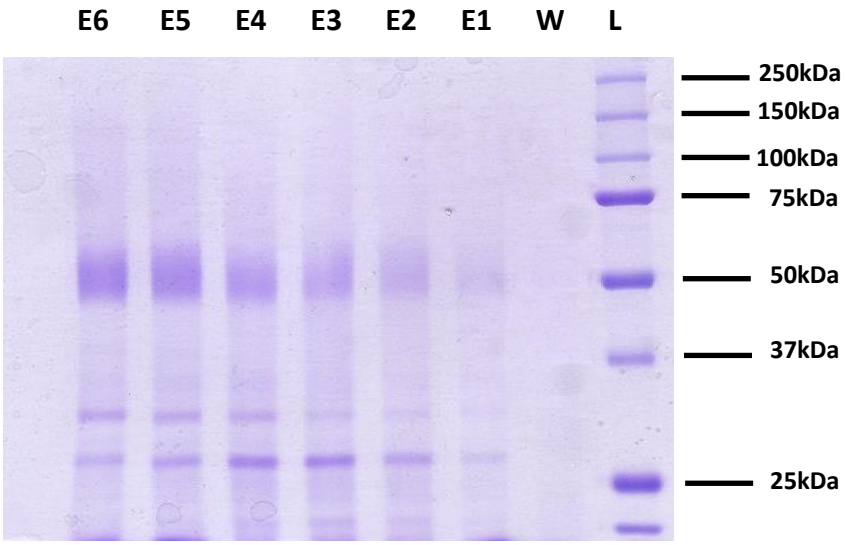


Figure 4.6.2.1.1: Stained SDS-PAGE at 12% with comassie blue. Eluted samples from the gravity Ni-NTA were loaded in the gel. Eluted fractions of 500µl (E1 to E6) are loaded from lower elution volumes (right) to higher (left).First lane, W; is the final wash before eluting with imidazole. Same volume was loaded for all the samples. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl. The Flow through (FT) is not represented on the below comassie since it did not change from a construct to another.

As for the elution, all elution aliquots show the same band pattern (E1-E6). A main diffuse band near 50kDA is appreciated. This band matches with the size of the WT hSGLT1 (**65, 36, 92**). The WT hSGLT1 band is even wider and more diffuse than the hSGLT1+eGFP band seen before (fig. 4.6.1.1.1). The band spreading may be due to heterogeneity caused by post-translational modifications or due to a SDS-PAGE artifact. Other bands at lower molecular weight are also seen in the gel, one below 37 kDa and another one above 25kDA. Those bands are very likely to be impurities. In the last elutions (E5-E6), a smear at high molecular between 75kDa and 150kDa is appreciated. This smear is very similar to what was seen before with the hSGLT+eGFP Ni-NTA purification.

In order eliminate the impurities all the elution fractions were pooled together, concentrated and injected into the SEC column. It must be pointed out that during the concentration and, before injecting into the SEC column, some low molecular weight proteins will be eliminated because the cutoff of the concentrator was 50kDa. Figures 4.6.2.1.2 and 4.6.2.1.3 show the resulting chromatograms and the corresponding SDS-PAGE images.

In the SEC chromatogram a main peak at 11 mL and a shoulder at around 9.7 mL are appreciated. This chromatogram is very similar to the one in fig. 4.6.1.1.6 corresponding to the construct with

GFP. The peak and the shoulder appear at the same position although, in this case, the protein is 25kDa smaller. This probes that trying to calculate the molecular weight of the peaks is very difficult with the SEC and, therefore, another technique is required to determine the molecular weight of the sample.

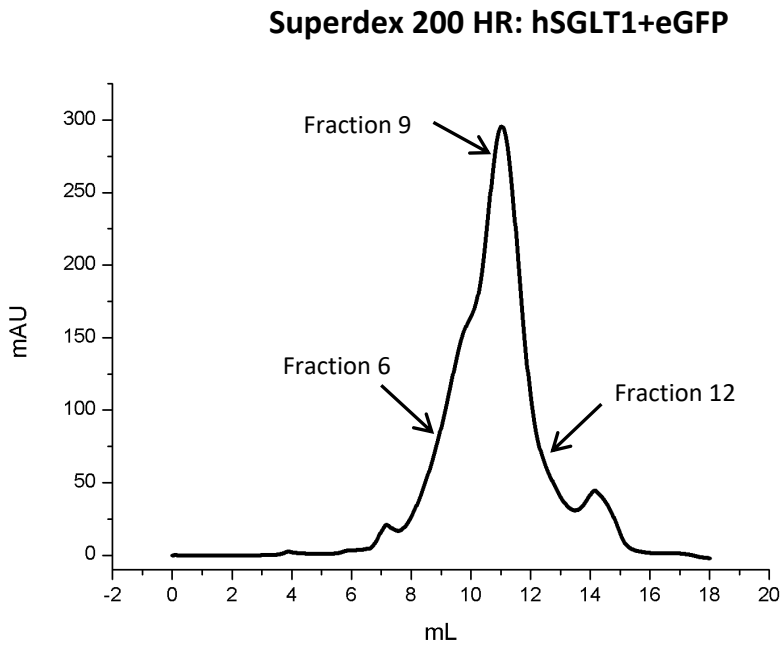


Figure 4.6.2.1.2: SEC chromatogram of WT hSGLT after Ni-NTA purification. The size exclusion was done using a Superdex 200 HR column of 25mL. Y axis is representing absorbance units and the Black arrows represent the fraction number of the elution.

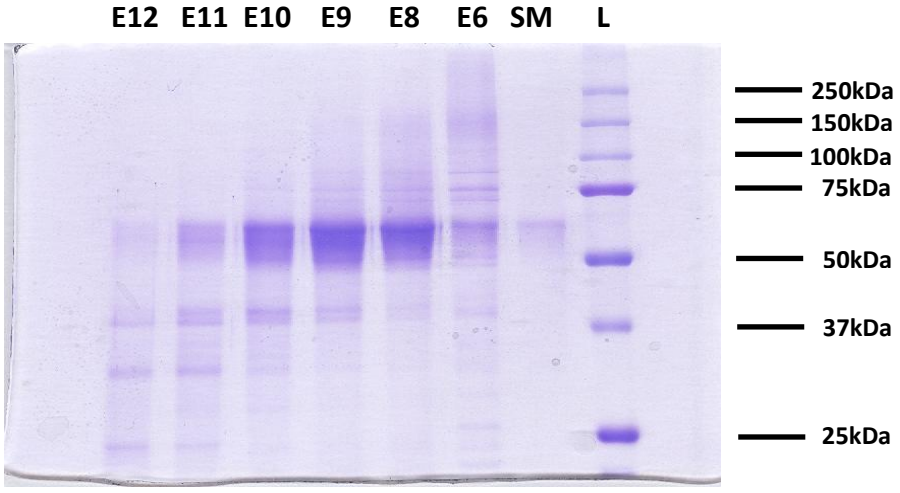


Figure 4.6.2.1.3: Stained SDS-PAGE at 12% with comassie blue. Eluted samples from the SEC were loaded in the gel. Starting material (SM) is the pooled samples of the Ni-NTA before concentrating and injecting. Eluted fractions (E6-12) are loaded from lower elution volumes (right) to higher elution volumes (left). Same volume was loaded for all the samples except SM. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl.

Overall, the chromatogram is very homogenous and promising for further crystallization trials or other structural studies but, like the eGFP construct, it's necessary to fractionate the eluted protein and run it on a SEC chromatogram to find out the presence or not of contaminant proteins. On the figure below 4.6.2.1.3, a SDS-PAGE stained with comassie of several fractions of the SEC is represented.

In the SDS-PAGE run with the eluted fractions fraction, E6 gives a high molecular weight smeared band around 150 kDa. This smeared band is very similar to what was seen in the previous SEC profile for the GFP construct (4.6.1.1.2). Other bands are observed at around 75, 85and 100kDa. Those bands were also observed after the His-tag purification of the hSGLT+eGFP construct, figure 4.6.1.1.1. Another band just above 50kDa is also observed. This band is diffuse and similar to one detected in the hSGLT+eGFP gel (fig. 4.6.1.1.5) and, at around 25kDa, nearly no band is appreciated which is what it would be expected for WT hSGLT1 construct (**63, 34, 90**). In elutions E8-10, the hSGLT1 band is more intense. The band heterogeneity might be due to post-translational modifications such as glycosylation, or to be a simple gel artifact (**49, 120**). Lower molecular bands are visible in elutions (E9-11)) and faint in the rest. Some of those low molecular bands appear also in the previous purification with the eGFP construct, like the 37kDa band, and maybe protein interactors or impurities.

The starting material (before concentrating the protein) was run in the lane as SM. Although this sample is much diluted, it gives an idea on how the sample looks like before concentration. The results were that only the hSGLT1 band can be appreciated. It's interesting that the smeared-band appears to be less intense in this lane than the rest. Since the smeared band can already be appreciated after Ni-NTA purification, it's possible that the amount of protein loaded in the gel effects the run and, therefore, causes the smearing. When the WT hSGLT1 band intensity increases, the smearing also increases. Overall, this data suggest that the band smearing or heterogeneity might be a gel artifact. It was reported in the past that SDS might interfere with hSGLT1 in an acrylamide gel (**93**). To confirm this possibility a mass spectrometry is necessary.
Overall, purification by SEC gave, in the case of hSGLT1, quite the same results as in the case of hSGLT1-eGFP, that is, some pure protein that gives a spread (thick) band in the SDS-PAGE accompanied by some impurities.

4.6.2.2 Affinity chromatography: FLAG-tag

In order to try to improve the final protein preparation by getting rid of the impurities observed in previous SDS-PAGE, other than the one corresponding to the hSGLT1, an extra purification step was introduced in the form of a FLAG-tag affinity chromatography. As explained in the vector design section (figure 4.1.1), the hSGLT1 insert was modified so a flag tag sequence was inserted at position 574 (**65**).

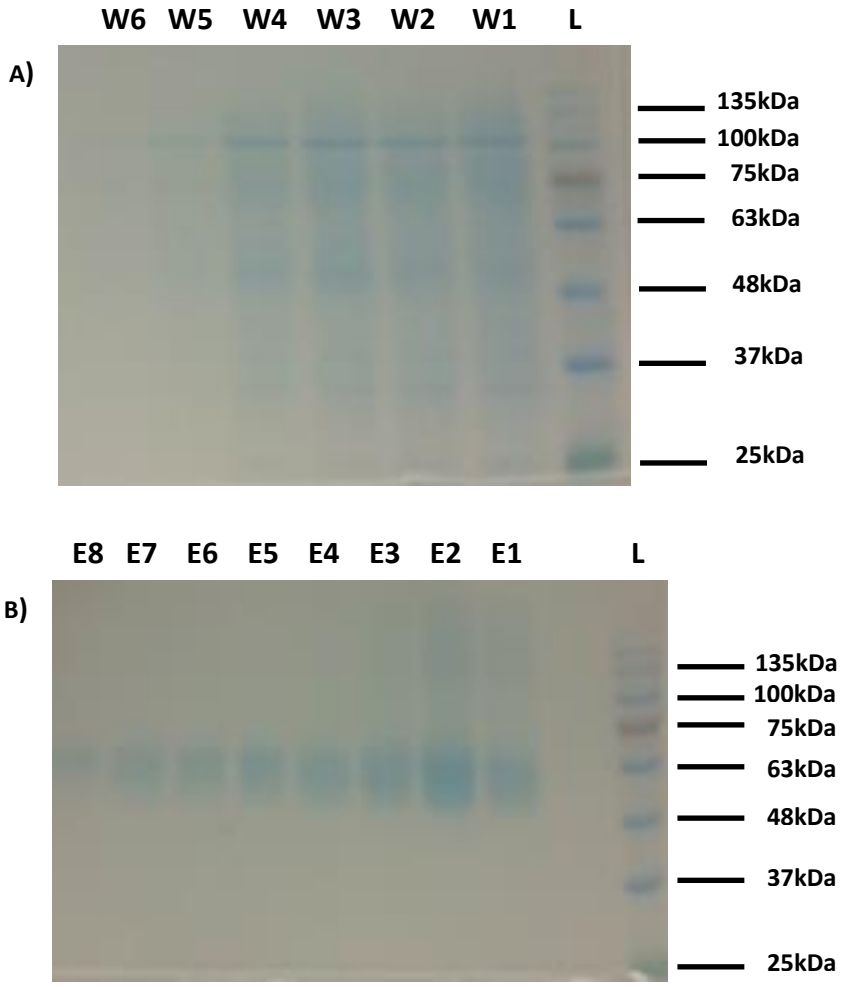


Figure 4.6.2.2.1: Stained SDS-PAGE at 12% with comassie blue. A) Washes of the flag tag purification. Same volume was loaded for all the samples. All samples are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl. B) Eluted samples of the flag tag purificaton.

The FLAG-tag purification was carried out before the SEC purification and after the Ni-NTA column. It is important to remember that FLAG-tag purification is more selective than Ni-NTA, because only proteins with the specific amino acid sequence YKDDDDK (flag tag sequence) will bind to the resin (**121**), therefore, it is expected that the impurities which go through the Ni-NTA column are removed. The results from the FLAG-tag column are presented in figure 4.6.2.2.1.

