



Universidad de Concepción
Facultad de Farmacia

**RECONOCIMIENTO DE INHIBIDORES DEL
CO-TRANSPORTADOR SODIO:GLUCOSA 2 HUMANO
MEDIANTE MÉTODOS DE DOCKING MOLECULAR Y
ANÁLISIS ESTADÍSTICO**

POR JOSÉ MIGUEL LEMA FIERRO

Tesis presentada en la Facultad de Farmacia de la Universidad de Concepción para
optar al título profesional de Bioquímico

Profesor Co-Guía

Dr. Felipe Zúñiga Arbalti
Laboratorio Biomarcadores
Depto. de Bioquímica Clínica e Inmunología
Facultad de Farmacia
Universidad de Concepción

Profesor Patrocinante y Guía

Dr. Alexis Marcelo Salas Burgos
Laboratorio de Diseño de Fármacos
Depto. de Farmacología
Facultad de Ciencias Biológicas
Universidad de Concepción

Enero, 2021

Concepción, Chile

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.



1. INDICE

1. Indice	I
2. Abreviaturas	V
3. Indice de Figuras	VI
4. Indice de Tablas	IX
5. Resumen	1
6. Abstract	2
7. Introducción	3
7.1. Análisis Económico de la Diabetes Mellitus.....	3
7.2. Necesidades insatisfechas en el desarrollo de fármacos antidiabéticos.....	4
7.3. Costos de desarrollo y patente	5
7.4. SGLT2 como blanco terapéutico	9
7.4.1. El Gen SC5A2 codifica a SGLT2	10
7.4.2. El Gen SC5A1 codifica a SGLT1.....	11
7.4.3. Estructura de los transportadores sodio:glucosa.....	13
7.4.4. Mecanismo de transporte alternante <i>Rocking-Bundle</i>	17
7.4.5. Moldes estructurales: el transportador vSGLT	19
7.5. Inhibición de SGLT2	24
7.5.1. Historia del desarrollo de los inhibidores de SGLT2	25
7.6. Docking y estrategias de diseño de fármacos basadas en estructura	33
7.7. Planteamiento del problema de investigación	36
8. HIPÓTESIS	38

9. OBJETIVOS	38
9.1. Objetivos generales.....	38
9.2. Objetivos específicos.....	38
10. MATERIALES Y MÉTODOS	39
10.1. Búsqueda de secuencias homólogas	39
10.1.1. Obtención de secuencias desde UniProt.....	39
10.1.2. Búsqueda de homólogos remotos en bases de datos	39
10.1.3. Filtrado de resultados de búsqueda	40
10.2. Alineamientos múltiples de secuencias	41
10.2.1. Software utilizado en alineamiento múltiple.....	41
10.2.2. Análisis de alineamientos múltiples de secuencias	43
10.3. Reconstrucción filogenética	43
10.4. Predicción de segmentos transmembrana	44
10.5. Construcción de modelos comparativos.....	44
10.5.1. Construcción del modelo	44
10.5.2. Evaluación de los modelos.....	45
10.6. Dinámica molecular	45
10.6.1. Preparación de los modelos para las simulaciones	45
10.6.2. Dinámica de minimización.....	46
10.6.3. Dinámica de equilibración	47
10.6.4. Dinámica de producción	47
10.6.5. Análisis de las dinámicas moleculares	47
10.7. Creación de biblioteca de inhibidores.....	48
10.8. Docking Molecular.....	48
10.8.1. Detección de cavidades en las estructuras de los	

modelos	48
10.8.2. Docking molecular	48
10.9. Identificación de inhibidores desde búsqueda en base de datos	49
10.9.1. Docking masivo con Dock 6	49
10.9.2. Evaluación estadística de moléculas similares y con mejores puntajes	49
11. RESULTADOS	52
11.1. Secuencias homólogas obtenidas desde SwissProt y Protein Data Bank	52
11.2. Alineamiento múltiple de secuencias y conservación	53
11.3. Predicción de segmentos transmembrana	58
11.4. Reconstrucción filogenética.....	59
11.5. Resultados de modelado comparativo.....	61
11.6. Dinámica molecular para refinamiento	65
11.7. Agrupamiento jerárquico y extracción de estructuras centroides....	68
11.8. Creación y caracterización de biblioteca de inhibidores	71
11.9. Docking masivo	75
12. DISCUSIÓN.....	87
12.1. Discusión preliminar	87
12.2. Búsqueda de homólogos en bases de datos.....	90
12.3. Alineamiento múltiple de secuencias.....	92
12.4. Filogenia	95
12.5. Modelado comparativo	97
12.6. Dinámica Molecular	98

12.7. Biblioteca de diversidad de inhibidores de hSGLT2	100
12.8. Docking masivo con Dock6 e identificación de moléculas similares	101
12.9. Evaluación estadística de observaciones de docking.....	101
13. CONCLUSIONES Y PROYECCIÓN DEL TRABAJO	104
14. BIBLIOGRAFÍA.....	107



2. ABREVIATURAS

APC: Aminoácido-Poliamina-Organocación

CANVAS: CANagliflozin cardioVascular Assessment Study

DM: Diabetes Mellitus

DPP4: Dipeptidil peptidasa 4

EMA: European Medical Association

EMPA-REG: Empagliflozin-Removing Excess Glucose

FDA: Food and Drugs Administration

GLP1R: Receptor de péptido similar a glucagón 1

GLUT: Glucose Transporter

LeuT: Leucine Transporter

SGLT: Co-Transportador Sodio-Glucosa (*Sodium-GLucose co-Transporter*)

SLC: Solute Carriers

SSF: *Sodium Symporter Family*

QSAR: Quantitative Structure-Activity Relationship

UKPDS: United Kingdom Prospective Diabetes Study



3. ÍNDICE DE FIGURAS

Figura 1: Blancos terapéuticos contra diabetes según número de patentes	8
Figura 2: Familias SLC y sus clasificaciones según clanes de Pfam	10
Figura 3: Topología de LeuT	14
Figura 4: Plegamiento LeuT	15
Figura 5: Arquitectura general de LeuT... ..	16
Figura 6: Residuos que conforman las puertas celulares en LeuT	17
Figura 7: Mecanismo de acceso alternante <i>rocking bundle</i>	18
Figura 8: Alineamiento estructural de moldes de vSGLT	20
Figura 9: Residuos de las puertas intracelular y extracelular en vSGLT	21
Figura 10: Residuos en el sitio de unión a sodio en vSGLT y LeuT	22
Figura 11: Estructura molecular de la florizina.....	25
Figura 12: O-aril glucopiranosidos inhibidores de SGLT2	32
Figura 13: C-aril glucopiranosidos inhibidores de SGLT2.....	32
Figura 14: Otros inhibidores de SGLT2 derivados de carbohidratos	32
Figura 15: Sección del alineamiento de las secuencias obtenidas con MAFFT G-INS-i y matriz SLIM.....	55
Figura 16: Sección del alineamiento de las secuencias de la familia SSS con MAFFT G-INS-i y matriz SLIM	56
Figura 17: Logo de los sitios 240-290 del alineamiento con MAFFT G-INS-i.	58
Figura 18: Esquema de los resultados de predicción de segmentos transmembrana entregados por MEMSAT-SVM	58

Figura 19: Resultados de predicción de segmentos transmembrana entregados por TOPCONS.....	58
Figura 20: Árbol filogenético realizado mediante método de máxima verosimilitud con RAxML para el alineamiento de secuencias homólogas de la familia SSS con MAFFT G-INS-i y matriz SLIM	60
Figura 21: Modelo de hSGLT2 (modelo número 3) producido con MODELLER...	63
Figura 22: Análisis de Ramachandran de modelo 3 de hSGLT2 previo a la minimización.....	64
Figura 23: Residuos que interactúan con la Galactosa en el modelo de hSGLT2...	64
Figura 24: Sistema de hSGLT2-membrana.....	65
Figura 25: Análisis de RMSD y RMSF para 50 ns dinámica molecular de hSGLT2.....	66
Figura 26: Distribución conjunta de valores de RMSD y Radio de giro durante las Dinámica Molecular de hSGLT2.....	67
Figura 27: Análisis de poro en 50 ns de Dinámica Molecular de hSGLT2.....	67
Figura 28: Determinante de poro de agua en hSGLT2.....	68
Figura 29: Agrupamiento de conformaciones de hSGLT2.....	69
Figura 30: Análisis de Ramachandran de centroide D de hSGLT2 posterior a la equilibración y dinámica.....	70
Figura 31: Distribución de puntajes de docking de inhibidores conocidos en docking con hSGLT2	71
Figura 32: Análisis de componentes principales con reducción dimensional a 2 dimensiones del espacio químico de biblioteca de inhibidores de hSGLT2 a partir de Fingerprints de Morgan	73

Figura 33: Análisis de componentes principales, con reducción dimensional a 3 dimensiones, del espacio químico de biblioteca de inhibidores de hSGLT2 a partir de Fingerprints de Morgan	73
Figura 34: Moléculas centroides resultantes del agrupamiento de Butina de la base de datos de inhibidores de hSGLT2	74
Figura 35: Distribución de puntajes de docking para base de datos STK/STL con Dock6 y modelo de hSGLT2	75
Figura 36: Distribución de puntajes de docking con modelos de hSGLT2 para moléculas con índice de similaridad de Dice superior a 0.6	76
Figura 37: Bloxpot de los conjuntos de puntajes de docking ‘Similares’, ‘STK/STL’ y ‘Conocidos’ con outliers.....	77
Figura 38: Bloxpot de los conjuntos de puntajes de docking ‘Similares’, ‘STK/STL’ y ‘Conocidos’ sin outliers.....	77
Figura 39: Distribuciones de puntajes de docking de los conjuntos de moléculas ‘Similares’, ‘Conocidos’ y ‘STK/STL’.....	78
Figura 40: Muestra de moléculas idénticas del conjunto ‘Conocidos’ en distribución de puntajes de conjunto ‘Similares’.....	78
Figura 41: Mejores 16 moléculas respecto a puntaje de docking de VitasM	79
Figura 42: Distribución, boxplot y gráfico QQ de conjunto ‘STK/STL’.....	81
Figura 43: Distribución, boxplot y gráfico QQ de conjunto ‘Similares’.....	81
Figura 44: Distribución, boxplot y gráfico QQ de conjunto ‘Conocidos’.....	81
Figura 45: Alineamiento de secuencias con POA.....	93
Figura 46: Histograma de longitud de alineamientos y porcentaje de gaps para los programas de alineamiento múltiple empleados.....	94



4. INDICE DE TABLAS

Tabla I: Residuos funcionales importantes en vSGLT, hSGLT1 y hSGLT2.....	23
Tabla II: Software utilizado en alineamiento múltiple de secuencias	41
Tabla III: Métodos de búsqueda y especificaciones para la búsqueda en la base de datos SwissProt.....	52
Tabla IV: Resultados de alineamiento y análisis de conservación.....	53
Tabla V: Sitios del alineamiento sobre 90% de conservación, con sus respectivas entropías de Shannon, residuos equivalentes en hSGLT1 y hSGLT2	57
Tabla VI: Resultados de reconstrucción filogenética con RAxML para los alineamientos	59
Tabla VII: Resultados ordenados según puntaje DOPE para modelado de hSGLT2.....	61
Tabla VIII: Interpretación de tamaño del efecto de d de Cohen	82

5. RESUMEN

El desarrollo de fármacos para el tratamiento de la Diabetes Mellitus (DM) se ha visto históricamente dificultado por la laboriosa tarea de encontrar y validar moléculas *in vitro* o *in vivo*, con procedimientos costosos en tiempo y dinero. Existe una constante demanda de nuevos fármacos para el tratamiento de la DM, debida a la progresiva pérdida de efectividad de las monoterapias y a las nuevas regulaciones respecto a efectos sobre la calidad cardiovascular. En particular, los inhibidores de los cotransportadores de sodio y glucosa (SGLTs), proteínas que acoplan el transporte azúcares al gradiente electroquímico del sodio, son considerados un novedoso blanco terapéutico para múltiples patologías, y poseen una larga historia de desarrollo, ensayo y comercialización, en especial durante la última década. Se predice que las proteínas miembros de la familia de simportadores de sodio (SSF), a los que pertenecen los SGLTs, comparten un tipo de plegamiento estructural conocido como plegamiento LeuT y el mecanismo de acceso alternante *rocking bundle*, lo que abre la posibilidad de un modelado estructural comparativo. Esta aproximación es de interés por la dificultad que presentan las proteínas de membrana para su determinación estructural. En esta tesis se plantea reconocer la capacidad de los métodos de biología computacional para identificar inhibidores de SGLTs, demostrando sus ventajas para acortar los costes de tiempo y dinero asociados al desarrollo de nuevos fármacos. Mediante la aplicación extensiva de métodos de screening virtual a partir de modelado comparativo de proteínas, dinámica molecular y docking molecular, se plantea generar una metodología universal, replicable y robusta para la exploración económica y rápida del espacio químico en búsqueda de moléculas que interaccionen con éstas y otras proteínas transportadoras transmembrana.