The FLAG-tag column was washed to remove unbound protein until no protein was detected in the flow through. Each of the washing aliquots are represented in the above figure A, as W1to W6. As it can be appreciated, several bands are observed in all the lanes. The bands are less intense in the last washing elutions. Those bands were actually seen also in the purification of hSGLT1+eGFP after Ni-NTA and SEC (figure 4.6.1.1.1). As mentioned in the material and methods section, the

elution of the Flag-tag column can be carried out either with a flag-tag peptide or with glycine at very low pH (Sigma-aldrich protocol)**.** Lowering the pH breaks the electrostatic interactions between the flag-tag antibody and the protein. The other method breaks the interaction by the competition caused by the addition of free flag peptide. This last method was attempted but surprisingly no protein was eluted. So, the results showed in panel B of 4.6.2.2.1 figure come from elutions carried out with glycine at low pH.

The elution was fractionated until no protein was detected and each fraction was run in SDS-PAGE. In all elutions, a thick smeared-band at around 50kDa is appreciated which matches the expected hSGLT1 band. Interestingly, a smeared-band at higher molecular weight is also seen in the first elutions, E1 to E3. This band matches with the 'oligomerization' band seen and postulated in the hSGLT1+eGFP SEC profile analysis. This result clearly indicates that that band is not a contamination or otherwise it would not bind to the flag tag column.

The eluted protein is highly pure and, if necessary, a further SEC can be carried out to remove the 'oligomers'. Overall, the flag-tag purification was a successful method. The washed bands are actually interactors or contaminants. If the bands were interactors, they'll still be bound to the protein after the flag-tag purification, therefore, those interactors proteins will not elute in the washes. Since all non-hSGLT1 bands were eluted in the wash and not in the elution, they must be simple impurities. Same buffer was used for the Ni-NTA and the flag tag so it must be assumed that the protein should remain bound to the column.

4.6.3 Purification of N248A mutant of hSGLT1

As mentioned before, the apparent heterogeneity (smeared band) of hSGLT1 and the hSGLT1+eGFP in SDS-PAGE can be a consequence of several reasons. One reason is post-translational modifications such as glycosylations, palmitoylation, hydroxylation and others. In yeast, glycosylation is a very common post-translational modification which can add up to 200 sugars chains to a single protein (**122**), although *P. pastoris* doesn't glycosylate as much as *Saccharomyces cerevisiae*, the glycosylation is still very relevant (**123**). Glycosylation is well known to cause heterogeneity and making the protein run poorly in a SDS-PAGE (**120**). That heterogeneity can be a problem for crystallization or other structural studies. Asparagine (N) 248 of hSGLT1 is well known to be a target for glycosylation (**48, 49**) so it is possible that this residue could be hyper-glycosylated by the heterologous system and causing heterogeneity of the expressed protein.

Two methods were used to test this possibility: Purification of the WT hSGLT1 with a Concanavalin A resin and preparation of a mutant of hSGLT1 where the N248 is substituted for an Alanine (A). Concanavalin A is an agarose resin linked to a lectin protein which binds carbohydrates (**123**). If hSGLT1 is glycosylated, it should therefore bind to the column. The purification was attempted with hSGLT1 and no binding at all was detected (data not shown). The N248A mutation was

introduced into the pJIN_hSGLT1 vector. Expression and purification of the mutant was carried out according to the protocols suited up for hSGLT1-eGFP and hSGLT1+eGFP. Results after Ni-NTA injected to a SEC (chromatogram and gel image) are shown in figures 4.6.3.1 and 4.6.3.2.

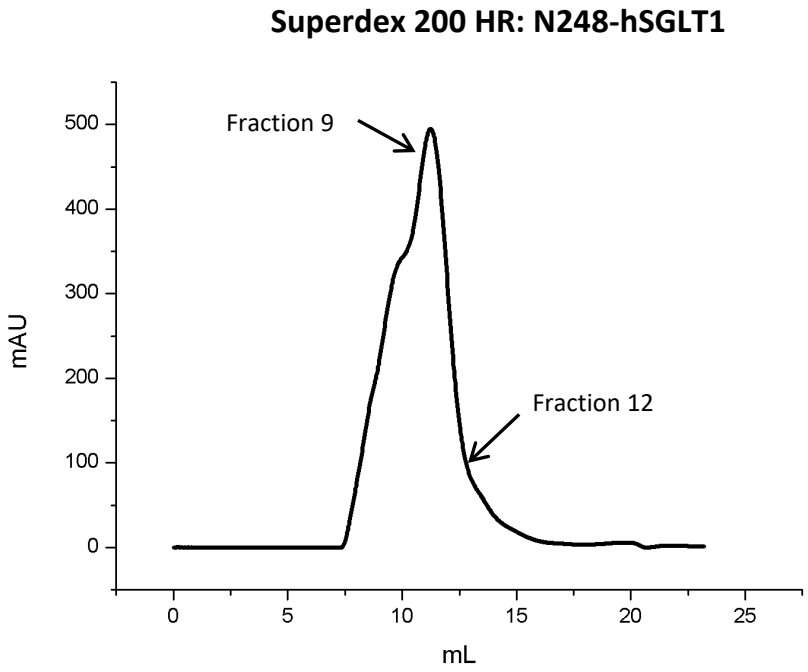


Figure 4.6.3.1.: SEC chromatogram of N248A-hSGLT mutant after Ni-NTA purification. The size exclusion was done using a Superdex 200 HR column of 25mL. Y axis is representing absorbance units and the Black arrows represent the fraction number of the elution.

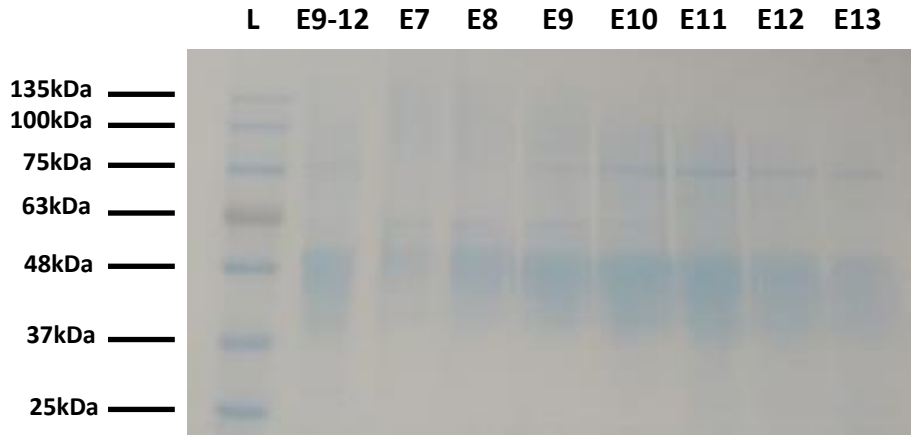


Figure 4.6.3.2 Stained SDS-PAGE at 12% with comassie blue. Eluted samples from the SEC were loaded in the gel. E9-12 lane represents the concentrated sample before injecting in SEC. Eluted fractions (E7-13) are loaded from lower elution volumes (left) to higher elution volumes (right). Same volume was loaded for all the samples and all are buffered with 20mM Tris-HCl at pH 7.6 and 150mM KCl.

The chromatogram is very similar to the hSGLT1 and the hSGLT1+eGFP chromatograms. The main results from the SDS-PAGE analysis is that the band corresponding to the mutant protein is as diffused (smeared) as it was hSGLT1 and hSGLT1+GFP.

The final conclusion after flag tag column and the N248A-hSGLT1 mutant experiments is that glycosylation are not the cause of the band heterogeneity

4.7 Mass spectroscopy of WT hSGLT1

hSGLT1 purified using the flag-tag column followed by SEC purification was used from mass spectroscopy. The mass spectrum is shown in figure 4.7.1.

The mass spec reveals a main peak at 74.483 kDa with another smaller peak at 74.747. Both peaks are consistent with the size of hSGLT1. The mass spectrum is not compatible with the sample having major post-translational modifications such as glycosylations or a non-resolved mass spectrum would be expected.

In order to check the possibility that the translational modification where rather small, and therefore with no significative effect on the molecular weight of the protein, a second mass spec analysis was carried out with the same sample but, instead of doing a full whole time fly of the sample, the protein was cut and the pieces were analyzed. This method is very useful because it reveals if those pieces have small modifications, as mentioned, such as hydroxylation or others. However, no changes were appreciated.

Overall this data indicate that the protein doesn't contain important post-translational modifications so glycosylation is not the reason why the protein band looks so smeared and blurry in the SDS-PAGE.

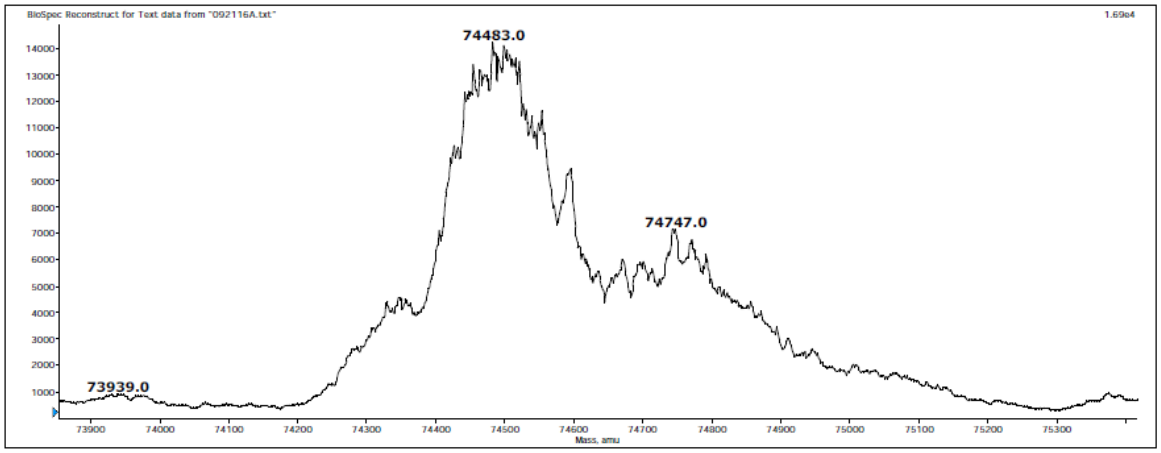


Figure 4.7.1 MALDI-TOF of purified WT hSGLT1 in100mM DTT.

4.8 Circular dichroism of WT hSGLT1 (CD)

Circular Dichroism (CD) is a spectroscopic technique that provides information on the secondary structure of proteins (**125, 126**). Membrane proteins like hSGLT1 with 14 transmembrane domains are expected to be mostly α-helical (helical up-down bundle structure) (**127**).

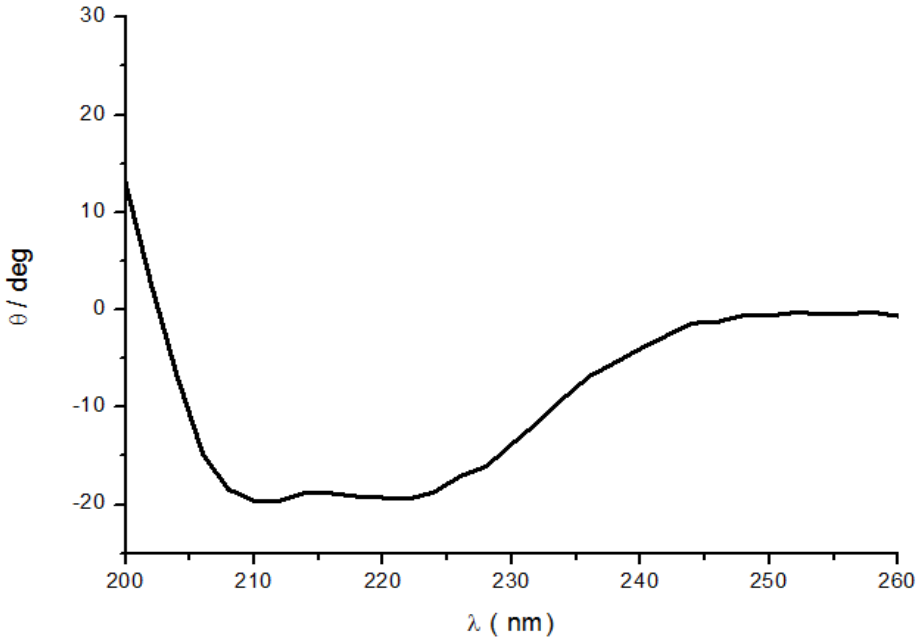


Figure 4.8.2: CD spectrum of WT hSGLT1 at 15 µM in 150mM TKCL and 20mM Tris-HCl pH 7.6. Measures were taken in cuvette with an optical path of 1 mm.

The CD spectrum of hSGLT1 is shown in figure 4.8.2. The spectrum has 2 minimums at around 210 and 220 nm which correspond to a protein predominantly α-helical.

The spectrum can be deconvoluted and, by doing so, it's possible to know the contribution of α-helix and β-sheet of it. This deconvolution was done with the previous WT hSGLT1 C.D spectrum. The results show that the secondary structure is ≈ 85% α-helix. These results are consistent with the C.D. spectra reported for vSGLT1 (**128**) and the data derived from the crystal structure of vSGLT1 (**33**).