6. ABSTRACT

The development of drugs for Diabetes Mellitus (DM) treatment has been historically hindered by the laborious task of finding and validating *in vitro* or *in vivo* molecules, with costly procedures, both in time and money. There is a constant demand for new drugs for DM treatment, due to the progressive loss in effectiveness of monotherapies and new regulations over improved cardiovascular quality. In particular, sodium-glucose transporters (SGLTs) inhibitors, proteins that couple transport of sugars to the electrochemical gradient of sodium, are considered a novel therapeutic target for various pathologies, and have a long history of development, assay and commercialization, especially during the last decade. It is predicted that proteins members of the Sodium-solute symporter family (SSF), to which SGLTs belong, share a kind of structural folding known as LeuT folding and the alternate access mechanism *Rocking Bundle*, which opens the possibilities for a comparative structural modeling. This approximation is of interest for the difficulty that membrane proteins offers for structural determination. In this thesis we propose to recognize the capacity of computational biology methods to identify SGLTs inhibitors, showing their advantages to shorten the time and money costs associated to the development of new drugs through the extensive application of virtual screening methods from comparative modeling of proteins, molecular dynamics and molecular docking, generating a universal, replicable and robust methodology for economic and fast exploration of the chemical space in searching for molecules that interact with these and others transmembrane transporters proteins.

7. INTRODUCCIÓN

7.1. Análisis económico de la Diabetes Mellitus

Diabetes Mellitus es una condición seria de largo plazo que ocurre cuando el organismo presenta altos niveles de glucosa sanguínea debido a que no es capaz de producir insulina o es incapaz de usar efectivamente la insulina que produce. Se estima que 436 millones de adultos de edades entre 20 y 79 años sufren diabetes a nivel mundial, lo que corresponde a 9.3% de los adultos en este grupo etario. La proyección para 2030 es que 578.4 millones de personas en este rango etario padecerán diabetes, y para 2045 esta cifra se elevará hasta 700,2 millones. En Chile, en 2019 este número asciende a cerca de 1 millón 200 mil adultos, un 9.8% de la población en el rango etario de 20 a 79 años (Saeedi *et al.* 2019)

Esta alta prevalencia convierte a la diabetes en una alta carga tanto a nivel personal como de salud pública. En Estados Unidos, una persona con diabetes mellitus tipo 2 diagnosticada incurre en un gasto médico 2,3 veces más grande que el de una persona sin diabetes, sin incluir los costos indirectos por ausencias, productividad reducida y desempleo. El costo público estimado de la diabetes diagnosticada en 2017 fue de USD 327 mil millones, de los que USD 237 mil millones fueron por costos médicos directos y USD 90 mil millones en productividad reducida. Esto es un 25% del dinero asignado a cuidados de salud (American Diabetes Association, 2018). En Chile, el costo de un solo paciente diabético sin complicaciones puede ascender a cerca de 7 millones de pesos y un gasto público estimado de \$2,289,495,854 (Rojas, 2019)

7.2. Requerimientos de desarrollo de fármacos antidiabéticos

La primera línea de tratamiento de la diabetes tipo 2 es la intervención en el estilo de vida, estimulando un mejoramiento de la dieta, reducción de peso corporal e incremento del ejercicio. Si esto falla en reducir la hemoglobina glicosilada (HbA1c) en el tiempo, se inicia una monoterapia usualmente con metformina o sulfonilureas, debido a su efectividad y bajo costo.

Sin embargo, la efectividad de las monoterapias suele disminuir con el tiempo, debido al progresivo declive de la función de las células beta pancreáticas. El estudio United Kingdom Prospective Diabetes Study (UKPDS) mostró que, luego de una disminución inicial en HbA1c, tanto pacientes tratados con metformina como los que fueron tratados con sulfonilurea “experimentaron deterioro progresivo del control glicémico”, llevando a que dentro de 3-5 años de iniciado el tratamiento ~50% de los pacientes requiriesen otro fármaco para mantener una HbA1c < 7.0% (Turner *et al.* 1999, Mittermayer *et al.* 2015). Además, las causas del aumento de HbA1c se relacionan también con las características patológicas multifactoriales de la diabetes: no sólo los defectos de las células beta, sino la insensibilización de los tejidos musculares y hepáticos a la insulina, y el deterioro en adipocitos, tracto gastrointestinal, riñones y cerebro, contribuyen a la inadecuada respuesta orgánica a la insulina (DeFronzo 2009). Por esto, los pacientes diabéticos necesitan una intervención con nuevos agentes reductores de glicemia cada 3-4 años para obtener o retener un buen control, ya sea mediante biterapia, o incluso triterapia, de metformina en combinación con alguna de las siguientes opciones farmacológicas vigentes:

- 1) Sulfonilureas: Secretagogos, estimulan secreción de insulina

2) Tiazolidinedionas: Sensibilizadores de insulina en tejido adiposo

3) Inhibidores del co-transportador de sodio-glucosa 2 (SGLT2): Bloqueadores de la recaptación de glucosa filtrada en la orina, reduciendo glicemia mediante excreción urinaria de azúcares.

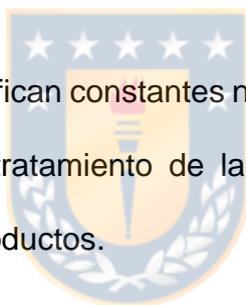
4) Agonistas del receptor de péptido similar a glucagón 1 (GLP1R): Estimulan secreción de insulina e inhiben secreción de glucagón

5) Inhibidores de la dipeptidil peptidasa 4 (DPP4): Inhiben degradación de GLP1, manteniendo su acción secretagoga de insulina.

6) Análogos sintéticos de la insulina.

Finalmente, cuando todos estos tratamientos pierden efectividad, se utiliza terapia de reemplazo de insulina.

Los factores mencionados justifican **constantes** necesidades médicas insatisfechas en el terreno farmacológico del tratamiento de la diabetes tipo 2, requiriéndose una constante oferta de nuevos productos.



7.3. Tiempo y costos en el desarrollo de fármacos

Los costos de investigación y desarrollo de un nuevo fármaco pueden llegar a ser muy altos, tanto en términos monetarios como temporales, lo que se suma a la carga económica del tratamiento médico directo de la diabetes. DiMasi *et al.* en 2016 estimaron, a partir de un estudio de 106 nuevos fármacos, que el costo total de un fármaco hasta su llegada al mercado ronda los USD 2,87 mil millones, de los que \$2,6 mil millones se gastan en desarrollo y aprobación, y \$312 millones en investigaciones post-aprobación sobre dosificación y otros factores, en un proceso que toma 10-12 años hasta el visto bueno por la Food and Drugs Administration (FDA). De los

fármacos ensayados sólo existe un 11,83% de probabilidades de éxito clínico; esto es, que el fármaco que ingresa a ensayos clínicos resulte en su eventual aprobación. DiMasi *et al.* hallaron también que los nuevos fármacos antidiabéticos resultaban estar entre los más costosos, en especial durante su Fase III de ensayos clínicos, donde alcanzaban costos sobre el 92% respecto al promedio de los otros fármacos (Viereck and Boudes 2011). Esto se debe en gran parte a la incorporación el 2008 de nuevas exigencias en las guías entregadas por la FDA para el desarrollo de esta clase de compuestos, particularmente respecto a exigencias respecto a evaluación y caracterización de sus posibles efectos cardiovasculares (CV). Los pacientes diabéticos presentan de por sí un elevado riesgo CV, y los requerimientos de neutralidad (o beneficio) cardiovascular aumentaron luego de la publicación de un meta-análisis que relacionaba a un medicamento previamente aprobado (Avandia®, Rosiglitazona) con un incremento en el riesgo de ataques cardíacos, y que causó eventualmente su retiro de los mercados (Nissen y Wolski 2007). Existen 3 inhibidores que han mostrado efectos positivos a nivel cardiovascular: estos incluyen dos inhibidores de SGLT2, empagliflozina y canagliflozina, con efectos demostrados en reducción de muerte cardiovascular (estudios EMPA-REG y CANVAS, respectivamente) y un agonista de GLP1R, liraglutida, que ha mostrado efectos en prevención de eventos cardiovasculares adversos principales (ataques cardíacos, ataque cerebrovascular y muerte cardiovascular) (Boehm *et al.* 2018). Estudios recientes han mostrado también efectos protectores sobre la función renal de los inhibidores de SGLT2 (Nespoux y Vallon 2020).

El centro Tufts para el estudio de desarrollo de drogas (*Tufts Center for the Study of Drug Development*, Tufts University), en su estudio de impacto de Septiembre/Octubre de 2016, destacó el gran riesgo de los fármacos anti-diabéticos en comparación a

drogas de otras áreas. En este estudio, se determinó que sólo 1 de 13 drogas en estudio recibieron aprobación por la FDA, pese a que 44% de las nuevas drogas endocrinas aprobadas en USA entre 1995 y 2015 fueron para diabetes y condiciones relacionadas. Los tiempos de desarrollo han aumentado entre 2002 a la fecha desde 4.7 años a 6.7 años, y el número de pacientes en pruebas aleatorias estandarizadas ha subido en un factor entre 2.5 a 4.0 veces respecto a antes del 2008. Es decir, el incremento en requisitos aumentó no solo los costos, sino también los tiempos de desarrollo.

Estas nuevas barreras de entrada han llevado a una disminución en el número total de patentes contra blancos farmacológicos de pequeñas y grandes moléculas para el tratamiento de la diabetes en el periodo de tiempo de seis años entre 2011-2016 (1149 patentes) comparado al período de 3 años entre 2008-2010 (1318 patentes) (Figura 1). Incluso, grandes jugadores en el terreno del desarrollo de fármacos anti-diabéticos, como Bristol-Myers Squibb (USA), desarrolladores de dapagliflozina (Farxiga®), han abandonado completamente el mercado de fármacos anti-diabéticos, vendiendo la mayoría de sus activos del área a AstraZeneca (UK). Pese a esto, la necesidad médica por nuevos fármacos aún existe y varias grandes compañías farmacéuticas incluyendo Merck (principalmente), Eli Lilly, Boehringer-Ingelheim y Novo Nordisk continúan activamente involucradas en el descubrimiento de nuevas drogas (Boehm *et al* 2018). De estas, sólo Boehringer-Ingelheim y Sanofi han incrementado su portafolio del área entre 2011-2016.

La investigación actual se enfoca en 15 blancos farmacológicos, que poseen sobre 20 nuevas patentes cada uno. De estos, los 5 que concentran el 30% de las patentes son, en orden, la dipeptidil peptidasa-4 (DPP-4), el receptor de péptido similar a

glucagón 1 (GLP1R), el receptor acoplado a proteína G 119 (GPR119), el receptor de insulina (INSR) y el transportador de sodio-glucosa 2 (SGLT2).

Boehringer-Ingelheim y Eli Lilly son las compañías que más invierten en desarrollo de inhibidores de SGLT2. Éstas desarrollaron una alianza estratégica en 2011 para desarrollar nuevas terapias para tratar diabetes en sobre 20 blancos farmacológicos, de los que DPP4 y SGLT2 son los principales (Boehm *et al.* 2018).