Overall, these results are very interesting because they indicate that the protein is folded in the presence of detergent and its secondary structure is somehow similar to a homologous well studied protein such as vSGLT1.

4.9 Protein reconstitution of WT hSGLT1 in liposomes

Protein reconstitution is a well-established and powerful methodology for transport/functional studies of purified membrane proteins because, if the protein is engulfed in a membrane instead of the detergent, it is expected to behave similar to in vivo conditions. (**129**)

There are several methods to reconstitute a protein into a membrane mainly differing on how the detergent is removed (**130**). The most widely detergent removal method is the use of bio beads usage. Bio beads are hydrophobic polymers to which amphiphilic detergents can bind with higher affinity than lipids and proteins (**129, 131**). This method has been reported to be successful for hSGLT1 (kinne) and vSGLT. The method of reconstitution described for hSGLT1 in proteoliposomes uses Triton X-100 (TX100) (**65**). However, since the protein seems stable and well folded in the presence of F12 (see D.C. data and SEC profiles), removing F12 with bio beads instead of TX-100 seems a reasonable approach. Although protein detergent removal using biobeads is a process that has been characterized by a good deal of different detergents commonly used in protein purification and reconstitution, no studies existed on the elimination of F12. Following what is described in the literature in terms of characterization of the protein reconstitution in model membranes we undertook the study of (1) how F12 interacts with model membranes and (2) how effectively the detergent is removed by biobeads in the presence of lipids. The characterization of the interaction of foscholine with model membranes (liposomes), technically a solubilization, can be done by analyzing in which solubilization state the process is going to take place.

4.9.1 Liposome solubilitzation with F12

Protein reconstitution can be carried out at 3 different stages, depending on the amount of detergent used to solubilize the chosen lipids: Onset of solubilization, where mixed detergent lipid micelles and vesicles coexist, total solubilization where no vesicles are present and only micelles are formed and, last, below the onset of solubilization where no micelles are formed yet (**129, 130**). These three stages are defined in what is known as the lipid solubilization curve for any given detergent at a given lipid concentration. The solubilization curve of 1 mg lipid/mL of asolectin liposomes is shown in figure 4.9.1.

As it can be seen in the figure, the membranes at this lipid concentration are readily solubilized from quite low F12 concentrations and total lipid solubilization is reached at around 15 mM F12. Total lipid solubilization is the most usual starting condition for membrane protein reconstitution. F12 has a CMC of 1.5mM (anatrace ref) in water so, at around 15mM of F12, the final detergent concentration is around 10 times CMC. This information serves to settle the conditions for F12 removal with bio-beads. In order to work at total solubilization, it's necessary to settle the F12 concentration around 15 mM.

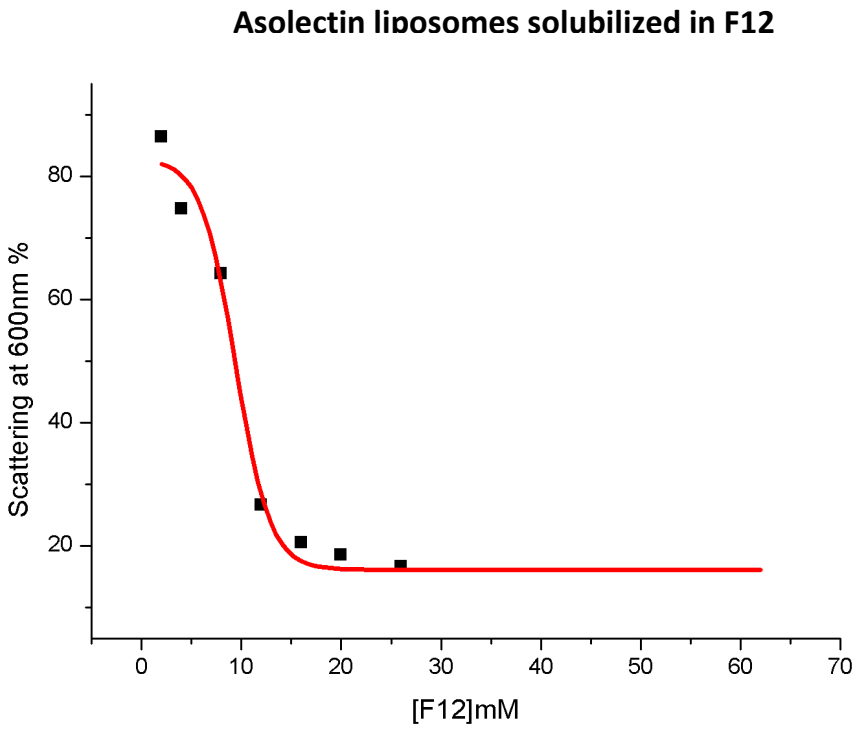
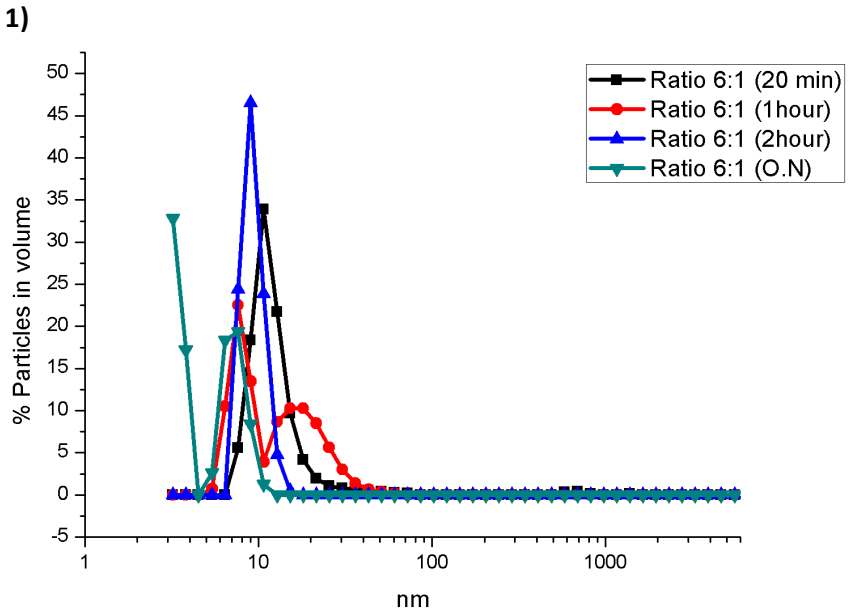


Figure 4.9.1: Turbidity measurements of 1mg/mL of asolectin lipids at different concentration of Fos-choline 12. The turbidity was measured with a fluorimeter by reading emitted light at 600 nm. The scattering measured at 600 nm is expressed in percentage of signal respect the initial value. Lipids were suspended in TKCL pH 7.6

4.9.2 Detergent removal from detergent/lipid micelles using bio-beads

This pilot experiment was carried out without protein and the results are shown in figure 4.9.2. Different biobeads/detergent (w/w) ratios were tested: 6:1, 12:1 and 24:1 and the diameters of the particles is suspension was measured at different times by dynamic light scattering.



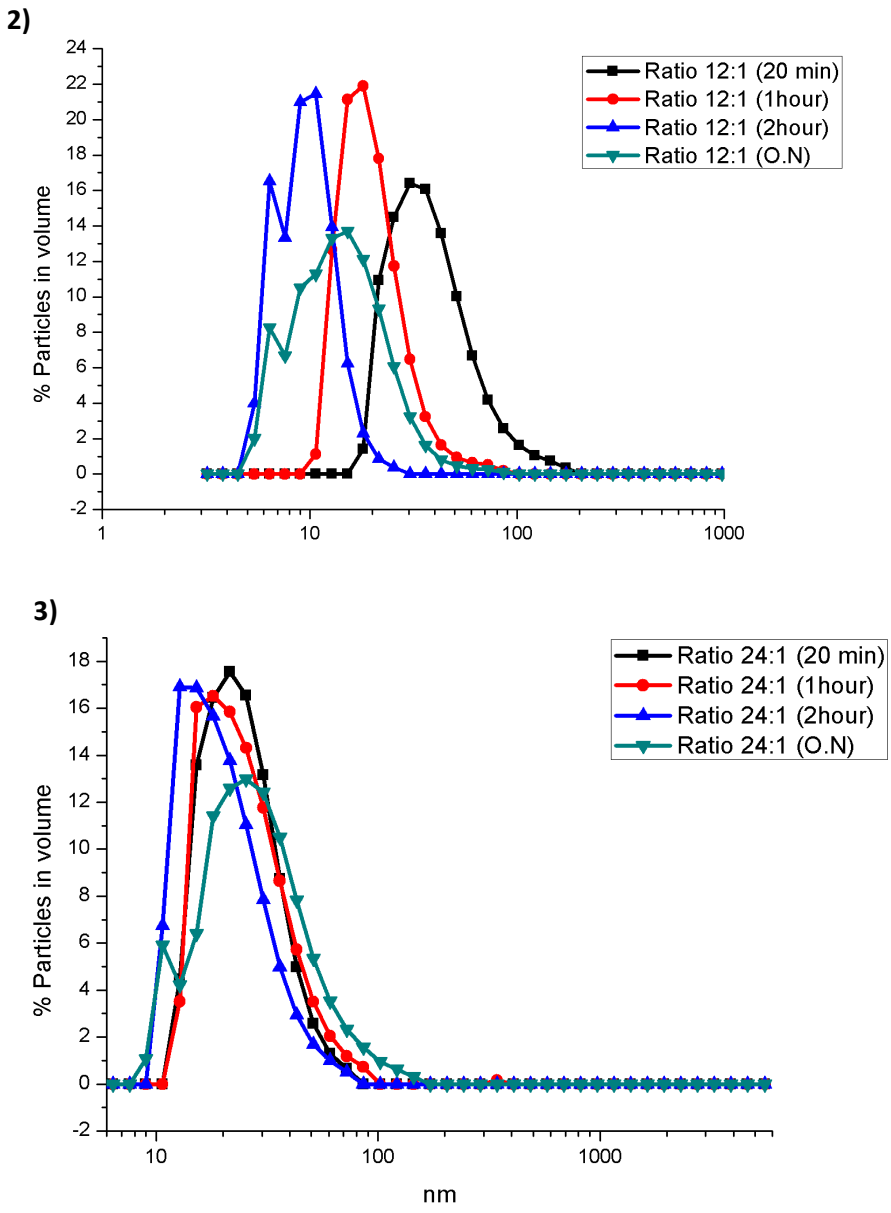


Figure 4.9.2: Dynamic Light Scattering (DLS) of reconstituted empty liposomes. Y axis represents a percentage of particles in volume. X axis represents the distribution of the particles in size (nm). 1) Ratio 6:1 (w/w) of bio-bead and detergent 2) Ratio 12:1 (w/w) of bio-bead and detergent 3) Ratio 24:1 (w/w) of bio-bead and detergent. Asolectin lipids were used mixed with Cholesterol (10:1) (w/w) lipid, cholesterol ratio.

At low ratios, 6:1, the particles mean size is at 10-20 nm at any given time, which is the size of micelles so, the detergent wasn't removed at all. At higher ratios, the mean size of the particles is higher. At 24:1 ratio, the size is well centered at 50nm, while at 12:1 ratio, the size distribution is broader and a little shifted to smaller size. Those results indicate that at 24:1 and 12:1 ratio the detergent was removed and forming liposomes because the size of the particles is consistent with the size (**133**) of Small-Unilamellar Vesicles (SUV). Also, it looks like the addition of the bio-beads

at different times didn't' affect the reconstitution because the size of the particles is very similar when adding at different times (2 hours, 1 hour, 20 minutes or directly over night).

Another very important variable is the amount of protein used for the reconstitution but the previous pilot reconstitution, where no protein was added, this variable couldn't be tested. Depending on what those proteoliposomes are going to be used for, it's recommended to work at higher or lower concentration. Normally, the amount of protein is expressed as a ratio of protein to lipid (w/w). For IR experiments, a low ratio 1:10 or even lower is needed but, for transport studies, a higher ratio of 1:100 is recommended instead. Although, working at low ratio like 1:10 or lower causes less protein incorporation (**134**).

4.9.3 hSGLT1 reconstituted with bio-beads in F12

Figure 4.9.3 shows a representative DLS graph after reconstitution of hSGLT1 with bio beads in the presence of F12 and asolectin. Several conditions were carried out without much changes on particle distribution (data not shown).

Although particles with a size (100 nm) in diameter are obtained which are characteristic (for their dispersity as well) of proteoliposmes, no transport, measured using radioactive gluocose could be observed with those liposomes (data not shown). This lack of transport could be a consequence of no protein incorporation or protein being incorporated incorrectly in the membrane, so, it's possible that the reconstitution did not work with F12 due to the detergent's nature.