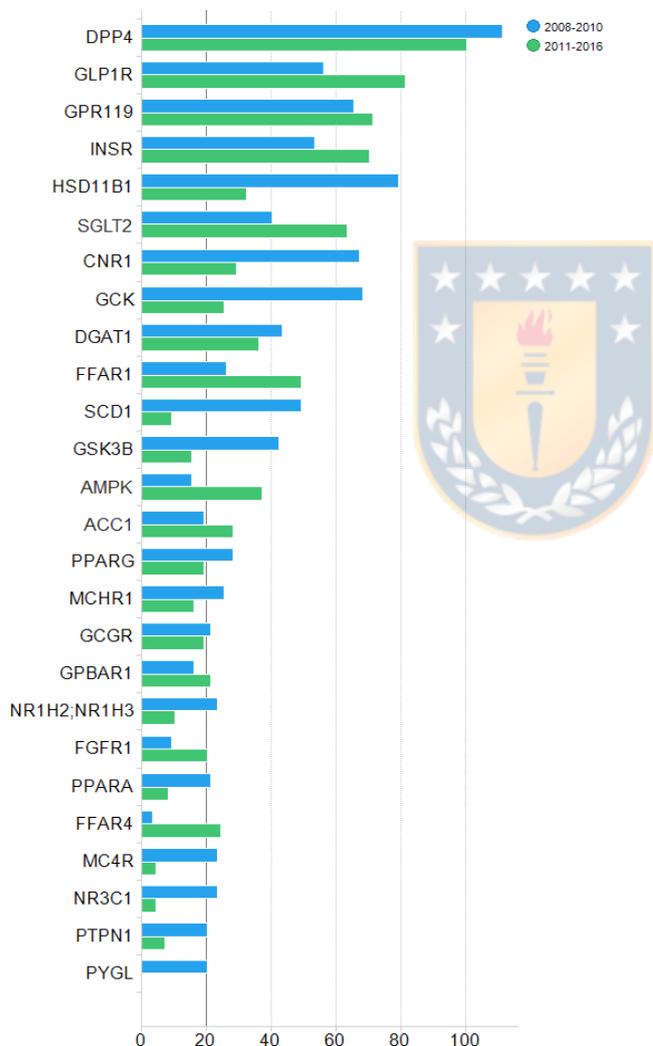


Figura 1: Blancos terapéuticos contra diabetes según número de patentes. Principales blancos terapéuticos anti-diabéticos con conteo de patentes >20 durante 2008-2010 y 2011-2016 (Boehm *et al.* 2018)

En conclusión, los inhibidores de SGLT2 se muestran como drogas novedosas y algunas demuestran efectos CV y renales favorables, pero con un desarrollo limitado por los altos costos de tiempo y dinero.

7.4. SGLT2 como blanco terapéutico

El interés en usar SGLT2 como blanco terapéutico surge como una alternativa a los mecanismos metabólicos usados por otros fármacos para el control de la glicemia. Mientras la mayoría de los fármacos anti-diabéticos utilizan blancos pertenecientes a las rutas metabólicas de la glucosa y otros azúcares, o intentan potenciar el efecto de la insulina mediante aumento de su secreción o de su sensibilización orgánica, los inhibidores de SGLT2 apuntan a la reducción directa de los niveles sanguíneos sin la utilización de la glucosa por el organismo, aumentando su excreción a través de la orina. Esto contribuye a los ya mencionados efectos benéficos a nivel cardiovascular, debido a que alivia el estrés mecánico de la hiperglicemia sobre los vasos sanguíneos y reduce la carga metabólica (Kanwal *et al.* 2013).



7.4.1. El gen SC5A2 codifica a SGLT2

El co-transportador de sodio-glucosa 2 humano es una proteína transportadora integral de membrana, que pertenece a la familia *SLC5* (*Solute Carrier 5*) de genes. Al mismo tiempo, la totalidad de los miembros de la familia *SLC5* pertenecen, en la clasificación estructural de Pfam, a la familia de simportadores de sodio:soluto SSF (*Sodium:solute symporter family*), que pertenece al clan APC (Aminoácido-Poliamina-Organocación, CL0062). (Figura 2)

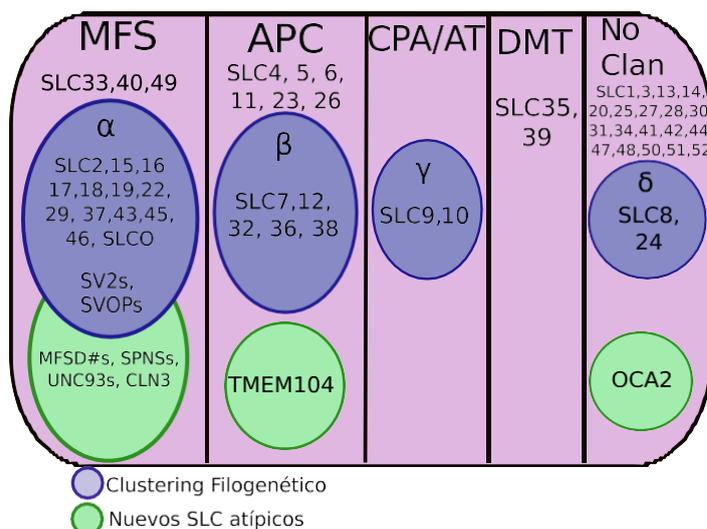


Figura 2: Familias SLC y sus clasificaciones según clanes de Pfam.

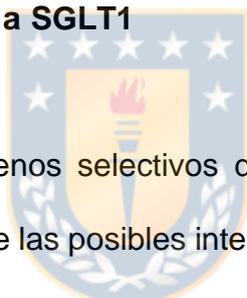
Las varias familias de genes Solute Carriers (SLC) pueden clasificarse en varios clanes estructurales catalogados por la base de datos Pfam y grupos filogenéticos. Los *SLC5*, que codifican los SGLTs, se ubican en la superfamilia APC (Amino-Poliamida-Catión)

El gen *SLC5A2* se ubica en el cromosoma 16 (16p11.2) y expresa la proteína hSGLT2 de 670 aminoácidos y 73 kDa que se encuentra principalmente en el túbulo proximal de las nefronas del riñón, con 59% de identidad y 79% de similitud con SGLT1 (Wright 2013, Wells *et al.* 1992). hSGLT2 es el responsable del 90% de la reabsorción de glucosa desde el plasma filtrado, siendo el resto reabsorbido por hSGLT1 en secciones más distales de la nefrona. Es un cotransportador de sodio:glucosa de baja afinidad ($K_M=6$ mM) y alta capacidad (saturación ≥ 35 mM), que utiliza la energía almacenada en el gradiente electroquímico del sodio, generada por transportadores primarios, para acoplar el transporte de azúcares, principalmente glucosa, con una estequiometría sodio:glucosa de 1:1, a diferencia de hSGLT1 que es de 2:1 (Wright *et al.* 2011). En el riñón funciona aproximadamente al 50% de su capacidad. A diferencia de SGLT1, SGLT2 presenta escasa afinidad por la galactosa, probablemente dado

porque las diferencias de secuencia causan diferencias sutiles en las distancias de enlaces de hidrógeno entre hidroxilos de los azúcares y los sitios de unión, resultando en diferencias de selectividad (Hummel *et al.* 2011). Posee 14 hélices transmembrana determinadas experimentalmente, aunque no se dispone de estructuras tridimensionales cristalizadas (Wright 2013).

Mutaciones en hSGLT2 causan glicosuria renal familiar (GRF), una rara enfermedad autosómica recesiva en que la glucosa es excretada por la orina en tasas >1 g/día cuando los niveles de glucosa sanguínea son normales (Wright *et al.* 2011, Wright *et al.* 2007, Santer y Calado 2010)

7.4.2. El gen *SC5A1* codifica a SGLT1



Muchos de los inhibidores menos selectivos de SGLT2 también suelen afectar a SGLT1, por lo que el estudio de las posibles interacciones con este transportador (y a veces con otros transportadores de azúcares como GLUT1) es un requerimiento en todos los ensayos clínicos. Además, SGLT1 fue tempranamente clonado y secuenciado (Hediger *et al.* 1987) y es el co-transportador más ampliamente caracterizado de la familia SSF, y desde donde se extrapolan las características estructurales y funcionales al resto de la familia. Por tanto, se describirá brevemente a SGLT1 también.

El gen *SLC5A1* se localiza en el cromosoma 22 (22q12.3) y codifica la proteína de membrana hSGLT1 de 664 aminoácidos y de 73 kDa, que se expresa de manera principal en el borde en cepillo de la mucosa del lumen del intestino delgado y en menor medida, en el túbulo proximal de las nefronas del riñón, y también en tráquea, riñón, corazón y colon (Wright *et al.* 2004, Wright *et al.* 2011). SGLT1 transporta de

forma indistinta tanto glucosa como galactosa, ambas con una K_M de 0.5 mM y un coeficiente de Hill cercano a 2 respecto a Na^+ (Wright *et al.* 2007, Wright *et al.* 2011), lo que lo vuelve un transportador de alta afinidad por glucosa, pero de baja capacidad. SGLT1 posee también 14 hélices transmembrana. Como todos los cotransportadores de sodio, comparte una región de unión a sodio, Na_2 , y un segundo sitio de unión a sodio Na_1 cercano al sitio de unión del sustrato (Wright *et al.* 2017). Se ha postulado un enlace disulfuro entre C255 y C511 en los lazos extracelulares 4 y 7 que tendrían implicancias en el cambio conformacional del transportador (Gagnon *et al.* 2006, Gagnon *et al.* 2007).

Mutaciones en el gen de hSGLT1 producen una enfermedad conocida como malabsorción de glucosa-galactosa (GGM), una enfermedad autosómica recesiva que se presenta en infantes recién nacidos como una diarrea osmótica que amenaza la vida por acumulación de azúcares sin absorber en el intestino, y que cesa tras 1 hora luego de remover la entrega oral de lactosa, glucosa y galactosa (Wright *et al.* 2002, Al-Suyufi *et al.* 2017). Por otra parte, hSGLT1 ha sido hallado en cardiomiocitos y se cree que puede cooperar en el desarrollo de cardiomiopatía diabética (Zhou *et al.* 2003).

7.4.3. Estructura común de los transportadores sodio:glucosa

La mayoría de los transportadores secundarios, incluidos los APC, poseen dos repeticiones de unidades de entre cuatro a siete segmentos transmembrana (TMs) cada una, que conforman dos dominios simétricos o pseudosimétricos (Forrest *et al.* 2011).

La superfamilia APC, específicamente, presenta un plegamiento estructural conocido como LeuT (desde *Leucine Transporter* de *Aquifex aeolicus*, un homólogo del

transportador de leucina humano), que consiste en un núcleo de 2 unidades repetidas de 5 segmentos TMs cada una, con algunos segmentos TM laterales “extra” dependiendo de la familia. (Yamashita *et al.* 2005, Krishnamurthy y Gouaux 2012, Singh *et al.* 2007, Zhou *et al.* 2009, Singh *et al.* 2008). Se ha reconocido una relación evolutiva desde el molde LeuT hacia las estructuras del resto de las familias que conforman la superfamilia APC y se considera que en general las estructuras de repeticiones invertidas se han originado por procesos de duplicación y fusión de genes (Vastermark y Saier 2014, Keller *et al.* 2014, Kinch y Grishin 2002). Se predice que hay 71 distintos SLCs humanos que poseen el plegamiento LeuT (Edwards *et al.* 2017).

Las repeticiones de 5 segmentos TM (TM1-TM5 y TM6-TM10) de LeuT se organizan respecto a un eje de pseudo-simetría casi perpendicular al plano de la membrana, con un RMSD de 5.3 Å. Además, una característica funcional clave de este plegamiento es que los segmentos TM1 y TM6, que poseen la mayor conservación de residuos, son hélices antiparalelas discontinuas a la mitad de la membrana. En las zonas de discontinuidad se exponen átomos de oxígeno y nitrógeno de la cadena principal que se utilizan para generar enlaces de hidrógeno y núcleos de coordinación de iones (Figura 3). Tanto TM3 como TM8 tienen también residuos altamente conservados, que en conjunto con TM1 y TM6 cooperan en la unión al sustrato y al ion. El resto de los segmentos TM sirven como soporte a este núcleo estructural (Yamashita *et al.* 2005).

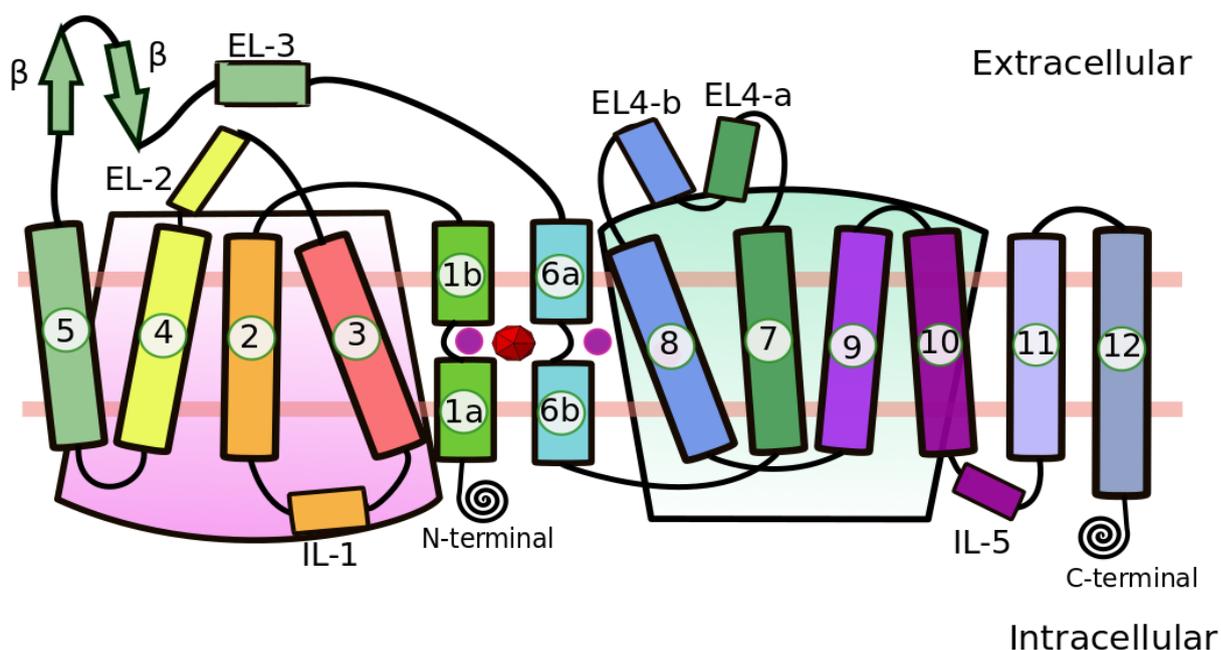


Figura 3: Topología de LeuT. Los sitios de unión de leucina y de los sodios se representan como un poliedro rojo y con dos puntos violetas, respectivamente. Las dos repeticiones invertidas de 5 segmentos transmembrana se muestran coloreadas en rosado y verde

Las dos unidades repetidas se ensamblan siguiendo patrones de plegamiento comunes: las hélices TM4 y TM5, de la primera repetición, forman una estructura en “V” en la que se inserta TM3, y del mismo modo lo hacen TM9 y TM10 con TM8, en la segunda repetición; en tanto, los segmentos TM1b y TM3 (1° repetición), y TM6a y TM10 (2° repetición), contribuyen a la formación de un vestíbulo acuoso expuesto a la superficie extracelular de la proteína a la que se unen uno o dos iones Na⁺ y el sustrato, y los TMs 1a, 5 (1° repetición), 6b, 8 (2° repetición), a la vez forman simétricamente un vestíbulo intracelular, generando la estructura que permite la

captación y liberación de los ligandos durante el proceso de transporte alternativo (Figuras 4 y 5).

Las hélices forman 3 motivos estructurales visibles: un haz o núcleo (core) de 4 hélices con las dos primeras hélices de cada repetición (TMs 1, 2, 6 y 7), un dominio con la forma de un signo “hash” (#) (TMs 3-4 y 8-9) y dos brazos auxiliares que estabilizan las hélices de ambos dominios (TMs 5 y 10) (Figura 5)

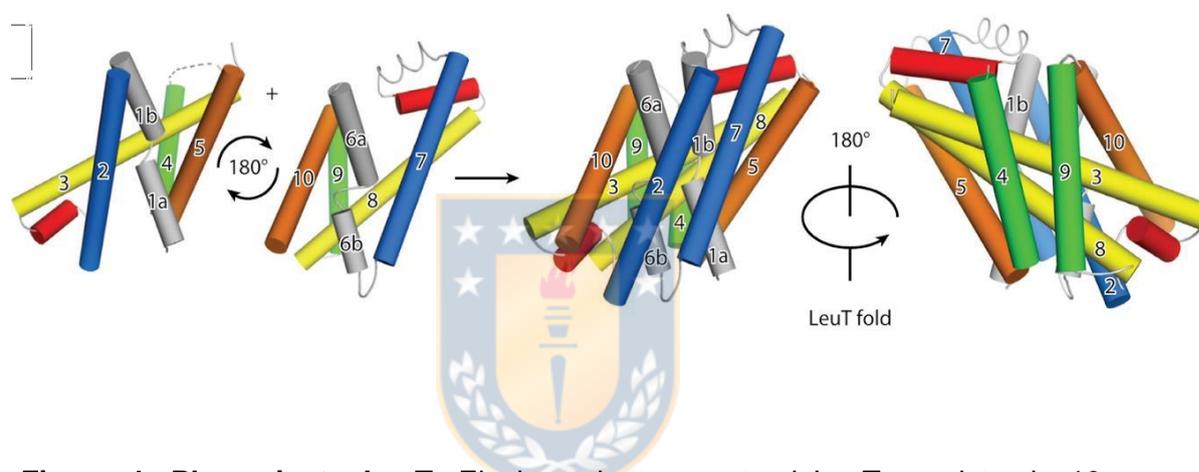


Figura 4: Plegamiento LeuT. El plegamiento estructural LeuT consiste de 10 segmentos transmembrana organizados en dos repeticiones estructurales invertidas, cada una de 5 segmentos TM. Las primeras TMs de cada repetición, TM1 y TM6 son discontinuas y consisten de dos alfa hélices cortas conectadas por un segmento desenrollado altamente conservado.

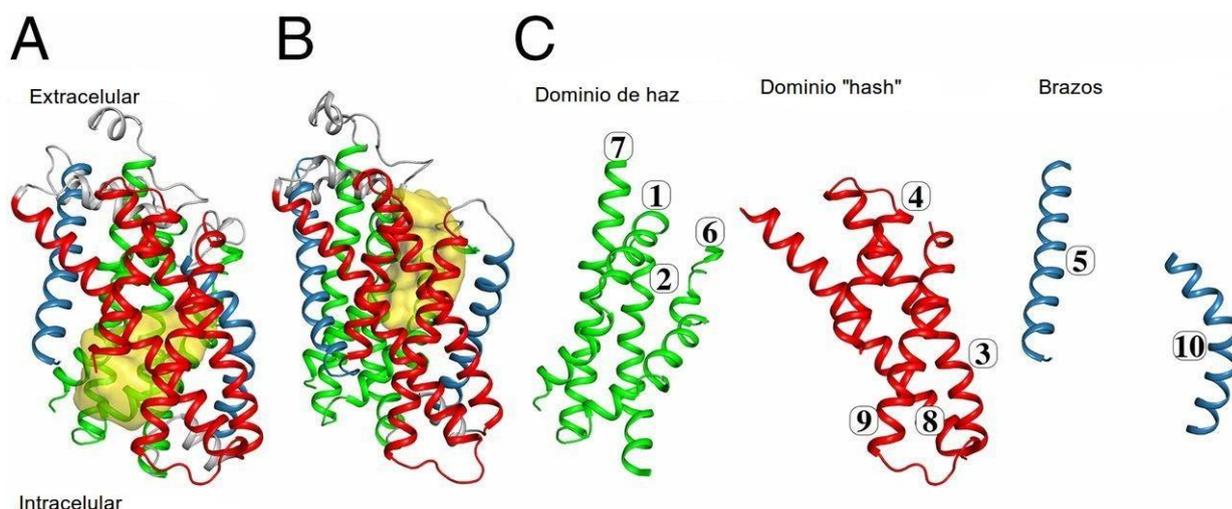


Figura 5: Arquitectura general de LeuT. (A) Representación en cinta de estructura de LeuT, en conformación endofacial, con la cavidad de unión intracelular de galactosa mostrada en amarillo. El motivo de haz se representa en verde, el motivo "hash" se representa en rojo y los brazos de sostén en azul. (B) Conformación exofacial (C) Motivos estructurales mostrados por separado (Forrest *et al.* 2011)

Este plegamiento causa que un grupo de no más de 3 residuos formen una puerta extracelular que obstruye la salida a la solución extracelular en las conformaciones ocluida y citoplasmática, que interactúan con los solutos e impiden su difusión al medio extracelular luego de unirse al sitio de unión de sustrato. Estos residuos están altamente conservados en la familia SSF (Yamashita *et al.* 2005, Jiang *et al.* 2012) (Figura 6).

La puerta intracelular es más sustancial que la extracelular, con 20 Å de longitud, y está conformada por el esqueleto de TM1a, TM6b y TM8, y un par cargado conservado R5 y E369 que forman un puente salino. Todos estos residuos están estrictamente conservados en la familia SSF (Yamashita *et al.* 2005).

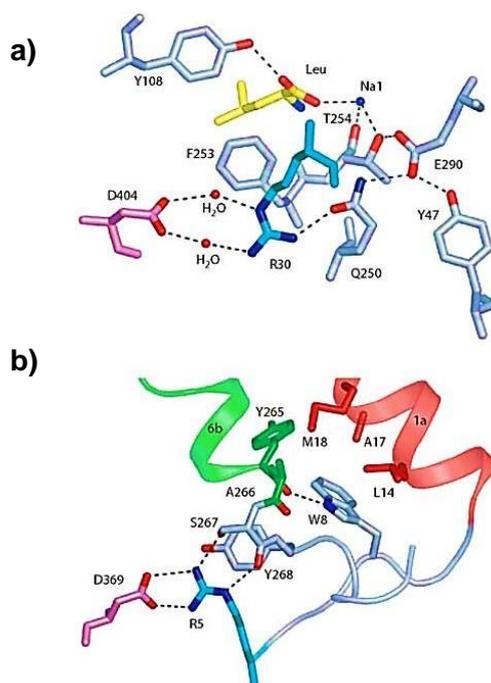


Figura 6: Residuos que interactúan en la formación de las puertas celulares. a) Residuos de la puerta extracelular b) Residuos de la puerta citoplasmática. En azul se representan los residuos de arginina, púrpura para los residuos aromáticos, rojo los de aspartato y en amarillo se representa la leucina (Yamashita *et al.* 2005)

7.4.4. Mecanismo de transporte alternativo *Rocking-Bundle*

Se postula que el transporte secundario de solutos se produce mediante un mecanismo alternante, en que el transportador alterna entre conformaciones abiertas hacia el ambiente externo y el ambiente intracelular (Mitchell 1957, Jardetzky 1966). Distintos tipos de mecanismos se han propuesto para diferentes familias (Forrest *et al.* 2011). El transporte de solutos en LeuT se produce mediante un mecanismo alternante conocido como *rocking-bundle*. En este, tanto el motivo *haz* como el motivo “*hash*” se comportan casi como estructuras rígidas, con 0.7 Å and 0.9 Å de RMSD respectivamente, que rotan una respecto a la otra, permitiendo el cierre y la apertura

de los vestíbulos intra- y extracelulares (Kazmier *et al.* 2017). La unión de los sustratos y los iones desde el ambiente extracelular en la conformación exofacial, donde la puerta extracelular está abierta, produce un cambio conformacional en que el “hash” de 4 hélices sufre una rotación en torno a un eje cercano a los sitios de unión de sustrato y de iones, formando un puente salino que cierra la puerta extracelular y lleva a una conformación ocluida en que el sustrato y los iones en sus sitios de unión se encuentran encerrados por las puertas extracelular e intracelular. Un nuevo cambio conformacional lleva a la conformación abierta hacia el interior y permite la liberación de los iones y el sustrato al citoplasma (Rudnick 2013, Forrest *et al.* 2011, Schweikhard y Ziegler 2012) (Figura 7).