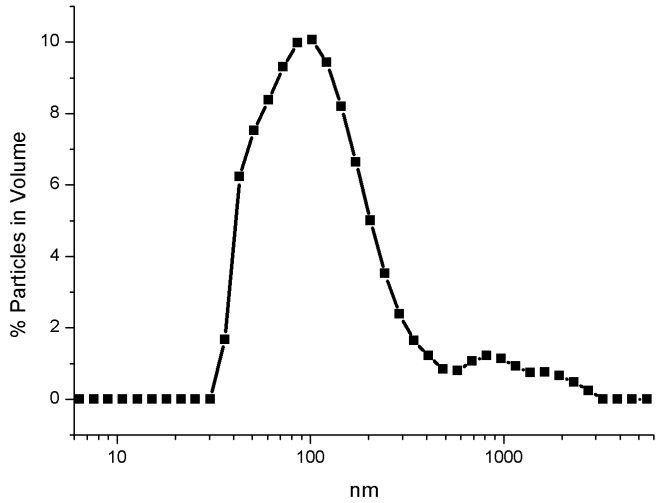


Figure 4.9.3: Dynamic Light Scattering (DLS) after a reconstitution of hSGLT1 with bio-beads in the presence of F12. Y axis represents percentage of particles in volume. X axis represents the distribution of the particles in size (nm). Lipids were prepared from asolectin at 1mg/mL with Cholesterol, 10:1 (w/w). The ratio of lipid to protein was 25:1 (w/w). A ratio of 24:1 biobeads/detergent was used.

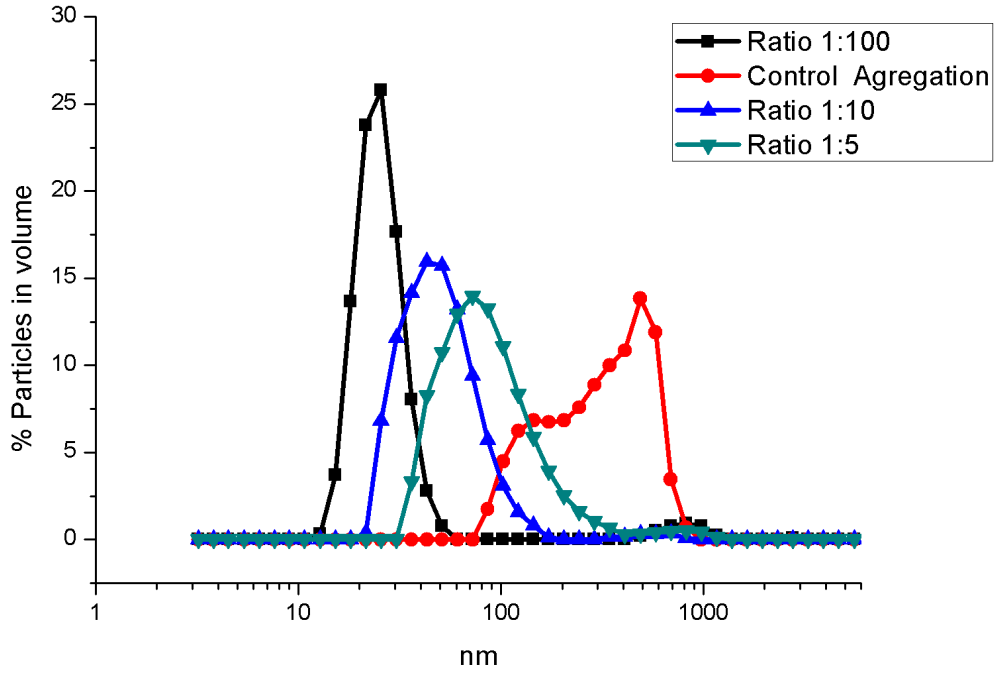


Figure 4.9.4.1: Dynamic Light Scattering (DLS) after a reconstitution of hSGLT1 with bio-beads in the presence of TX-100. Y axis is expressed as percentage of particles in volume. X axis is representing the distribution of the particles in size (nm). Lipids were prepared from PC+PS (4:1) at 1mg/mL with Cholesterol, 10:1 (w/w). In black, reconstitution with a protein lipid ratio of 1:100 (w/w). in blue; a ratio of 1:10 (w/w) and red; ratio 1:5 (w/w). Last, in red, an aggregation control was added where no lipid was added

4.9.4 hSGLT1 reconstituted with bio beads in TX-100

Since TX-100 has been widely used for protein reconstitution (**136, 137**) even in the case of hSGLT1 (**65**), a reconstitution attempt with TX-100 was carried out. Reconstitution was carried out at a ratio of 24:1 (w/w) bio-bead detergent with different ratios of protein-lipid.

The results show that it is as well possible to obtain particles which size is around 100 nm when reconstituting in the presence of T-X100, see figure 4.9.4.1. However, as it was the case for the reconstitution in the presence of F12, no glucose transport was detected.

4.9.5 Fourier transform infrared spectroscopy (FT-IR)

Since no transport study worked out with the product of any reconstitution trials, we checked structure of the reconstituted protein using Fourier-transform infrared spectroscopy (FTIR), a

technique which yields information on the secondary structure of the protein. An infrared spectrum of hSGLT1 reconstituted in asolectin liposomes is shown in figure 4.9.5.1.

Proteins have several bonds which absorb infrared light but the amide I (1600-1700 cm^{-1}) is the most widely used to get information on the structure of the protein and SGLT1 is no exception (**137**).

The amide I band (between 1600 and 1700 cm^{-1}) is mainly associated with the C=O stretching vibration (70-85%) and is directly related to the backbone conformation. Depending on the secondary structure of the protein those bands absorb at different wavelength and are sensitive to conformational changes which can provide information of the protein structure. Absorbance bands at 1660 and 1653 cm^{-1} are specific for α-helix while β-sheet structure has bands at 1640, 1634 and 1628 cm^{-1}.

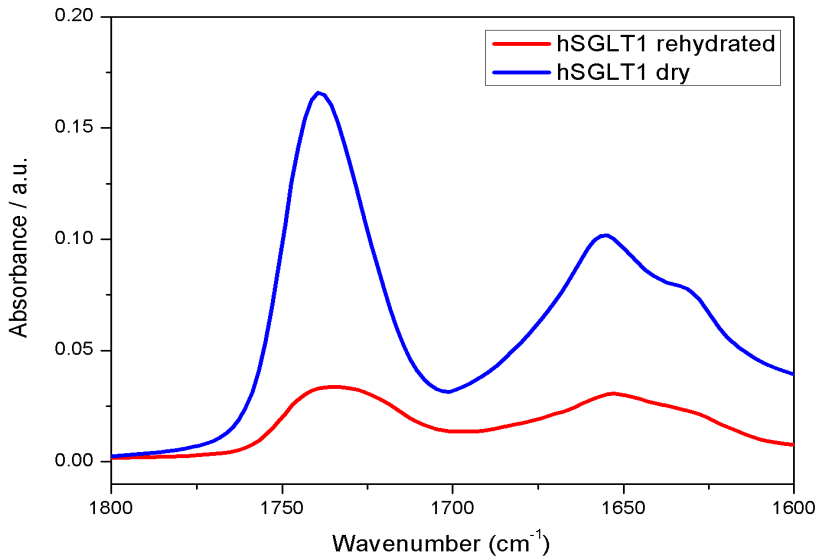


Figure 4.9.5.1: FT-IR spectrum of reconstituted WT hSGLT1 with TX-100 in PC+PS (1:4) (w/w) with Cholesterol (10:1) (w/w). In blue, the spectrum of dry WT hSGLT1 reconstituted proteoliposomes. In red, the rehydrated spectrum of the same

In the spectra in figure 4.9.5.1 (and the corresponding second derivative in figure 4.9.5.2) a band centered at 1730 cm^{-1} corresponds to the carbonyl of the ester bond from the lipid. The amide I band has a maximum at 1655 cm^{-1} and a shoulder at 1629 cm^{-1}. The 1655 cm^{-1} band is characteristic of α-helix structure but the 1629 cm^{-1} band is characteristic of antiparallel β strands caused from protein aggregation. There are other typical smaller bands as a result of protein aggregation like the 1694 cm^{-1} band which is also observed in the spectra. The intensity of the aggregation band at 1629 cm^{-1} is very high for a protein with expected high α-helix (85%) content.

Overall, these results are concluding that, even the protein is being reconstituted in the membrane; the secondary structure did change from micelles to liposomes causing protein aggregation making the protein not functional.

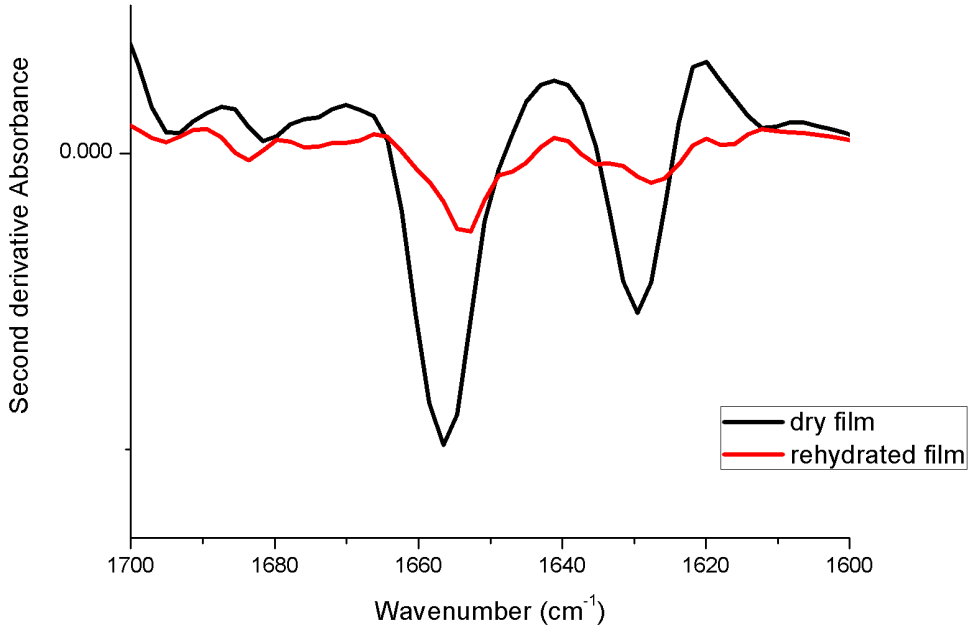


Figure 4.9.5.2: Second derivative of FT-IR spectrum of reconstituted WT hSGLT1 with TX-100 in PC+PS (1:4) (w/w) with Cholesterol (10:1) (w/w). In black, the spectrum of dry WT hSGLT1 reconstituted proteoliposomes. In red, the rehydrated spectrum of the same sample of WT hSGLT1 reconstituted liposomes. The 1629 cm^{-1} band and the 1650 cm^{-1} band are seen more clearly with the second derivative.

4.10 Fluorescence tryptophan spectra of hSGLT1

So far, no functional study was obtained but there's another described method that measures binding indirectly. The fluorescence spectra of WT hSGLT1 varies when substrates (glucose or sodium) are added and, also, when the inhibitor, phlorizin, is added (**63**). Glucose increases slightly (10-15%) the fluorescence intensity maximum while phlorizin causes the fluorescence to decay (50-60%). Phlorizin also causes a red shift in the fluorescence maximum while glucose doesn't (**138**).

These fluorescence experiments were carried out for reconstituted liposomes and for protein in detergent. Since FT-IR revealed that the protein is somehow incorporated incorrectly to the membrane, the results presented here are for WT hSGLT1 in micelles.

As observed in the figure, no changes are appreciated after adding sodium or glucose. The fluorescence intensity after adding glucose didn't change as it was reported in the past (**65, 138**).

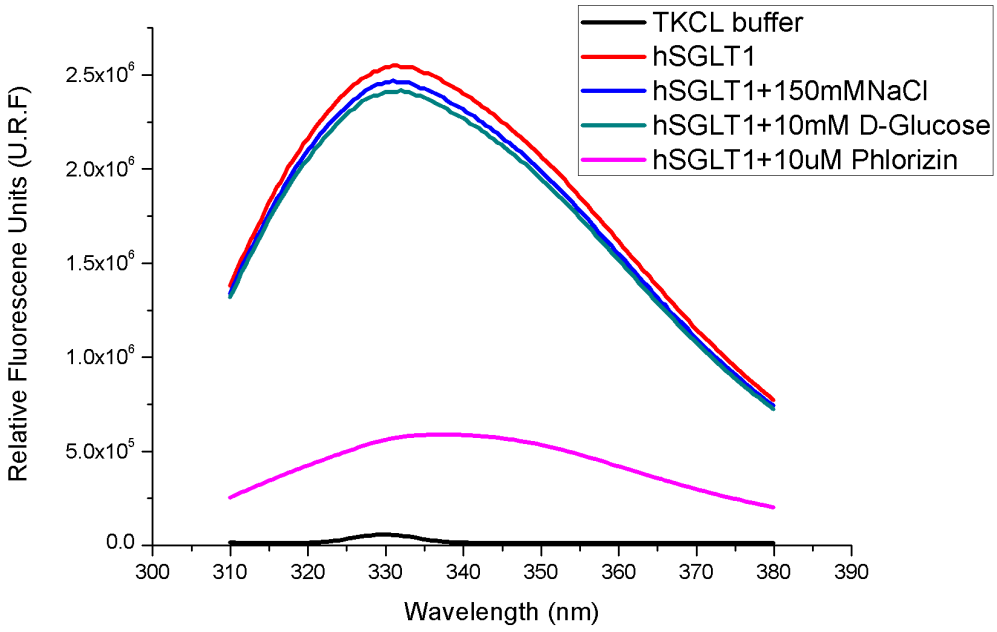


Figure 4.10.1: Fluorescence spectra of WT hSGLT1 at 1.6µM in TKCL buffer with 0.2% of F12.