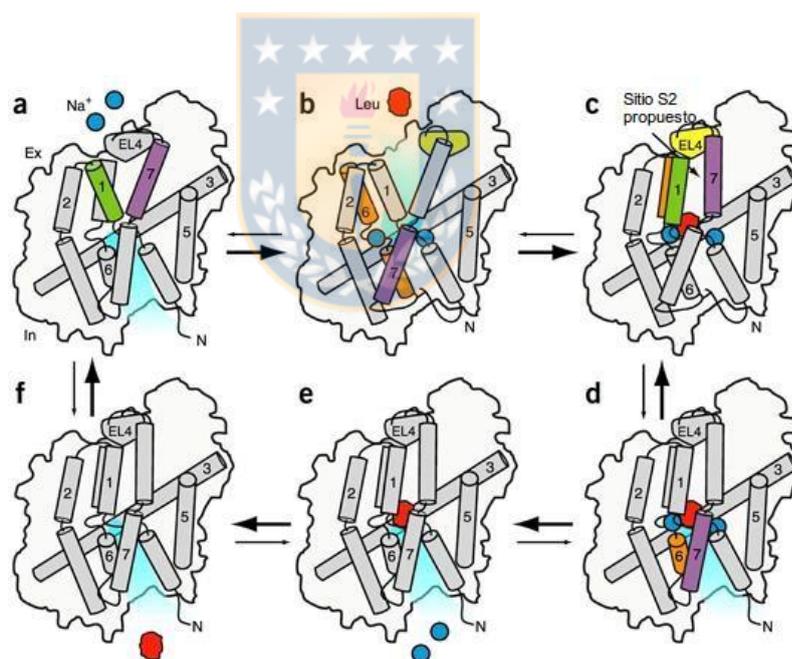


Figura 7: Mecanismo de acceso alternante *rocking bundle*.

Esquema del movimiento alternante *rocking bundle* en 6 pasos propuesto para LeuT (Kazmier *et al.* 2017)

Este mecanismo es la base funcional para el transporte de las familias de la superfamilia APC (Drew y Boukder 2016, Weyand *et al.* 2011, Adelman *et al.* 2011, Krishnamurthy *et al.* 2009, Ressler *et al.* 2009, Abramson y Wright 2009), y difiere de los mecanismos de MFS y otras superfamilias. Cabe mencionar que una de las diferencias principales en el mecanismo de transporte entre otras estructuras con plegamiento LeuT es la orientación del haz en distintas estructuras en la misma conformación, yendo desde +15° en LeuT hasta -14° en vSGLT (Forrest y Rudnick 2009).

7.4.5. Moldes estructurales: el transportador vSGLT

Las proteínas integrales de membrana presentan serios desafíos para su cristalización. Estos desafíos incluyen generar grandes cantidades de la proteína, usualmente recombinante (del orden de mg), solubilización con detergentes que conserven la estructura nativa de la proteína, estabilizaciones adicionales (anticuerpos/nanocuerpos, fusión, mutación) y cristalización como complejos proteína-detergente (Ishchenko *et al.* 2017, Bill *et al.* 2011). Por esto, no es de sorprender que la determinación de los aspectos mecánicos y estructurales del transporte mediado por SGLTs tuviese que esperar a que las técnicas de cristalización de proteínas avanzaran lo suficiente para generar modelos a resolución atómica.

El mecanismo estructural del transporte a nivel de proteína no pudo caracterizarse hasta la determinación de las estructuras de LeuT, descritas anteriormente, y del transportador de sodio:glucosa de *Vibrio parahaemolyticus*, vSGLT (Faham *et al.* 2008). Este transportador comparte un 32% de identidad y un 60% de similaridad de secuencia con hSGLT1 y posee 14 segmentos TM.

El 2008 fue esclarecida la primera estructura cristalográfica de vSGLT (PDBID: 3DH4) en conformación exofacial, con una molécula de galactosa co-cristalizada en el sitio de unión (Faham *et al.* 2008), y el 2011 una segunda estructura en conformación exofacial-ocluida (PDB:2XQ2) (Watanabe *et al.* 2011).

En la Figura 8 se muestra un alineamiento estructural de las dos estructuras de vSGLTs disponibles en PDB en diferentes conformaciones, mostrando la mínima variación conformacional entre ellas.

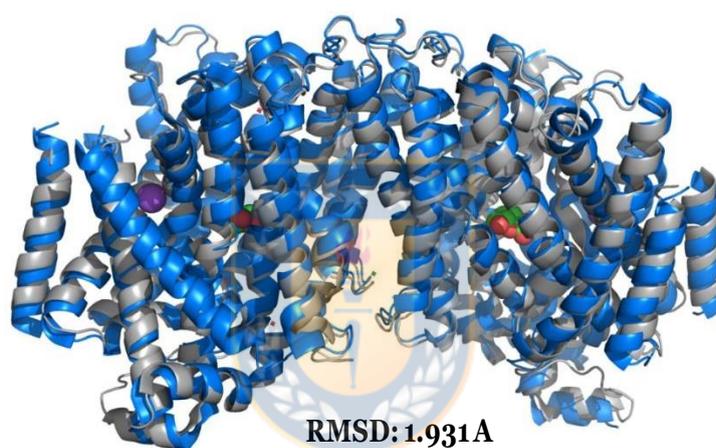


Figura 8: Alineamiento estructural de moldes de vSGLT.

En azul se muestra la estructura en conformación abierta (PDB: 3DH4) y en gris en conformación abierta-ocluida (PDB:2XQ2).

Como se observa en la Figura 9, la puerta extracelular está formada por una triada de residuos hidrofóbicos (M73, Y87 and F424), mientras que la intracelular está formada por Y262, Y263 y W264, que producen una interacción de apilamiento con la glucosa, común a todas las estructuras de unión a azúcar (p.ej. también se observa en GLUT1) (Sujatha y Balaji 2004, Abramson *et al.* 2003). En conjunto con los segmentos TM hacia el exterior, conforman una gran masa proteica de aproximadamente 20 Å de

grosor que se extiende desde el sitio de unión de ligando a la superficie extracelular, bloqueando el acceso de la región de unión a azúcar desde el medio extracelular. (Sujatha y Balaji 2004, Abramson *et al.* 2003).

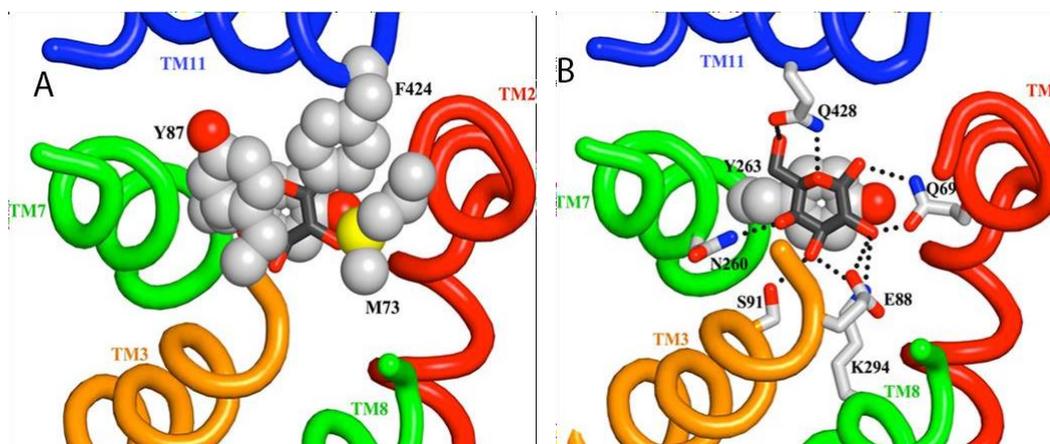


Figura 9: Residuos de las puertas intracelular y extracelular en vSGLT.

A) Residuos de la puerta intracelular (Y263) y extracelular (M73, Y87 y F424) representados como esferas. Galactosa se representa como varas. B) Misma vista, con la puerta hidrofóbica extracelular removida para mostrar el sitio de unión a galactosa. Residuos del sitio de unión se muestran como varas coloreadas por tipo de átomo. Enlaces de hidrógeno se representan como líneas negras punteadas. Notar interacciones de apilamiento entre la puerta intracelular y la molécula de galactosa. (Faham *et al.* 2008).

En el sitio de unión a Na⁺, el catión está débilmente coordinado con los oxígenos de los carbonilos de I65, A361 y el hidroxilo de la cadena lateral de S365. El oxígeno del carbonilo de A62 y del hidroxilo de cadena lateral de S364 también están en cercana proximidad (Figura 10).

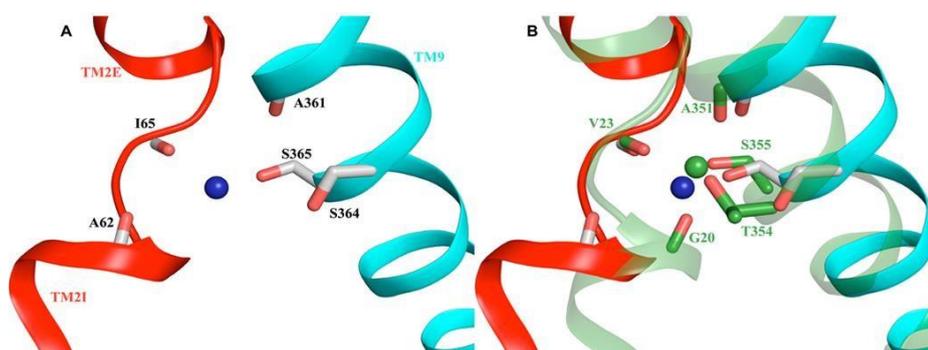


Figura 10: Residuos en el sitio de unión a sodio en vSGLT y LeuT. A) Residuos en el sitio de unión a sodio se muestran como varas coloreadas por tipo de átomo y el ion sodio como un círculo azul. B) Superposición del sitio de unión a sodio de vSGLT sobre el segundo sitio de unión a sodio de LeuT (en verde).



En la Tabla I se resumen los residuos funcionalmente importantes en el molde vSGLT, SGLT1 humano y SGLT2 humano:

Tabla I: Residuos funcionales importantes en vSGLT, hSGLT1 y hSGLT2

Función	vSGLT (soluto:sodio=2:1)	hSGLT1 (1:2)	hSGLT2 (1:1)
Puerta extracelular	M73 (TM2B)	L87(TM2B)	L84(TM2B)
	Y87 (TM3)	F101(TM3)	F98(TM3)
	F424 (TM11)	F453 (TM11)	F453 (TM11)
Puerta intracelular	Y263 (TM7E)	Y289 (TM7)	Y289 (TM7)
	Y262 (TM7E)	Y290 (TM7)	Y290 (TM7)
	W264 (TM7E)	W291 (TM7)	W291 (TM7)
Sitio de unión a sustrato	E68, Q69, (TM2), Y87, S91(TM3), A259, N260 (TM7), Q428(TM10)	H83 y N78 (TM2), E102 (TM3), Y290, W291 (TM7),K321 (TM8) y Q457 (TM11)	H80 y N75 (TM2), E99 (TM3), Y290, W291 (TM7), K321 (TM8) y Q457 (TM11)
Sitio de unión a sodio (s)	Na2: A62, I65, A361, S364 y S365	Na2: A76, I79, S389, S392 y S393	Na2: A73, I76, S389, S392, S393
Otros	D189 (Interacción con sodio al salir)	D204	D201

7.5. Inhibición de SGLT2

Siendo SGLT2 el principal responsable de la reabsorción de glucosa desde el filtrado glomerular, el objetivo de la inhibición de sus mecanismos de transporte es reducir los niveles de glucosa sanguínea a través de su excreción directa por la orina. Sin embargo, como se dijo, desarrollar nuevos fármacos anti-diabéticos implica un alto coste, superior al promedio de los fármacos dedicados a otras áreas. Esta dificultad es especialmente notoria cuando se toma en cuenta la prolongada historia del desarrollo de los inhibidores de SGLT2. A continuación, detallamos esta historia, explicando paso a paso muchas de las características que han llevado al estado del arte actual.



7.5.1. Historia del desarrollo de los inhibidores de SGLT2

El desarrollo de los inhibidores del co-transportador de sodio-glucosa 2 tiene una historia que se remonta hasta la primera mitad del siglo XIX, si bien los esfuerzos farmacológicos para desarrollarlos adquirieron mayor alcance a principios de los 2000s hasta la fecha.

- **1835: Descubrimiento de la florizina:** Christian Petersen, químico alemán, aisló desde la corteza de raíz de manzana (*Malus*) por primera vez el que se convertiría en el inhibidor clásico de SGLT2: **florizina** (Petersen 1835). La florizina es un flavonoide bicíclico; específicamente es un glucósido de floretina, la cual es una dihidrochalcona compuesta de dos anillos aromáticos unidos por un espaciador alquilo de 3 carbonos. El grupo glucósido y la floretina están

conectados por un átomo de oxígeno (Figura 11). El grupo glucósido le permite a la florizina interactuar con el sitio de unión de SGLT2 y ejercer una acción de inhibición competitiva.

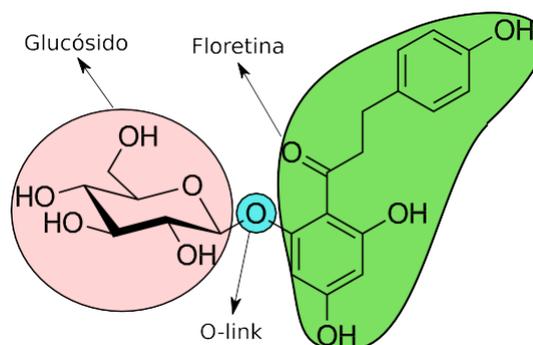


Figura 11: Estructura molecular de la florizina

Florizina, así como su aglicona, floretina, pasan por una tautomerización cetoenólica, en la que la forma ceto es la configuración con mayor afinidad por SGLT1 (Wright *et al.* 2011, Fuhrmann *et al.* 1992, Wielert-Badt 2000). Florizina es un inhibidor no específico de SGLT1 y SGLT2, con una K_i de 200-300 y 10-39 nM para estos respectivamente (Wright *et al.* 2011, Pajor *et al.* 2008, Bezanilla *et al.* 1991).

- **1886: Descubrimiento de efectos glucosúricos:** Usando la receta de Petersen, Josef von Mering (von Mering 1886) descubre el efecto glucosúrico de la florizina al administrar sobre 1.0 g de esta en perros. Un año después reporta que en personas con diabetes, 15-20 g de florizina inducen una glicosuria de 6-8%, y establece su acción sobre los riñones. Siendo que en ese momento se pensaba que la diabetes era una afección renal, von Mering nota la diferencia entre la diabetes producida por florizina y la diabetes patológica

en que la última se presentaba con hiperglicemia, mientras que en la primera la glicemia disminuía.

- **1903: Estudio en modelos murinos:** Stiles estudia el efecto renal de la administración de florizina en ratones. (Stiles 1903)
- **1933: Estudio en humanos:** Chasis estudia el efecto de la florizina en humanos y concluye que la dosis intravenosa mínima para producir un efecto transiente positivo es de 10-20 mg (Chasis 1933)
- **1961: Proposición de mecanismo de transporte secundario:** Crane propone por primera vez el mecanismo de transporte secundario sodio:glucosa para explicar el ingreso de glucosa a la célula intestinal desde el lumen de intestino delgado durante el Simposio sobre Transporte y Metabolismo de Membranas en Praga (Crane *et al.* 1961). A partir de esta propuesta, durante los 1960s y 1970s, la hipótesis fue ampliamente comprobada y extendida a otros sistemas con otras moléculas e iones, logrando identificar otros cotransportadores de sodio:soluto, que pasarían a conformar la familia SSF (Wright *et al.* 2011, Schultz y Curran 1970)
- **1962: Florizina como inhibidor competitivo:** Alvarado y Crane determinan la acción de florizina como inhibidor competitivo de la reabsorción intestinal de azúcares mediante estudios de cinética de transportadores.
- **1973: Identificación de membrana en cepillo de riñón como sitio de inhibición:** Vick, Deidrich y Baumann identifican la membrana en cepillo de los túbulos proximales del riñón como la zona donde se produce la inhibición por florizina al aislar distintos tejidos renales y estudiarlos *in vitro*. (Vick *et al.* 1973)
- **1987 y 1990s: Clonación de SGLT1 intestinal:** Hediger *et al.* logran clonar SGLT1 e inician una serie de investigaciones, en conjunto con Wright, para

caracterizar su fisiología y estructura putativa durante los 1980s y 1990s (Hediger *et al.* 1987a, Hediger *et al.* 1987b, Hediger *et al.* 1989, Hediger *et al.* 1991, Hirayama y Wright 1992, Lostao *et al.* 1994, Turk *et al.* 1994, Turk y Wright 1997, Eskandari *et al.* 1997, Eskandari *et al.* 1998). En 1992, se realiza la clonación de SGLT2 (Wells *et al.* 1992)

- **1999: T-1095, primera gliflozina:** El link de oxígeno entre el grupo glucósido y la floretina de la florizina, vuelve a esta vulnerable a las beta-lactamasas intestinales, y la floretina liberada posee acción inhibitoria sobre otros receptores, en particular sobre el ubicuo GLUT1, lo que causa complicaciones en intestino delgado, túbulo renales, cerebro y alvéolos pulmonares, resultando tóxica al consumo, además de demostrar baja biodisponibilidad y pobre absorción intestinal (Kanwal *et al.* 2013, Chao y Henry 2010, Ehrenkranz *et al.* 2005). Por esto, se han desarrollado nuevos fármacos O-, N- y S-glucósidos en base a florizina, fármacos conocidos en conjunto como **gliflozinas**. Si bien se ha planteado el uso de estos fármacos para el tratamiento de la diabetes mediante acción sobre hSGLT1, se ha demostrado que se producen efectos indeseados como inhibición de la absorción de glucosa intestinal y liberación de péptido GLP-1 (Song *et al.* 2016), por lo que los estudios de inhibición de SGLTs se han centrado en el tratamiento de la DMT2 por inhibición de hSGLT2 renal. El primer inhibidor efectivo reportado fue **T-1095** (Figura 12), sintetizado en 1999 por Mitsubishi Tanabe Pharma (Japón). Esta estructura es 4 veces más selectiva para SGLT2 y al ser metabolizado en el intestino, toma su forma activa T-1095A, la que es absorbida (Oku *et al.* 2000). Sin embargo, esta forma es aún muy poco selectiva y no prosiguió en estudios clínicos.

- **2000: Método para preparación de C-glucósidos:** Link y Sorensen publican su método para la preparación de análogos C-glicósidos de florizina usando el método de acople Suzuki de Johnson (Link y Sorensen 2000)
- **2007 y 2008: Síntesis de sergliflozina y remogliflozina:** Kissei Pharmaceutical Company (Japón) desarrollan **sergliflozina** (y su forma activa, sergliflozina A) y **remogliflozina** etabonato (prodroga de remogliflozina), que poseen mayor selectividad para SGLT2 (210:1) y mayor afinidad (Ki 2-10 nM), pero de estos sólo sergliflozina progresó a ensayos clínicos (Katsuno *et al.* 2007, Fujimori *et al.* 2008). Sin embargo, por razones no bien conocidas ninguno de los dos pasó de la fase II, posiblemente porque al ser ambos O-glucósidos son susceptibles a degradación por enzimas beta-glucosidasas (Kanwal *et al.* 2013, Pajor *et al.* 2008).
- **2008: Desarrollo de Dapagliflozina, primer C-glucósido: Dapagliflozina** (Figura 13), desarrollado en 2008 por AstraZeneca (UK) y Bristol-Myers Squibb Co. (USA), es un C-aril glucopiranosido, por lo que no sufre de la degradación de los O-glucósidos. Esta modificación aumenta la selectividad por SGLT2 respecto a SGLT1 en un factor de 1200 (Wright *et al.* 2011, Calado 2009, Meng *et al.* 2008).
- **2008: Primera estructura de vSGLT2:** Se publica la primera estructura de vSGLT cristalizada (Faham *et al.* 2008), con código PDBID: 2XQ2, dando los primeros indicios acerca de los mecanismos de transporte e inhibición de los SGLTs.
- **2010: Se sintetiza Canagliflozina:** Desarrollado por Mitsubshi Tanaba Company (Japón) y Johnson & Johnson (USA), **canagliflozina** es otro C-

glucósido que ha demostrado una reducción en niveles de HbA_{1c} desde 0.77% a 1.16% (Nomura *et al.* 2010)

- **2010: Segunda estructura de vSGLT:** Watanabe *et al.* publican la segunda estructura disponible en la actualidad de vSGLT, PDBID: 3DH4.
- **2012: Desarrollo de Ipragliflozina y Empagliflozina:** Otras modificaciones del grupo C-aril de la glucopiranosas llevaron a la síntesis de inhibidores de SGLT2 como *ipragliflozina*, desarrollado por Astellas Pharma y Kotobuki Pharmaceutical (Imamura *et al.* 2012) y *empagliflozina* (Jardiance), desarrollada por Boehringer Ingelheim y Eli Lilly Company (Grempler *et al.* 2012).
- **2012: Aprobación comercial de dapagliflozina en la Unión Europea:** La EMA (European Medical Association) aprueba la comercialización de dapagliflozina bajo el nombre comercial Forxiga® (Prasanna-Kumar *et al.* 2017)
- **2013: Aprobación comercial de canagliflozina en E.E.U.U.:** La FDA (Food and Drugs Administration) de EEUU aprueba por primera vez un inhibidor de SGLT2, entrando canagliflozina al mercado con el nombre comercial de Invokana® (Prasanna-Kumar *et al.* 2017)
- **2014: El año de las gliflozinas:** Este año se aprueban para comercialización 5 gliflozinas: Dapagliflozina es aprobada por la FDA en EEUU como Farxiga®; Empagliflozina es aprobada por la FDA y EMA como Jardiance®; Ipragliflozina es aprobada en Japón como Suglat®; Tofogliflozina (CSG452), una nueva clase de inhibidores de SGLT2 C-arilglucósido O-espirocetal, desarrollado por Kowa y Sanofi en colaboración con Chugai Pharma, es aprobada en Japón como Deberza® y actualmente está bajo fase III de

ensayos clínicos (Ohtake *et al.*, 2012); y Luseogliflozina (TS-071), un tioglicósido que inhibe SGLT2 y SGLT1, es aprobada en Japón como Lusefi® (Mishra *et al.* 2017, Prasanna-Kumar *et al.* 2017, Nisly *et al.* 2012, Takasu *et al.* 2015, Poole y Dongo 2014) y está actualmente bajo ensayos clínicos de fase III por Taisho Pharmaceuticals (Kakinuma *et al.*, 2010)