Nevertheless, the fluorescence intensity after adding the inhibitor did decay as it was described and the maximum shifted to higher λ (blue shift). Surprisingly, it seems that the protein is somehow responding to the inhibitor but not the substrate. It must be added that phlorzin quenches the signal of tryptophan even in solution; therefore, the observed fluorescence decay could be unspecific.

Overall, these experiments couldn't be reproduced as it was reported except with the inhibitor responds.

4.11 Voltage Clamp in planar lipid membranes

Since no transport was achieved with proteoliposomes a different approach was decided in order to test the protein functionality: a planar lipid membrane system in which the protein is incorporated.

The Protein in detergent and lipids are inserted into a small hole (100µM) in plate of pretreated Teflon (polytetrafluoroethylene) with hexadecane that separates two different compartments. The hexadecane makes the Teflon hydrophobic which will cause lipids and hydrophobic compounds

like membrane proteins to attach to it. Two electrodes are positioned into each compartment of the chamber and those are connected to a signal amplifier. The amplified signal is connected to a recording system that registers any electric current flowing through the membrane. The system is set up so a constant voltage can be applied to between the two compartments while the current flowing through the membrane is measured. Since the Teflon hole is very small (10μm in diameter), it is possible to record the electric current caused by one or few protein molecules inserted into the membrane. Another clear advantage is that this technique requires a very low amount of protein (nmols). Initially the protein is added in one side of the chamber (with the lipididc membrane already formed in the hole) and inserts in the membrane spontaneously due to detergent dilution. The preformed artificial membrane is made up of DPHPC lipids which are very rigid and non-permeable (**139**).

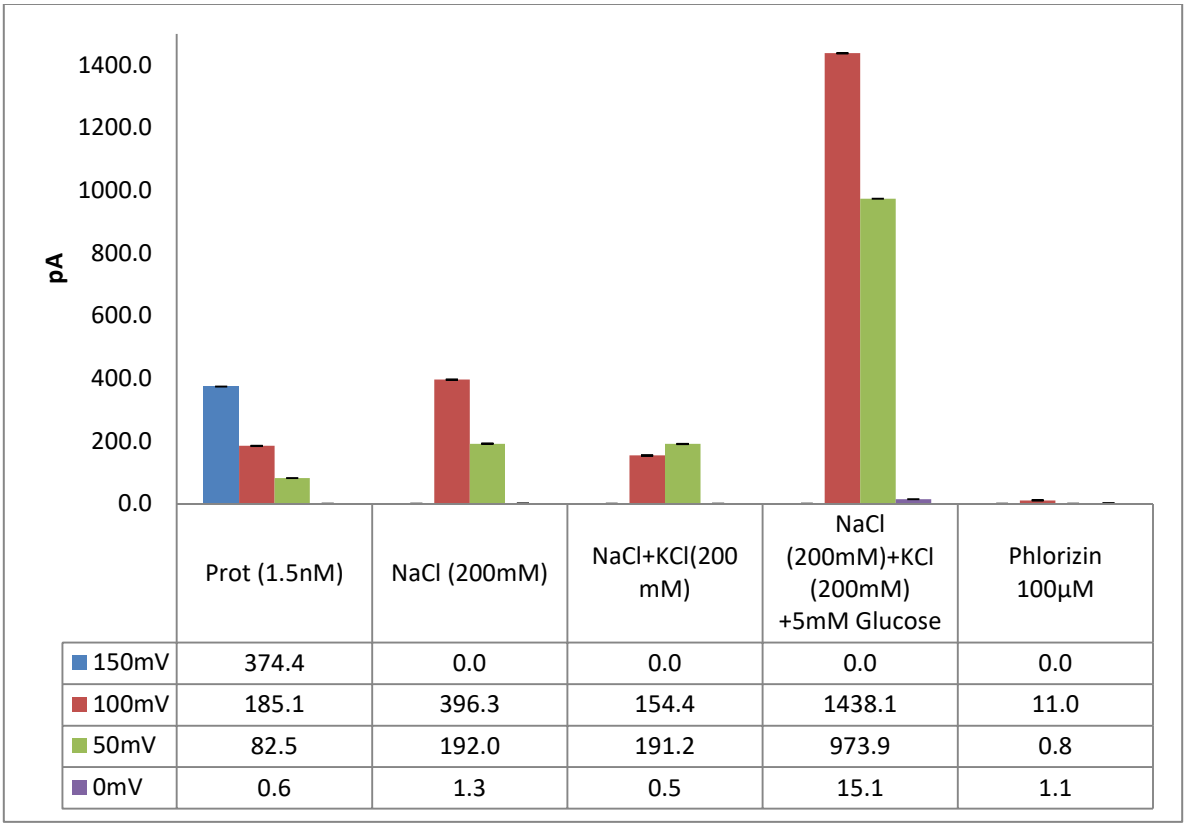
The results of an experiment with hSGLT1 inserted in a planar lipid membrane are shown in figures 4.11.1 and 4.11.2. When there's no protein, no current is flowing due to the high impermeability of the DPHPC lipids but, The insertion of the protein is detected in the form of a small electric current (ionic flow from one compartment to the other) caused by the destabilization that the protein insertion causes in the lipid bilayer.

As observed, when protein is added, the current depends on the applied voltage: at 150mV the current is 375 pA, at 100mV; the current is 185 pA and at 50mV is 82 pA. So, initially, it can be concluded that the protein was inserted in the membrane. The idea at this stage is to add the protein into the compartment in a way (for a voltage difference) that causes the smallest possible detectable electrical current; One that will permit stablishing that the protein is inserted into the membrane.

hSGLT1 works as a glucose transporter in the presence of a Na^+ gradient. In order to check whether hSGLT1 is functional in the planar membranes system, in which we can only measure ion flows we need to establish that: (1) there is electric current when we establish a Na^+ gradient between the two compartment in the presence of glucose; there is no electric current when we establish a K^+ gradient across the membrane (the transporter is specific for Na^+); (3) there is no electric current in the presence of an specific inhibitor of glucose transport such as phlorizin. This is what is illustrated in figures 4.11.1 and 4.11.2. In the presence of only a K^+ gradient, in the absence of glucose or in the presence of the inhibitor, only currents equivalent to those caused by the mere protein insertion are detected. However, in the presence of a Na^+ gradient, a clearly higher electric current is detected, which confirms that hsGLT1 functions as a transporter when inserted into the planar lipid membrane.

A remarkable feature in these experiments is that after adding the glucose without applying any voltage, a current is still appreciated. At 0mV the current is 15 pA and, even is a low current, no current is appreciated in any tested conditions at 0mV. This is very interesting because, in this situation, no extra work is applied to the system a part from the electrochemical gradient generated by adding sodium chloride to one side of the chamber. When no sugar was added, the

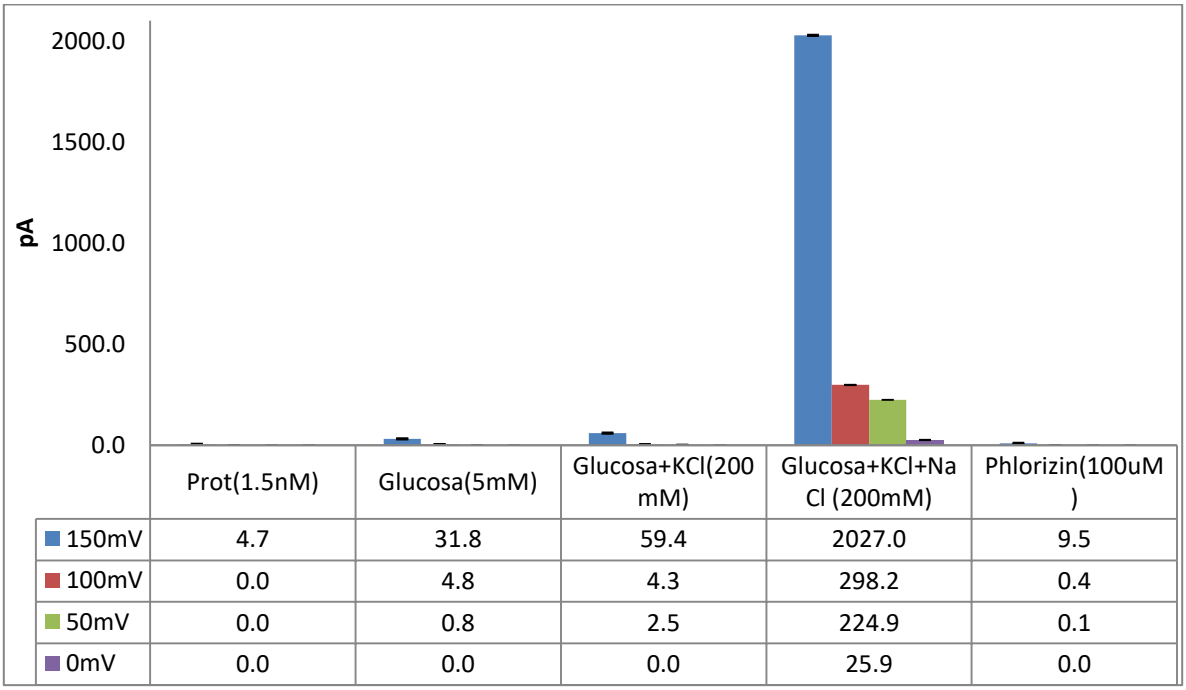
current at 0mV was 1pA but the electrochemical gradient was still present, therefore, the current appreciated after adding sugar at 0mV it has to be a transport process.



	Prot (1.5nM)	NaCl (200mM)	NaCl+KCl(200 mM)	NaCl (200mM)+KCl (200mM) +5mM Glucose	Phlorizin 100µM
150mV	374.4	0.0	0.0	0.0	0.0
100mV	185.1	396.3	154.4	1438.1	11.0
50mV	82.5	192.0	191.2	973.9	0.8
0mV	0.6	1.3	0.5	15.1	1.1

Figure 4.11.1: Current of planar lipid membranes of WT hSGLT1 at different voltages. 50 µg of DPHPC (1,2-diphytanoyl-sn-glycero-3-phosphocholine) was the chosen lipid. The current is expressed in the y axis as pico Amp (pA). In the x axis each tested conditions. All additions were done subsequently. First column; only protein was added, second column; 200mM sodium chloride, third column; 200mM KCl, fourth lane; 5mM D-glucose and last 100µM of hSGLT1 inhibitor was added (phlorizin). Each chamber was already in 2mL of 150 mM KCl pH 7.3

Overall, both experiments confirm that only when sodium and glucose are added a huge current is observed which clearly indicates a mechanism of transport that is inhibited specifically with phlorizin.



	Prot(1.5nM)	Glucosa(5mM)	Glucosa+KCl(200 mM)	Glucosa+KCl+NaCl (200mM)	Phlorizin(100uM)
150mV	4.7	31.8	59.4	2027.0	9.5
100mV	0.0	4.8	4.3	298.2	0.4
50mV	0.0	0.8	2.5	224.9	0.1
0mV	0.0	0.0	0.0	25.9	0.0

Figure 4.11.2: Current of planar lipid membranes of WT hSGLT1 at different voltages. 50 µg of DPHPC (1,2-diphytanoyl-sn-glycero-3-phosphocholine) was the chosen lipid. The current is expressed in the y axis as pico Amp (pA). In the x axis each tested conditions. All additions were done subsequently. First column; only protein was added, second column; 200mM sodium chloride, third column; 200mM KCl, fourth lane; 5mM D-glucose and last 100µM of hSGLT1 inhibitor was added (phlorizin). Each chamber was already in 2mL of 150 mM KCl pH 7.3.

4.12 General discussion

In a general perspective, hSGLT1 was purified successfully for further structural studies such as protein crystallization and spectroscopic studies, but with some unexpected drawbacks. The yield of the expression system was good enough to obtain protein in the mg range, which is needed for crystallization trials. hSGLT1 expression was actually published once before by a group in Germany, and this was very encouraging (**65**). These previous reports permitted to set up our own expression system and investigate, by using the hSGLT1 gene fused to eGFP, the optimal expression and purification conditions in our own lab. Following the green fluorescence of the GFP facilitates the control and optimization of the protein expression. Although we did not succeed in cleaving GFP from hSGLT1, the results with the GFP construct were useful to undertake the expression and purification of the non-fluorescent WT hSGLT1 construct.

Screenings were carried out in order to enhance and optimize the final yield of protein production, including: expression conditions (methanol induction time, incubation temperature, expression media, etc.) and solubilization screening. As our results indicate, the expression conditions were successful since higher expression levels have been achieved**.** As far as the solubilization screening is concerned, we could not identify a better detergent than the one (F12) described in the literature. Nevertheless, some details were improved, such as the total amount of detergent used respect to the membranes or the ability of the DTT to extract more protein from the membrane.

It must be pointed out that most of the previous results, especially technical details regarding hSGLT1 expression and detergent screening were not published. Therefore, it's difficult to compare or evaluate our expression system in the literature since it seems no other group has attempted to reproduce those results (**65)**.

In our hands, the secondary structure of purified protein, as seen in the CD spectra, is consistent to what was described in the past (**128**) which clearly indicates that the protein is properly folded after purification in a predimoninantly α-helical structure, therefore, the protein should be functional within the detergent-micelles complex. Our functional transport study, carried out with planar lipid membranes revealed clearly that the protein is active because it responds to both substrates, also to the inhibitor, which is very promising for future pharmacological experiments. Isothermal titration calorimetry (ITC) was done with glucose as a substrate to check for binding, but results were inconclusive. As observed previously, some discrepancies between the published data of WT hSGLT1 expressed in *P. pastoris* and our data were observed and are actually an issue that must be addressed. In the choice of a human sugar transporter, the main reason why hSGLT1 was chosen for expression towards structural biology was due to previous positive results.

In a recent editorial by Nature (**141**) on a survey on reproducibility (understood as different scientists using the same methods obtaining the same results), we can read: 'The ability to reproduce experiments is at the heart of science, yet failure to do so is a routine part of research'. The editorial concludes that 'more steps are needed starting with a discussion in the research community on how to properly credit, and talk to each other about, attempted replications'.

This is an issue with which we have come face to face at different points during the time dedicated to the expression of hSGLT1 in *P. pastoris*, its purification and its characterization. hSGLT1 was selected as a transporter to study since it was suitable for the type of biophysical studies we have expertise with (**140**) and mainly because it was published once before (**65**).

Although we were successful in expressing and purifying the transporter in *P. pastoris*, there are some issues, one related to the purification process and a few others related to the functional/structural characterization of the protein, for which we were not able to reproduce the results that had been previously described.

In relation to the hSGLT1 purification, although we got excellent chromatograms when re-purifying the protein through a SEC column, the electrophoretic analysis (SDS-PAGE, Western Blots) of the chromatography outcome yielded always a thick but fuzzy band which corresponds to hSGLT1 as

analyzed by mass spectroscopy. Apart from that, other bands are also present even after re-purification, but not after FLAG-tag purification. Therefore these additional bands must be impurities migrating together with the hSGLT1 fuzzy band. Moreover, the fuzzy band is not related to any kind of major post-translation modification, such as glycosylation, since it is still observed with the N248A mutant purification and the mass spectroscopy of the WT. Therefore, we concluded that such unusual band smearing is probably caused by some sort of artifact generated by the SDS-PAGE.

The observed impurities and the fuzzy band of hSGLT1 were not observed in the previous publication (**65**). In this sense, the impossibility to observe such and electrophoretic band meant that we lacked of the details to reproduce the neat and extremely thin electrophoretic band previously reported (**65**). Also, the purity of hSGLT1 after the Ni-NTA yielded pure protein in a single purification step, which we could not get. It must be pointed out that no chromatogram was shown in the previous work since no SEC purification was done.

Lack of reproducibility has been an issue in other aspects such asthe reconstitution into model membranes (yielding functional proteoliposomes) and our failure to get any response to glucose binding via changes in the detergent-solubilized protein's intrinsic fluorescence.

The impossibility to reproduce previously described results lies in two main aspects: expression of human membrane proteins is challenging, and in our lab we have previous experience in expressing mammalian channels, but not transporters, in *P. pastoris.* Thus, we asked for the collaboration of a world-renowned laboratory in hSGLT1 physiology and structural biology at UCLA, under the supervision of Professor Ernest Wright with the goal to get functional insightful information from transport assays. The efforts in Professor's Wright lab to get hSGLT1 correctly embedded into model membranes or further purify the protein did not yield any positive results either.

We spent nearly four months at UCLA to evaluate the behavior of hSGLT1, but also to certify that the work done in handling hSGLT1 so far was properly done. Finally, the main conclusion of this stay helped to cement our opinion that the problems we were facing with hSGLT1 were shared by the community dedicated to the study of this transporter, since professor Wright has been totally dedicated to this transporter for 40 years.

The impossibility to reproduce the results from previous studies (**65**), now confirmed by world-renowned experts. This has to be taken as a window for criticism, but in regards to hSGLT1, has to be taken as an opportunity to value what we have achieved, and what information can be derived from discrepancies between previous studies and our study.

Firstly, it will be important to clarify the type and source of the contaminant proteins after the Ni-NTA step. The contaminant proteins are observed after Ni-NTA when using SMD1168H strain, but apparently not for GS115 strain (**65**), opening the perspective of a *P. pastoris* strain screening, using strains such as GS115, X33, K7H, etc.

Secondly, and more important is to investigate why the reconstitution has not worked, although it has been done as described in the literature. In the FT-IR spectrum after reconstitution, the protein it is partially aggregated, probably, due to low solubility in TX-100. If that is the case, trying to reconstitute the protein without TX-100 using F12 instead, in which the protein is already soluble, might solve the problem. Therefore, it would be mandatory to study the detergent removal with bio-beads of F12 with radioactive detergent like it was done with other detergents in the past. Another option would be to replace TX-100 for other known removable detergents like DDM. Also, it could be possible to work with different lipids and cholesterol compositions since the lipid environment affects the structure. Another approach would be to change the initial solubilization stage of the lipid-detergent-protein mixture. As mentioned before, the initial mixture can be totally solubilized (above the solubilization threshold) or not and, this might affect the activity of the protein. A completely different approach could be to change the method of reconstitution, for example, instead of using bio-beads, a fast dilution method can be done or, even, using cyclodextrin. All of the above are different approaches to study hSGLT1 reconstituiton into proteoliposomes, which for obvious time reasons fall out of the scope of the present Ph.D. thesis.

Thirdly, get more information and knowledge of the protein incorporation in planar lipid membranes. This technique has been successfully used to study protein channels such as VDAC but not transporters in detergent-micelles. The same concept used in proteliposomes can be used to study the binding key residues where the substrate binds and crosses the membrane. Thus, this technique opens the possibility for pharmacological and structure-function relationship studies. Some residues related to substrate binding are already described but not all of them. Planar lipid membranes requires low amounts of purified protein and, even although in some mutants the expression is very low, it would be still enough for this technique. Using site-directed mutagenesis the substrate binding sites for hSGLT1 can be further described, even towards the discovery of new hSGLT1 agonists and antagonists.

Different techniques and analysis were attempted to assess hSGLT1 functionality like: Isothermal tritation calorimetry (ITC), fluorescence sugar binding (2-NBDG), radioactive sugar for transport assays with membranes and cells and, last, voltage-clamp with planar lipid membranes. Most of those techniques are based on the preparation of proteoliposomes and did not work, except the voltage-clamp technique which gave positive results, so; the take-home message is, especially when working with recombinant mammalian proteins, to explore all possibilities and options before giving up and concluding that a protein is not functional. Clearly the results were not what it was initially expected but it must be emphasize that, despite all the hindrances present in the expression of hSGLT1, a good final yield of functional protein was achieved. This is a great starting point to keep working with the protein for structural studies like, for example, crystallization or other techniques. It must be point out that the voltage-clamp with planar lipid membranes can be applied in the future for hSGLT1 mutants or even perform more complex experiments; the potential of this technique must be addressed in the future, not only for hSGLT1, but also for other membrane proteins.

5. Conclusions

1. hSGLT1 was cloned successfully in pJIN vector (with and without eGFP) and the resulting vector was electroporated correctly. Electroporated colonies were screened for both vectors satisfactorily.
2. Expressed protein is localized in membrane compartments as seen in the confocal microscopy images.
3. Detergent screening was done with hSGLT1 and, a part from Fos-choline 12, no other detergent solubilized better the protein
4. Expression conditions and solubilization conditions were optimized to achieve the maximum protein as possible. A final yield of 1-2 mg/mL was obtained for both constructs.
5. Protein was successfully purified after a NI-NTA column, FLAG-tag purification and SEC.
6. Circular dichroism revealed a structure of a protein with a high α-helix (85%) content which is consistent for a 14 transmembrane domains
7. It was not possible to incorporate a fully functional structure of hSGLT1 in liposomes as described before.
8. A voltage-clamp experiment in planar lipid membranes with hSGLT1 showed an ion specific influx when substrates (NaCl and D-Glucosa) were added but not when KCl was added. Those results are consistent with a specific transport experiment of hSGLT1

6. Bibliography

1) McIntosh, T. J., and Simon, S. A. (2006) Roles of bilayer material properties in function and distribution of membrane proteins, *Annual. Rev. Biophys Biomol Struct*. 35, 177-198.

2) Locher, K. P., Bass, R. B., and Rees, D. C. (2003) Structural biology. Breaching the barrier, *Science* 301, 603-604.

3) Lappano R., Maggiolini M. (2011) G protein-coupled receptors: novel targets for drug discovery in cancer, *Nature Reviews Drug Discovery* 10(1), 47–60.

4) Sandra Tan, Hwee Tong Tan, Maxey C. M., Chung Professor.(2008) Membrane proteins and membrane proteomics, *Proteomics* 8 (19) ,3924-3932.

5) Stevens T. J. & Arkin I. T. (2000) Do more complex organisms have a greater proportion of membrane proteins in their genomes? , *Proteins* 39 (4), 417-420.

6) Alberts, B., Johnson, A., Lewis, J. (2002) Molecular Biology of the Cell 4th edition. Principals of membrane transporters.

7) Okada, Y. (2004) Ion channels and transporters involved in cell volume regulation and sensor mechanisms, *Cell Biochemistry and Biophysics* 41, 233-258.

8) Vandenberg, R. J., and Ryan, R. M. (2005) How and why are channels in transporters? , *Sci STKE* 280, pe17.

9) Le Coutre, J., and Kaback, H. K. (2000) Structure-function relationships of integral membrane proteins: membrane transporters vs channels, *Biopolymers* 55, 297-307.

10) Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H., and Bruford, E. A. (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins, *Pflugers Arch*. 447, 465-468.

11) Lei He, Vasiliou, K., Nebert, D. W. (2009) Analysis and update of the human solute carrier (SLC) gene superfamily, *Hum. Genomics* 3 (2), 195-205.

12) Hoglund, P. J., Nordstrom, K. J. V., Schioth, H. B., Fredriksson, R. (2010) The Solute Carrier Families Have a Remarkably Long Evolutionary History with the Majority of the Human Families Present before Divergence of Bilaterian Species, *Molecular Biology and Evolution* 28 (4), 1531–1541.

13) Ishida, N., and Kawakita, M. (2004). Molecular physiology and pathology of the nucleotide sugar transporter family (SLC35). *Pflugers Arch*. 447, 768-775.

14) Hagenbuch, B., and Meier, P. J. (2004) Organic anion transporting polypeptides of the OATP/ SLC21 family: phylogenetic classification as OATP/ SLCO superfamily, new nomenclature and molecular/functional properties, *Pflugers Arch*. 447, 653-665.

15) Wright, E.M. (2013) Glucose transport families SLC5 and SLC50, *Mol. Aspects Med*. 34 (2-3), 183-96.

16) Wright, E. M., and Turk, E. (2004) The sodium/glucose cotransport family SLC5. *Pflugers Arch* 447, 510-518.

17) Hediger, M.A., Coady, M.J., Ikeda, T.S., Wright, E.M. (1987). Expression cloning and cDNA sequencing of the Na+/glucose co-transporter. *Nature* 330 (6146), 379–381.

18) Mallee, J.J., Atta, M.G., Lorica, V., Rim, J.S., Kwon, H.M., Lucente, A.D., Wang, Y., Berry, G.T. (1997) The structural organization of the human Na+/myo-inositol cotransporter (SLC5A3) gene and characterization of the promoter. *Genomics* 46 (3), 459–465

19) Panayotova-Heiermann, M., Eskandari, S., Turk, E., Zampighi, G. A., and Wright, E. M. (1997) Five transmembrane helices form the sugar pathway through the Na+/glucose cotransporter. *J Biol Chem* 272, 20324-20327.

20) Vallon, V., Platt K.A., Cunard, R., Schroth, J., Whaley, J., Thomson, S. C., Koepsell, H., Rieg T. (2011) SGLT2 Mediates Glucose Reabsorption in the Early Proximal Tubule. *J. Am. Soc. Nephrology* 22 (1), 104-112.

21) Nagamori, S., Vázquez-Ibar, J.L., Weinglass, A.B., Kaback, H.R. (2003) In vitro synthesis of lactose permease to probe the mechanism of membrane insertion and folding, *Journal of Biological Chemistry* 278 (17), 14820-14826.

22) Bassilana, M., Pourcher T., Leblanc, G. (1988) Melibiose Permease of *E.coli*, *Journa of Biological Chemistry* 263 *(20), 9663-9667.*

23) Ganapathy, V., Thangaraju, M., Gopal, E., Martin, M. P., Itagaki, S., Miyauchi, S., Prasad, D. P. (2008) Sodium-coupled Monocarboxylate Transporters in Normal Tissues and in Cancer, *AAPS. J.* 10(1), 193.

24) Hediger, M.A., Coady, M.J., Ikeda, T.S., Wright, E.M. (1987) Expression clonning and cDNA sequencing of the Na/glucose co-transporter, *Nature* 330 (6146), 379-81.

25) Lee, W.S., Kanai, Y., Wells, R. G., and Hediger, M. A. (1994) The high affinity Na^+/glucose cotransporter. Re-evaluation of function and distribution of expression, *J. Biol. Chem*. 269, 12032–12039.