- **2015: Estudio EMPA-REG:** Se publican los resultados de los ensayos clínicos EMPA-REG realizado entre 2010 y 2013 a más de 7 mil pacientes con DM2 que habían padecido eventos cardiovasculares previos en Europa, Norteamérica y Asia (<https://clinicaltrials.gov/ct2/show/NCT01131676>). Se concluye que “pacientes que recibieron empagliflozina, comparado con quienes recibieron placebo, tienen una menor tasa de resultado cardiovascular primario compuesto y de muerte por cualquier causa cuando la droga de estudio fue añadida al cuidado estándar” (Zinman *et al.* 2015)
- **2017: Primera biterapia gliflozina/iDDP4:** Se aprueba en la Unión Europea la primera combinación de dosis fija de Saxagliptina/Dapagliflozina, bajo el nombre de Qtern®. Durante la fase clínica III saxagliptina + dapagliflozina + metformina demostró ser más efectiva en proveer que control glicémico que saxagliptina + metformina o dapagliflozina + metformina. La combinación se asocia a reducción del peso corporal y menor riesgo de hipoglicemia (Garnock-Jones 2017)
- **2018: Estudio CANVAS:** CANagliflozin cardioVascular Assessment Study (CANVAS) (<https://clinicaltrials.gov/ct2/show/results/NCT01032629>) y CANVAS-R (<https://clinicaltrials.gov/ct2/show/NCT01989754>), en 3486 pacientes sin eventos cardiovasculares previos y 6656 pacientes con eventos cardiovasculares previos, demostró evidencia de reducción de resultados

cardiovasculares y renales en ambos grupos; sin embargo, también mostró un ligero aumento en posibilidad de amputaciones en extremidades inferiores (Mahaffey *et al.* 2018)

- **2018: Aprobación de Ertugliflozina:** La FDA aprueba la comercialización de ertugliflozina, desarrollado por Pfizer en 2013, que pertenece a una nueva clase de inhibidores de SGLT2 (Mascitti *et al.* 2011) bajo el nombre comercial Stelagro® (Kovacich y Chavez, 2018)
- **Actualidad:** Sotagliflozina (LX4211) es un inhibidor dual SGLT2/1, xilopiranósido desarrollado por Lexicon Pharmaceuticals, que se halla actualmente bajo fase III de ensayos clínicos, pero ha sido aprobado por la EMA en la Unión Europea como Zynquista® para algunos pacientes con diabetes mellitus tipo 1 en combinación con insulina (Cefalo *et al.* 2019, Goodwin *et al.*, 2009). EGT0001442 (Bexagliflozina), desarrollada por Theracos, está actualmente bajo la fase II de ensayos clínicos y se ha encontrado que reduce significativamente los niveles de HbA1c glicosilada y mejora los niveles de glicemia (Xu *et al.* 2014, Zhang *et al.*, 2011, p. 284; Cai *et al.*, 2011 patent).

Un panorama general de las clasificaciones de los inhibidores de hSGLT2 se presenta en las Figuras 12, 13 y 14. Una revisión de los descubrimientos más recientes puede encontrarse en Palasz *et al.* 2019.

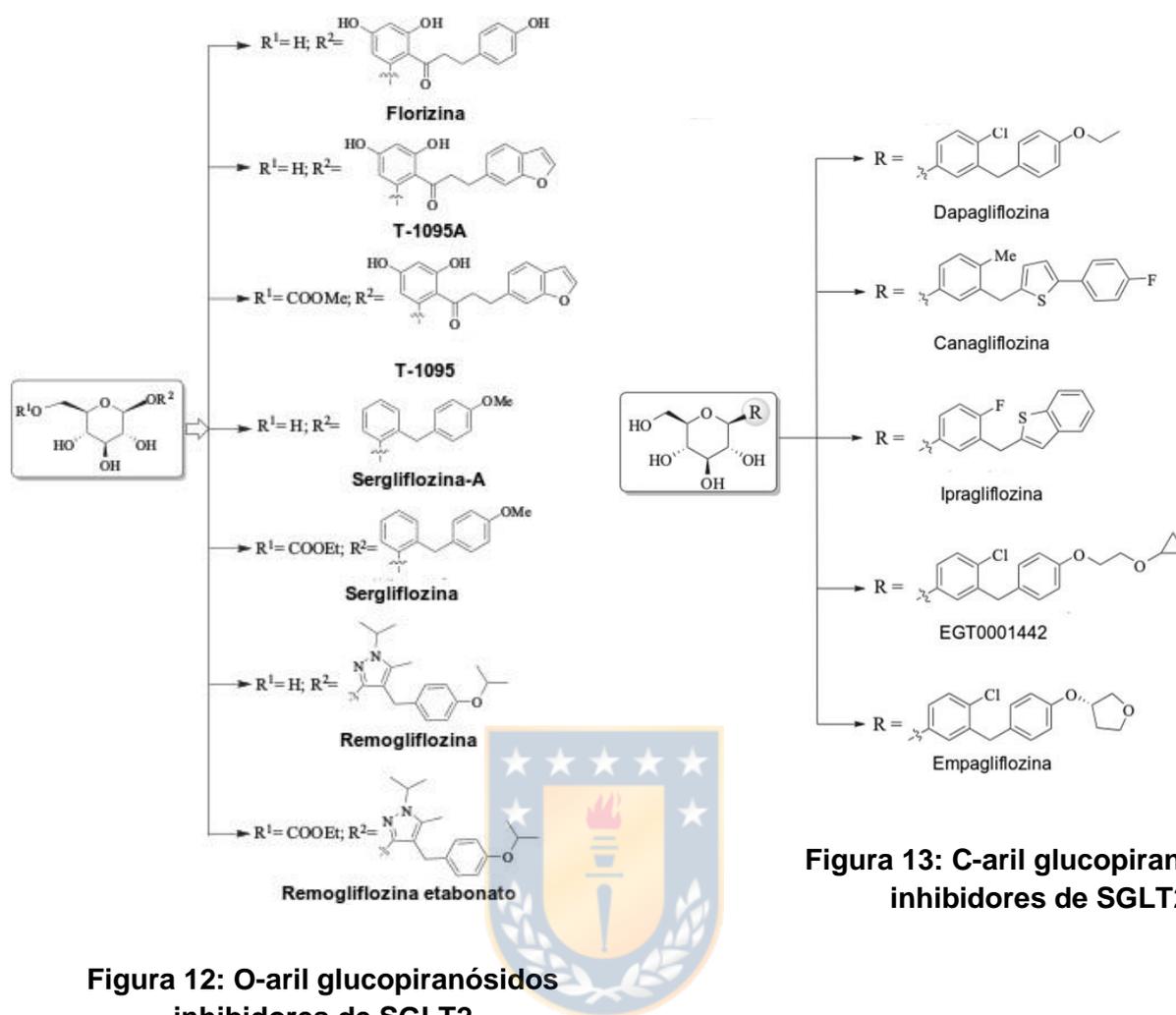


Figura 12: O-aryl glucopiranosidos inhibidores de SGLT2

Figura 13: C-aryl glucopiranosidos inhibidores de SGLT2

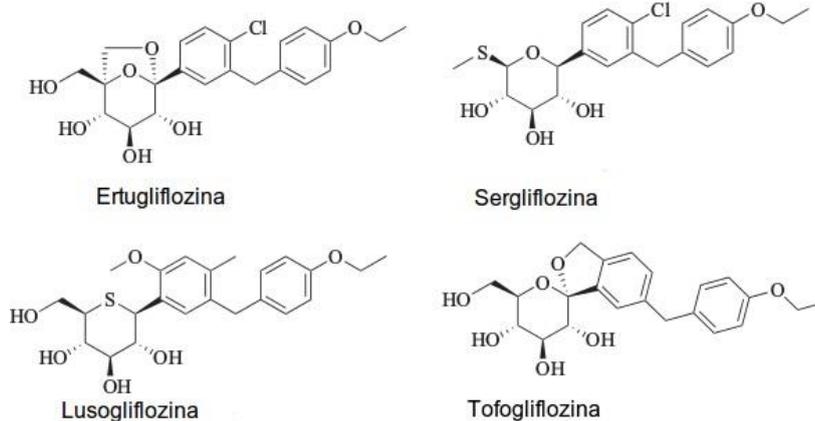


Figura 14: Otros inhibidores de SGLT2 derivados de carbohidratos

A diferencia de otros agentes antidiabéticos, los inhibidores de SGLT tienen bajo riesgo de hipoglicemia ya que no interfieren con los mecanismos de contraregulación normales para la glucosa (Kanwal *et al.* 2013). El desarrollo de estos fármacos ha crecido de modo exponencial y muchas otras modificaciones a grupos funcionales, afectando selectividad y acción, han sido propuestas y desarrolladas (Mishra *et al.* 2017, Choi 2016). Además, inhibidores no glucósidos también han sido descubiertos (Li *et al.* 2011, Du *et al.* 2011, Wu *et al.* 2010).

Más aún, el posible uso de inhibidores de SGLT2, tales como empagliflozina y canagliflozina en los ensayos EMPA-REG OUTCOME (Zinman *et al.* 2015) y CANVAS, respectivamente (Neal *et al.* 2017), han abierto la posibilidad del uso de estos fármacos en la prevención de enfermedades cardiovasculares (Rastogi y Bhansali 2017).

7.6. Docking y estrategias de diseño de fármacos basadas en estructura

Los métodos de diseño de fármacos *in silico* que utilizan estructuras tridimensionales (esto es, el uso de información estructural tridimensional obtenida desde objetivos biológicos) han sido incorporados de forma exitosa desde hace más de dos décadas en el proceso de desarrollo de fármacos, en combinación con datos farmacodinámicos (p. ej.: potencia, afinidad, eficacia, selectividad) y propiedades farmacocinéticas (ADMET: absorción, distribución, metabolismo, excreción y toxicidad) (Ferreira *et al.* 2015). De estos, el docking molecular, el screening virtual basado en estructura y las dinámicas moleculares son los métodos más prominentes debido a su amplio rango de aplicaciones en el análisis de reconocimiento molecular tales como energías de unión, interacciones moleculares y cambios conformacionales inducidos. Otros

métodos en desarrollo de fármacos se conocen como “basados en ligandos” y usan moléculas bioactivas conocidas en vez de una estructura tridimensional (p. ej.: Modelos de farmacóforos).

El docking molecular consiste en el modelado de la interacción entre una proteína (o cualquier macromolécula biológica) y una molécula pequeña en un sitio de unión previamente establecido, lo que permite caracterizar el comportamiento de las moléculas pequeñas en el sitio de unión así como dilucidar procesos bioquímicos fundamentales. El proceso del docking involucra dos pasos básicos: (1) Predicción de la conformación del ligando, así como su posición y orientación dentro de estos sitios (llamados usualmente “poses”) a través de métodos de muestreo y (2) evaluación de la afinidad de unión mediante esquemas de puntuación. La forma estándar de uso de docking, a causa de las limitaciones en recursos computacionales, se acomoda al “modelo de llave-cerradura”, usando un receptor inmóvil con una molécula móvil, aunque también se utilizan frecuentemente métodos de docking flexible en que a los residuos del sitio de unión se les permite un grado de movilidad. El proceso de docking se realiza de forma recursiva hasta converger a una solución de mínima energía (Meng *et al.* 2011). Es necesario comprender las limitaciones del docking como método: muchas aproximaciones de docking no toman en cuenta la solvatación de las moléculas ni el pH y sus contribuciones energéticas a la unión, lo que puede provocar que resultados no sean comparables con los obtenidos por dinámica molecular o métodos basados en ligandos (Chen 2015). Sin embargo, su utilidad en el descubrimiento de fármacos ha sido ampliamente demostrada y forma parte del procedimiento estándar de exploración inicial.

El screening virtual trata del uso de métodos *in silico*, ya sean basados en estructura o basados en ligando, para seleccionar compuestos promisorios desde bases de

datos de compuestos químicos. Puede usarse docking molecular para realizar un barrido extensivo sobre la base de datos y seleccionar los compuestos con mejores puntajes como posibles candidatos (Meng *et al.* 2011).