26) Chen, X. Z., Coady, M. J., Jackson, F., Berteloot, A., Lapointe, J.Y. (1995) Thermodynamic determination of the Na^+: glucose coupling ratio for the human SGLT1 cotransporter, *Biophys. J*. 69 (6), 2405-14.

27) Panayotova-Heiermann, M., Loo, D. D., Kong., C. T., Lever, J.E., Wright, E.M. (1996) Sugar Binding to Na /Glucose Cotransporters Is Determined by the Carboxyl-terminal Half of the Protein, *J. Biol. Chem*. 271 (17), 10029-34

28) Coady, M. J., Jalai, F., Bissonnette, P., Cartier, M., Wallendorff, B., Lemay G., Lapointe, J. (2000) Functional studies of a chimeric protein containing portions of the Na/glucose and N/myo-inositol cotransporters, *Biochim. Biophys. Acta.* 1466 (1-2), 139-50.

29) Tyagi, N.K., Puntheeranurak, T., Raja, M., Kumar, A., Wimmer, B., Neundlinger, I., Gruber, H., Hinterdorfer, P., Kinne, R. K. (2011) A biophysical glance at the outer surface of the membrane transporter SGLT1, *Biochim. Biophys. Acta.* 1808 (1), 1-18.

30) Neundlinger, I., Puntheeranurak, T., Wilding, L., Rankl, C., Wang, L.X., Gruber, H.J., Kinner, R.K., Hinterdorfer, P. (2014) Forces and Dynamics of Glucose and Inhibitor Binding to Sodium Glucose co-transporter SGLT1 Studied by Single Molecule Force Spectroscopy, *J. Biol. Chem.* 289 (31), 21673-83.

31) Smith, C.D., Hirayama, B. A., Wright EM. (1992) Baculovirus-mediated expression of the Na+/glucose cotransporter in Sf9 cells, *Biochim. Biophys. Act.* 1104(1), 151-9.

32) Quick, M., Wright, E.M. (2002) Employing Escherichia coli to functionally express, purify, and characterize a human transporter. *PNAS* 99 (13), 8597–8601.

33) Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht A., Hirayama, B.A., Wright, E.M., Abramson, J. (2008) The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na^+/Sugar Symport, *Science* 321 (5890), 810-4.

34) Wright, E.M., Loo, D. D., Hirayama, B.A. (2011) Biology of Human Sodium Glucose Transporters, *Physiol. Rev.* 91 (2), 733-94.

35) Sala-Rabanal, M., Hirayama, B.A., Loo, D.D., Chaptal, V., Abramson, J., Wright, E.M. (2012) Bridging the gap between structure and kinetics of human SGLT1. *Am. J. Phsyiol. Cell Physiol.* 302 (9), 1293-305.

36) Turk, E., Kim, O., le Coutre, J., Whitelegge, J.P., Eskandari, S., Lam, J.T., Kreman, M., Zampighi, G., Faull, K.F., Wright, E.M. (2000) Molecular Characterization of Vibrio parahaemolyticus vSGLT: a model for sodium-coupled sugar cotransporters, *J. Biol. Chem.* 275 (33), 25711-6.

37) Meinild, A. K., Loo, D.D., Hirayama, B.A., Gallardo, E., Wright, E.M. (2001) Evidence for the involvement of Ala 166 in coupling Na+ to sugar transport through the human Na+/glucose cotransporter*, Biochemistry* 40 (39), 11897-11904.

38) Díez-Sampedro, A., Barcelona S. (2011) Sugar Binding Residue Affects Apparent Na+ Affinity and Transport Stoichiometry in Mouse Sodium/Glucose Cotransporter Type 3B, *J. Bio. Chem*. 286 (10), 7975-82.

39) Napoli, R., Hirshman, M.F., Horton, E.S. (1995) Mechanisms and time course of impaired skeletal muscle glucose transport activity in streptozocin diabetic rats, *J. Clin Invest*. 96(1), 427-37.

40) Diez-Sampedro, A., Loo, D. D., Wright, E. M., Zampighi, G. A., and Hirayama, B. A. (2004) Coupled sodium/glucose cotransport by SGLT1 requires a negative charge at position 454, *Biochemistry* 43 (41),13175-84.

41) Diez-Sampedro, A., Wright, E. M., and Hirayama, B. A. (2001) Residue 457 controls sugar binding and transport in the Na+/glucose cotransporter, *J Biol Chem 276 (52)*, 49188-49194

42) Adelman, J.L., Ghezzi, C., Bisignano, P., Loo, D.D., Choe, S., Abramson, J., Rosenberg, J.M., Wright, E.M., Grabe, M. (2016) Stochastic steps in secondary active sugar transport. Joshua L.adelman, chiara ghezzi, paola bisignano, *PNAS* 113 (27), 3960-3966.

43) Loo, D. D., Hazama, A., Supplisson, S., Turk, E., and Wright, E. M. (1993) Relaxation kinetics of the Na+/glucose cotransporter, *PNAS* 90 (12), 5767-5771

44) Loo, D. D., Hirayama, B. A., Gallardo, E. M., Lam, J. T., Turk, E., and Wright, E. M. (1998) Conformational changes couple Na+ and glucose transport, *Proc. Natl. Acad. Sci.* 95 (13), 7789-7794.

45) Peerce, B. E., and Wright, E. M. (1984) Conformational changes in the intestinal brush border sodium-glucose cotransporter labeled with fluorescein isothiocyanate, *Proc. Natl. Acad. Sci.* 81 (7), 2223-6.

46) Lo, B., Silverman, M. (1998) Cysteine scanning mutagenesis of the segmentbetween putative transmembrane helices IV and V of the high affinityNa^+/Glucose cotransporter SGLT1. Evidence that this region participates in the Na^+ and voltage dependence of the transporter, *M.J Biol Chem*. 273 (45), 29341–29351.

47) Li, J., Tajkhorshid, E. (2012) A Gate free Pathway for Substrate Release from the Inward Facing State of the Na+ Galactose Transporter, *Biochim. Biophys. Acta* 1818(2), 263-271.

48) Hediger, M.A., Mendlein, J., Lee, H.S., Wright E.M. (1991) Biosynthesis of the cloned intestinal Na^+/glucose cotransporter. *Biochim. Biophys. Acta* 1064 (2), 360–364.

49) Hirayama, B.A., Wright E.M. (1992) Glycosylation of the rabbit intestinal brush border Na^+/glucose cotransporter. *Biochim. Biophys. Acta* 1103 (1), 37–44-

50) Eskandari, S., Wright, E.M., Kreman, M., Starace, D.M., Zampigh, G.A. (1998) Structural analysis of cloned plasma membrane proteins by freeze-fracture electron microscopy. *PNAS* 95 (19), 11235–11240.

51) Stevens, B.R., Fernandez, A., Hirayama, B., Wright, E.M., Kempner, E.S. (1990) Intestinal brush border membrane Na^+/glucose cotransporter functions in situ as a homotetramer. *PNAS* 87 (4), 1456–1460.

52) Strugatsky, D., Gottschalk, K. E., Goldshleger, R., Bibi, E., and Karlish, S. J. (2003) Expression of Na+, K+-ATPase in Pichia pastoris: analysis of wild type and D369N mutant proteins by Fe2+-catalyzed oxidative cleavage and molecular modeling. *J Biol Chem* 278 (46), 46064-46073.

53) Cai, J., Daoud, R., Georges, E., and Gros, P. (2001).Functional expression of multidrug resistance protein 1 in Pichia pastoris. *Biochemistry 40 (28)*, 8307-8316.

54) Cai, J., and Gros, P. (2003) Overexpression, purification, and functional characterization of ATP-binding cassette transporters in the yeast, Pichia pastoris. *Biochim Biophys Acta* 1610 (1), 63-76.

55) Bieszke, J. A., Spudich, E. N., Scott, K. L., Borkovich, K. A., and Spudich, J. L. (1999) A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. *Biochemistry 38* (43), 14138-14145.

56) Doring, F., Klapper, M., Theis, S., and Daniel, H. (1998) Use of the glyceraldehyde-3-phosphate dehydrogenase promoter for production of functional mammalian membrane transport proteins in the yeast Pichia pastoris. *Biochem Biophys Res. Commun.* 250 (2), 531-535.

57) Talmont, F., Sidobre, S., Demange, P., Milon, A., and Emorine, L. J. (1996) Expression and pharmacological characterization of the human mu-opioid receptor in the methylotrophic yeast Pichia pastoris. *FEBS Lett*. 394, 268-272.

58) Weiss, H. M., Haase, W., and Reilander, H. (1998) Expression of an integral membrane protein, the 5HT5A receptor. *Methods Mol Biol* 103, 227-239.

59) Cereghino, J. L., and Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev* 24, 45-66.

60) Cregg, J. M., Cereghino, J. L., Shi, J., and Higgins, D. R. (2000) Recombinant protein expression in Pichia pastoris. *Mol Biotechnol* 16, 23-52.

61) Higgins, D. R., and Cregg, J. M. (1998) Introduction to Pichia pastoris. *Methods Mol Biol* 103, 1-15.

62) McPherson, A. (2004) Introduction to protein crystallization. Volume 34, Issue 3, 254–265.

63) Mus-Veteau, I. (2010) Heterologous Expression of Membrane Proteins for Structural Analysis, *Methods Mol Biol*. 601, 1-16.

64) EasySelect Pichia Expression Kit for Expression of recombinant proteins using pPCIZ and pPICZalfa in pichia pastoris

65) Tyagi, N.K., Goyal, P., Jumar, A., Pandey, D., Siess, W., Kinner, R.K. (2005) High-Yield Functional Expression of Human Sodium/D-Glucose Cotransporter1 in Pichia pastoris and Characterization of Ligand-Induced Conformational Changes as Studied by Tryptophan Fluorescence, *Biochemistry* 44 (47), 15514-15524.

66) Nickoloff, J.A., Electroporation protocols for microorganisms. Methods in molecular Biology, volume 47, 22-26.

67) Calvin, N.M., Hanawalt, P.C. (1988) High-efficiency transformation of bacterial cells by electroporation, *J Bacteriol* 170 (6), 2796-2801.

68) Brooks, C.L., Morrison, M., Lemieux, M.J. (2013) Rapid expression screening of eukaryotic membrane proteins in Pichia pastoris, *Protein Sci*. 22(4), 425–433.

69) Bergksessel, M., Guthrie, C. (2013) Methods in Enzymology, chapter twentyfive-colony PCR. Volume 529.

70) Lõoke, M., Kristjuhan, A., Kristjuhan, K. (2011) Extraction of Genomic DNA from Yeast for PCR-based applications. *Biotechniques* 50(5), 325-328.

71) Marone, M., Mozzetti, S., De Ritis, D., Pierelli, L., Scambia, G. (2001) Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample, *Biol Proced Online* 3, 19-25.

72) Döring, F., Michel, T., Rösel, R., Nickolaus, M., Daniel, H. (1998) Expression of the mammalian renal peptide transporter PEPT2 in the yeast *Pichia pastoris* and applications of the yeast system for functional analysis. *Molecular Membrane Biology* 15 (2), 79-88.

73) Drew, D.E., Lerch, M., Kunji, E., Slotboom, D., Gier. J.W. (2006) Optimization of membrane protein overexpression and purification using GFP fusions, *Nature* Methods 3, 303-313

74) Preiss, T., Academic editor. (2007) Optimized Protein Extraction for Quantitative Proteomics of Yeasts. *PLoS ONE.* 2(10), e1078.

75) Drew, D.E., Von Heijne, G., Nordlund, P., de Gier, J.W. (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett*. 507 (2), 220–224.

76) Drew, D.E., Sjöstrand, D., Nilsson, J., Urbiq, T., Chin, C.N., de Gier, J.W., von Heijne, G. (2002) Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *PNAS*.99 (5), 2690–2695.

77) Jenny, R.J., Mann, K.G., Lundblad, R.L. (2003) A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa L. *Protein Expr Purif 31* (1), 1–11.

78) Waugh, S. D. (2011) An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expr Purif*. 80 (2), 283–293.

79) Akbarzadeh, A., Rezai-Sadabady, R., Davaran, S., Joo, S.W., Zarghami, N., Hanifehpour, Y., Samiei, M., Kouhi, M., Nejati-Koshki, K. (2013) Liposome: classification, preparation, and. *Nanoscale Res Lett* 2013, 8(1), 102.

80) Ricaud, J.L, Lévy, D. (2003) Reconstitution of Membrane Proteins into Liposomes, *Methods Enzymol.* 372, 65–86.

81) Cladera, J. Ricaud, J.L., Villa-Verde, J., Duñach, M. (1997)Liposome Solubilization and Membrane Protein Reconstitution Using Chaps and Chapso, *Eur. J. Biochem.* 243 (3), 798-804.

82) Bezrukov, S.M., Vodyanoy, I., (1993) Probing alamethicin channels with water-soluble polymers. Effect on conductance of channel states, *Biophy. J.* 64 (1), 16-25.

83) Kawai, S., Hashimoto, W., Murata, K. (2010) Transformation of Saccharomyces cerevisiae and other fungi. Methods and possible underlying mechanism. *Bioeng bugs.* 1(6), 395–403.

84) Potter, H., Heller, R. (2003) Transfection by Electroporation. *Curr Protoc Mol Biol.*

85) Goodsell, D. S. (2001) The Molecular Perspective: Ultraviolet Light and Pyrimidine Dimers. David S. Goodsell. *The Oncologist* 6 (3), 298-299.

86) Carmona-Gutierrez, D., Ruckenstuhl, C., Baurer, M.A., Eisenber, T., Büttner, S., Madeo, F. (2010) Cell death in yeast: growing applications of a dying buddy. *Cell Death and Differentiation* 17, 733–734.

87) Ahmad, M., Hirz, M., Pichler, H., Schwab, H. (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol.* 98 (12), 5301–5317.

88) Heap, J.T., Ehsaan, M., Cooksley, C.M., Ng, Y.K., Cartman, S.T., Winzer, K., Minton, N.P (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res*. 40(8), e59.

89) Reyrat, J.M., Pelicic, V., Gicquel, B., Rappuoli, R. (1998) Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis, *Infect Immun*. 66 (9), 4011–4017.

90) Aw, R., Polizzi, K.M. (2013) Can too many copies spoil the broth? *Microb Cell Fact*. 12, 128.

91) Vassileva, A., Chugh, D.A., Swaminathan, S., Khanna, N. (2001) Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast Pichia pastoris, *Protein Expr Purif.* 21 (1),71-80.

92). Hirayama, B.A., Wong, H.C., Smith, C.D., Hagenbuch, B.A., Hediger, M.A., Wright, E.M. (1991) Intestinal and renal Na+/glucose cotransporters share common structures, *Am J Physiol*. 261(2 Pt 1), 296-304.

93) Huang, W.C., Hsu, S.C., Huang, S.J., Chen, Y. J., Hsiao, Y.C., Zhang, W., Fidler, I.J., Hung, M.C. (2013) Exogenous expression of human SGLT1 exhibits aggregations in sodium dodecyl sulfate polyacrylamide gel electrophoresis, *Am J Transl Res* 5 (4), 441-449.

94) Crichton, P.G., Harding, M., Ruprechtt, J.J., Lee Y., Kunji, E.R. (2013) Lipid, Detergent, and Coomassie Blue G-250 Affect the Migration of Small Membrane Proteins in Blue Native Gels, *J Biol Chem.* 288(30), 22163-22173.

95) Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., Fotiadis, D., Kovar, K. (2015) Cultivation strategies to enhance productivity of Pichia pastoris: A review, *Biotechnol Adv*. 33(6 pt 2), 1177–1193.

96) Vasina, J.A., Baneyx, F. (1996) Recombinant Protein Expression at Low Temperatures under the Transcriptional Control of the Major Escherichia coli Cold Shock Promoter cspA, *Appl Environ Microbiol.* 62 (4), 1444–1447.

97) Siegel, R.S., Brierley, R.A. (1989) Meth ylotrophic Yeast Pichia pastoris Produced in High-Cell Density fermentations with High cell Yields as Vehicle for Recombinant Protein Production, *Biotechnol Bioeng*. 34 (3), 403-404.

98) Heyland, J., Fu, J., Blank, L.M., Shcmid, A. (2010) Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol Bioeng*. 107 (2), 357-368.

99) Boettner, M., Prinz, B., Holz, C., Stahl, U., Lang, C. (2002) High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*, *J.Biotechnol*. 99 (1), 51-62.

100) Wang, Y., Hu, Y., Lei, Y., Lv, Y., Wang, L., Wei, H. (2015) Survey of Intracellular Protein Extraction Methods from Pichia pastoris. *World Journal of Engineering and Technology* 3, 1-6.

101) Duquesne, K., Sturgis J.N. Membrane Protein Solubilization, heterologous Expression of Membrane Proteins. Springer protocols, Volume 601 of the series Methods in Molecular Biology, 205-217.

102) Quick, M., Wright, E.M (2002) Employing Escherichia coli to functionally express, purify, and characterize a human transporter. *PNAS* 99(13), 8597–8601.

103) Figler, R.A., Omote, H., Nakamoto, R.K., Al-Shawi, M.K. (2000)Use of Chemical Chaperones in the Yeast Saccharomyces cerevisiae to Enhance Heterologous Membrane Protein Expression: High-Yield Expression and Purification of Human P-Glycoprotein, *Arch. of Biochem Biophysics*. 376 (1), 34-46.

104) Adam, Y., Edwards, R.H., Schuldiner, S. (2008) Expression and function of the rat vesicular monoamine transporter 2, *Am. J. Physiol Cell Physiol*. 294 (4), C1004-C1011.

105) Leoir, G., Menguy, T., Corre, F., Montigny, C., Pedersen, P.A., Thinès D., le Maire, M., Falson, P. (2002) Overproduction in yeast and rapid and efficient purification of the rabbit SERCA1a $Ca2^{+}$-ATPase, *Biochim Biophys Acta*. 1560 (1–2), 67–83.

106) Sasseville, L.J., Morin, M., Coady, M.J, Blunck, R., Lapointe, J.Y. (2016) The Human Sodium-Glucose Cotransporter (hSGLT1) Is a Disulfide-Bridged Homodimer with a Re-Entrant C-Terminal Loop, *Plos One* 11 (5), e0154589.

107) Jetté, M., Vachon, V., Potier, M., Béliveau R. (1997) Radiation-inactivation analysis of the oligomeric structure of the renal sodium/d-glucose symporter. *Biochimi Biophy Acta.* 1327 (2), 242–248.

108) Stevens, B.R., Fernandez, A., Hirayama, B., Wright, E.M., Kempner, E.S. (1990) Intestinal brush border membrane Na+/glucose coransporter functions in situ as a homotetramer, *PNAS* 87 (4), 1456-1460.

109) Howell, S.C., Mittal, R., Huang, L., Travis, B., Breyer, R.M., Sanders, C.R. (2010) CHOBIMALT: A Cholesterol-Based Detergent, *Biochemistry*. 49(44), 9572–9583.

110) Kuliq, W., Tynkkynen, J., Javanainen, M., Manna, M., Roq, T., Vattulainen, I., Jungwirth, P. (2014) How well does cholesteryl hemisuccinate mimic cholesterol in saturated phospholipid bilayers, *J Mol Model*. 20 (2), 2121.

111) Chattopadhyay, A., Jafurulla, Md., Kalipatnapu, S., Pucadyil, T. J.,and Harikumar, K. G. (2005) Role of cholesterol in ligand bindingand G-protein coupling of serotonin 1A receptors solubilized from bovine hippocampus, *Biochem. Biophys. Res. Commun*. 327 (4), 1036-1041.

112) Vagenende, V., Yap, M.G., Trout, B.L. (2009) Mechanisms of protein stabilization and prevention of protein aggregation by glycerol, *Biochemistry*. 48 (46),11084-11096.

113) Murakami, C.J., Wall V., Basisty, N., Kaeberlein, M. (2011) Composition and Acidification of the Culture Medium Influences Chronological Aging Similarly in Vineyard and Laboratory Yeast, *PLoS one* 6 (9), e24530.

114) Routledge, S.J., Mikaliunaite, L., Patel, A., Clare, M., Cartwright, S.P, Bawa, Z., Wilks, M.D., Low, F., Hardy, D., Rothnie, A.J., Bill, R.M (2016) The synthesis of recombinant membrane proteins in yeast for structural studies, *Methods.* 95, 26-27.

115) Byrne, B. (2015) Pichia pastoris as an expression host for membrane prtein structural biology, *Curr. Opin. Struct. Biol*. 32, 9–17.

116) Snapp, E. (2005) Design and Use of Fluorescent Fusion Proteins in Cell Biology, *Curr. Protoc. Cell. Biol.*

117) Hefti, M.H., Van Vugt-Van der Toorn, C.J. Dixon, R., Vervoort, J. (2001) A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site. *Anal. Biochem*. 295, 180-185.

118) Jenny, R.J., Mann, K.G., Lundblad, R.L. (2003) A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. and Purif*. 31 (1), 1–11.

119) Sadilkova, L., Osicka, R., Sulc, M., Linhartova, I., Noval, P., Sebo, P. (2008) Single-step affinity purification of recombinant proteins using a self-excising module from *Neisseria meningitidis* FrpC, *Protein Sci*. 17(10), 1834–1843.

120) Van Berkel, H.C. patrick, van Veen, H.A., Geerts, E.J. Marlieke, de Boer, H.A., Nuijens, J.H. (1996) Heterogeneity in utilization of N-glycosylation sites Asn 624 and Asn 138in human lactoferrin : a study with glycosylation-site mutants, *Biochem. J*. 319 (1) 117-122.

121) Zhao, X., Li, G., Liang, S. (2013) Several Affinity Tags Commonly Used in Chromatographic Purification, *J Anal Methods Chem*, 581093.

122) Mattanoivch, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D. (2012). Recombinant Protein Production in Yeasts.*Methods, Mol. Biol.* 824, 329-358.

123) Ahmad, M., Hirz, H., Pichler, H., Schawab, H. (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production, *Appl. Microbiol. Biotechnol*. 98(12): 5301–5317.

124) Chen, G.Y., Chen., C.Y., Chang, M.D., Matsuura., Y., Hu, Y.C. (2009) Concanavalin A affinity chromatography for efficient baculovirus purification, *Biotechnol. Prog*. 25 (6):1669-77.

125) Norma J. Greenfield. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc*.; 1(6): 2876–2890.

126) Holzwarth, G., Doty, P. (1965) The Ultraviolet Circular Dichroism of Polypeptides, *JACS* 87 (2), 218-228.

127) Miles, A.J., Wallace, B.A. (2016) Circular dichroism spectroscopy of membrane proteins. *Chem.Soc.Rev*. 45, 4859

128) Turk, E., Gasymov, O.K., Lanza, S., Horwitz J., Wright, E.M. (2006) A Reinvestigation of the Secondary Structure of Functionally Active vSGLT, the Vibrio Sodium/Galactose Cotransporter. *Biochemistry* 45 (5), 1470–1479.

129) Rigaud, J.L., Pitard, B., Levy, D. (1995) Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins, *Biochmi. Biophys. Acta*. 1231 (3), 223-246.

130) Wang, L., Tonggu, L. (2015) Membrane protein reconstitution for functional and structural studies, *Sci. China. Life*. Sci. 58 (1) 66–74.

131) Rigaud, J.L., Mosser, G., Lacapere, J.J., Olofsson, A., Levy, D., Ranck, J.L (1997) Bio-Beads: An Efficient Strategy for Two-Dimensional Crystallization of Membrane Proteins. Jean-Louis Rigaud. Author links open the author workspace. *Journal of Structural Biology* (3), 226-235.

132) Paternostre, M.T., Roux, M., Rigaud, J.L. (1988) Mechanisms of Membrane Protein Insertion into Liposomes during Reconstitution Procedures Involving the Use of Detergents. 1. Solubilization

of Large Unilamellar Liposomes (Prepared by Reverse-Phase Evaporation) by Triton X-100, Octyl Glucoside, and Sodium Cholate, *Biochemistry* 27, 2668-2677.

133) Akbarzadeh, A., Rezai-Sadabady, R., Davaran, S., Joo, S.W., Zarghami, N., Hanifehpour, Y., Samiei, M., Kouhi, M., Nejati-Koshki, K. (2013) Liposome: classification, preparation, and. *Nanoscale Res Lett* 2013, 8(1), 102.

134) Parmar, M.M., Edwards, K., Madden, D.T. (1999) Incorporation of bacterial membrane proteins into liposomes: factors influencing protein reconstitution. *Biochim Biophy Acta*, 21 (1) 1421, 77–90.

135) Fang, G., Friesen, R., Lanfermeijer, F., Hagting, A., Poolman, B., Konings, W.N. (1999) Manipulation of activity and orientation of membrane-reconstituted di-tripeptide transport protein DtpT of *Lactococcus lactis*. *Mol Membr Biol*. 16(4), 297-304.

136) Rigaud, J.L., Paternostre, M.T.,Bluzat A. (1988) Mechanisms of Membrane Protein Insertion into Liposomes during Reconstitution Procedures Involving the Use of Detergents. 2. Incorporation of the Light-Driven Proton Pump Bacteriorhodopsin, *Biochemistry* 27 (8), 2677-2688.

137) le Coutre, J., Tuk, E., Kaback, R.H., Wright, E.M (2002) Ligand-Induced Differences in Secondary Structure of the Vibrio parahaemolyticus Na^+/Galactose Cotransporter, *Biochemistry* 2002, 41 (25), 8082-8086.

138) Tyagi, N.K., Kumar, A., Goyal, P., Pandey, D., Siess, W., Kinne, R.H. (2007) D-Glucose-Recognition and Phlorizin-Binding Sites in Human Sodium/D-Glucose Cotransporter 1 (hSGLT1): A Tryptophan Scanning Study, *Biochemistry*, 46, 13616-13628.

139) Baba, T., Toshima, Y., Minimaikawa, H., Hato, M., Suzuki, K., Kamo, N. (1999) Formation and characterization of planar lipid bilayer membranes from synthetic phytanyl-chained glycolipids, *Biochim. Biophys. Acta*. 1421 (1), 91–102.

140) Dave, N., Lórenz-Fonfría, V.A., Leblanc, G., Padrós, E. (2008) FTIR spectroscopy of secondary-structure reorientation of melibiose permease modulated by substrate binding, *Biophys. J*. 94 (9), 3659-70.

141) Nature (2016) 533:437 (Editorial)