En tanto, las simulaciones de dinámica molecular abarcan un conjunto de metodologías y algoritmos para simular los movimientos físicos de un sistema biológico molecular. Por ejemplo, es posible crear un sistema proteína-membrana en un medio acuoso a fin de estudiar la dinámica conformacional de la proteína. Las simulaciones de dinámica molecular se basan en métodos de integración que resuelven ecuaciones de movimiento para las moléculas del sistema a nivel atómico, en una serie de pasos de tiempo (generalmente picosegundos), los que conforman una “trayectoria”, y donde las fuerzas entre partículas y sus energías potenciales son calculadas usando potenciales interatómicos o campos de fuerza de mecánica molecular. Otros métodos usan muestreo de Monte Carlo para formar un colectivo de configuraciones del sistema, cuyas frecuencias determinan sus propiedades termodinámicas, sin la formación de una trayectoria. La solvatación del sistema puede ser de tipo explícita (cada molécula de agua expresada a nivel atómico) o implícita (como un medio continuo de interacciones polares). Esto permite estudiar las interacciones y conformaciones de múltiples moléculas en un ambiente que simula las condiciones biofísicas del mundo real (Tuckerman 2010). Aunque su gasto computacional es una limitante en el uso de las dinámicas moleculares en el descubrimiento de fármacos mediante screening virtual, se utiliza en el refinamiento de estructuras y cálculo preciso de energías de interacción que no son posibles con docking molecular (Zhao y Caflish 2015)

7.7. Planteamiento del problema de investigación

Como hemos visto en todo lo expuesto anteriormente, la historia de los inhibidores de SGLT2 ha sido larga, compleja y monetariamente costosa. El desarrollo de estos fármacos se ha logrado muchas veces mediante el ensayo y error, tanto en la síntesis como en las pruebas clínicas. En la última década, el uso de herramientas computacionales ha ayudado en este proceso de forma incremental. Por ejemplo, el descubrimiento de tofogliflozina por Ohtake *et al.* fue logrado mediante modelado computacional basado en modelos farmacóforos derivados de inhibidores previamente descubiertos (Ohtake *et al.* 2012). Asimismo, el descubrimiento de dapagliflozina fue ayudado en parte por métodos QSAR sobre modificaciones a estructuras reconocidas (Meng *et al.* 2008). Canagliflozina, en cambio, fue descubierta sólo mediante ensayos *in vivo* de modificaciones orgánicas realizadas "en el papel" (Nomura *et al.* 2010), al igual que ipragliflozina (Imamura *et al.* 2012). A pesar de estas técnicas adicionales, durante el 2018 y 2019 los intentos de síntesis novedosas ha ido disminuyendo, reemplazadas por el descubrimiento de fuentes naturales (Palasz *et al.* 2019).

La pregunta que nos plantearemos en esta tesis se refiere a la capacidad de los métodos exploratorios *in silico* de reconocer conclusivamente los inhibidores ya conocidos, y sus símiles moleculares, discriminando de manera significativa entre éstos y las moléculas no relacionadas. Esto se vincula a la capacidad de estos métodos de facilitar y disminuir los costos de descubrimiento, desarrollo y producción de inhibidores de hSGLT2, con proyección de uso a otras proteínas transportadoras.

Bajo la perspectiva expuesta de desarrollo histórico y retorno a la exploración de fuentes naturales, nos propusimos la tarea de cuantificar la eficiencia de los métodos de diseño de fármacos *in silico* basados en bioestructuras, nos planteamos las siguientes interrogantes:

Basados en una bioestructura refinada por dinámica molecular del transportador SGLT2 y utilizando docking molecular, ¿cuántos inhibidores ya descritos logramos identificar?



8. HIPÓTESIS

La implementación de metodologías de biología computacional para la selección y evaluación de moléculas utilizando modelos comparativos del co-transportador SGLT2 humano nos permite el reconocimiento y discriminación de un número significativo de moléculas con afinidad para este transportador.

9. OBJETIVOS

9.1. Objetivos generales

Identificar análogos de inhibidores de hSGLT2 mediante dinámica molecular, docking molecular y evaluación estadística-quimiométrica.

9.2. Objetivos específicos

- Generar y evaluar modelos tridimensionales funcionales de hSGLT2.
- Caracterizar análogos de inhibidores de hSGLT2 mediante docking molecular y análisis estadístico-quimiométrico.
- Discriminar significativamente entre análogos de inhibidores de hSGLT2 y moléculas sin potencial de interacción.

10. MATERIALES Y MÉTODOS

10.1. Búsqueda de secuencias homólogas

10.1.1. Obtención de secuencias desde UniProt

Se desarrolló un intérprete en el lenguaje de programación Python para la obtención de las 12 secuencias de aminoácidos de la familia *SLC5* desde la base de datos de secuencias SwissProt haciendo uso del programa Biopython (Cock *et al.* 2009).

10.1.2. Búsqueda de homólogos remotos en bases de datos

Se realizó una búsqueda de homólogos remotos en la base de datos de secuencias anotadas SwissProt para cada una de las secuencias de la familia, haciendo uso de los siguientes programas de búsqueda en bases de datos, utilizando distintos parámetros de E-value, penalidades de gap y matrices de sustitución:

- BLAST (BLOSUM62, -11/-1, E-value=0.001 (Altschul *et al.* 1990)
- PSI-BLAST (BLOSUM62, -11/-1, E-value=0.01) (Altschul *et al.* 1997)
- PSI-BLAST (BLOSUM62, -11/-1, E-value=0.001)
- PSI-BLAST (BLOSUM45, -14/-2, E-value=0.001)
- PSI-SEARCH (SSEARCH, BLOSUM45, -11/-1, E-value=0.001) (Li *et al.* 2012)
- SSEARCH (SLIM, -11/-1, E-value=0.01) (Pearson 1991, Müller *et al.* 2001)
- SSEARCH (SLIM, -15/-1, E-value=0.1)
- Jackhmmer (E-value=0.0001) (Johnson *et al.* 2010)

Se seleccionaron las secuencias homólogas para cada una de los 12 miembros de la familia SSS con un intérprete de Python mediante una aproximación por consenso de todos estos métodos y parámetros de búsqueda, a fin de evitar la aparición de falsos positivos y falsos negativos, dándole mayor peso a las secuencias que aparecen con más de un método y a los métodos iterativos, y se escribieron en un archivo multi-FASTA como una biblioteca de secuencias.

Además, se realizó una búsqueda con PSI-BLAST sobre PDB para encontrar posibles moldes para la construcción de modelos por homología, cuya secuencia se encuentra en la biblioteca inicial.

10.1.3. Filtrado de resultados de búsqueda

A fin de evitar las posibles redundancias de secuencias con alto porcentaje de identidad en los resultados de la búsqueda, se usó el programa CD-HIT 90 y CD-HIT 99 (Fu *et al.* 2012), los cuales identifican clústeres de secuencias con 90 y 99% de identidad respectivamente y retornan una secuencia representativa de cada uno de estos clústeres.

Además se filtró por longitud de secuencia, removiéndose las secuencias que difirieran en longitud bajo dos desviaciones estándar respecto a la media de las secuencias con un intérprete de Python. Esto debido a que las secuencias de baja longitud pueden afectar los análisis posteriores.

10.2. Alineamientos múltiples de secuencias

10.2.1. Software utilizado en alineamiento múltiple

Los programas utilizados para alineamiento múltiple se detallan en la Tabla II.

Tabla II: Software utilizado en alineamiento múltiple de secuencias

Programa	Método	Parámetros
ClustalO	Alineamiento progresivo global usando perfiles HMM y algoritmo mBed para aproximar construcción de árbol guía (Sievers <i>et al.</i> 2011)	--full (árbol guía desde matriz de distancias)
ClustalW2	Alineamiento progresivo global con árbol guía por Neighbor-Joining y penalidades de gap dinámicas (Thompson <i>et al.</i> 1994, Larkin <i>et al.</i> 2007)	Por defecto
DIALIGN-TX	Alineamiento progresivo local voraz usando segmentos consistentes (Morgenstern 1999, Subramanian <i>et al.</i> 2008)	Por defecto
Expresso	Alineamiento por consistencia incorporando información estructural 3D conseguida automáticamente (Armougom <i>et al.</i> 2006)	Por defecto
Kalign	Alineamiento progresivo global usando algoritmo de Wu-Manber (Lassmann y Sonnhammer 2005)	Por defecto
MAFFT E-INS-i	Alineamiento progresivo iterativo usando transformada rápida de Fourier (FFT), consistencia y penalidades de gaps afines generalizadas (Katoh <i>et al.</i> 2002, Katoh y Toh 2008)	Matrices utilizadas: 1) BLOSUM62 2) SLIM
MAFFT G-INS-i	Alineamiento progresivo iterativo global usando transformada rápida de Fourier (FFT) y consistencia (Katoh <i>et al.</i> 2002, Katoh <i>et al.</i> 2005)	Matrices utilizadas: 1) BLOSUM62 2) SLIM
MAFFT L-INS-i	Alineamiento progresivo iterativo local usando transformada rápida de Fourier (FFT) y consistencia (Katoh <i>et al.</i> 2002, Katoh y Toh 2008)	Matrices utilizadas: 1) BLOSUM62 2) SLIM
M-COFFEE	Meta-método que retorna consenso de 8 programas de alineamiento múltiple por consistencia (Wallace <i>et al.</i> 2006)	Por defecto
MUSCLE	Alineamiento progresivo iterativo usando refinamiento de árbol guía inicial (Edgar 2004a, Edgar 2004b)	Por defecto
MSAProbs	Alineamiento progresivo iterativo usando pares HMM y probabilidades posteriores desde	Por defecto

	función de partición (Liu <i>et al.</i> 2010)	
PicXAA	Alineamiento no progresivo probabilístico local que calcula el alineamiento de máxima exactitud esperada (Sahraeian y Yoon 2010)	Por defecto
PnpProbs	Alineamiento con mejor manejo de árboles guías: progresivo para secuencias relacionadas y no progresivo para secuencias distantes (Ye <i>et al.</i> 2016)	Por defecto
POAv2	Alineamiento local usando una representación gráfica del alineamiento (Alineamiento de Orden Parcial) (Lee <i>et al.</i> 2002)	Por defecto
PRALINE	Alineamiento usando bipartición entre segmentos TM y no TM, usando distintas matrices de puntaje para cada caso (Simossis y Heringa 2005)	
PRANK	Alineamiento que distingue entre eventos de inserción y deleción para mejor reconstrucción filogenética (Löytynoja y Goldman 2008)	Por defecto
ProbCons	Alineamiento progresivo usando un puntaje de probabilidad de consistencia a partir de pares HMM (Do <i>et al.</i> 2005)	Por defecto
PSI/TM-COFFEE	Alineamiento por consistencia usando extensión por homología sobre bases de datos de proteínas de membrana (Floden <i>et al.</i> 2016)	Por defecto
T-COFFEE	Alineamiento progresivo por consistencia (Notredame <i>et al.</i> 2000)	Por defecto
TM-ALIGNER	Alineamiento progresivo con predicción de segmentos transmembrana con TMHMM, alineamiento con programación dinámica y unión de los sitios con algoritmo de Wu-Manber (Bhat <i>et al.</i> 2017)	

Con los distintos MAFFT se usará también la matriz asimétrica de puntajes SLIM, diseñada para proteínas transmembrana.

10.2.2. Análisis de alineamientos múltiples de secuencias

Se desarrolló un intérprete en Python para obtener las posiciones conservadas en los alineamientos con porcentajes de conservación a elección, así como la longitud de los alineamientos y promedio de gaps. También en el mismo intérprete, haciendo uso del programa Prody (Bakan *et al.* 2014, Bakan *et al.* 2011) se calculó para todos los alineamientos la entropía de Shannon, ocupancia e información mutua de cada posición.

10.3. Reconstrucción filogenética

Se utilizó el software RaXML (Stamatakis 2014), el cual hace uso de una aproximación de máxima verosimilitud para proponer el mejor modelo evolutivo, topología y longitudes de ramas de un árbol filogenético a partir de un alineamiento múltiple de secuencias. Se usó un modelo Gamma de variabilidad de tasas de sustitución entre sitios, permitiendo al programa elegir automáticamente el modelo evolutivo mediante el criterio de información Bayesiana (BIC) (Posada y Buckley 2004) y realizando 1000 corridas de bootstrap. Estos parámetros se aplicaron a todos los alineamientos realizados anteriormente.

Los árboles filogenéticos generados son visualizados y anotados haciendo uso del servidor iTol (Letunic y Bork 2016) y el paquete ETEToolkit escrito en Python (Huerta-Cepas *et al.* 2016).

10.4. Predicción de segmentos transmembrana

La predicción de los segmentos transmembrana se realizó con TOPCONS (Bernsel *et al.* 2009), el cual ofrece un consenso de 5 programas de predicción (OCTOPUS, Philius, PolyPhobius, SCAMPI y SPOCTOPUS), MEMSAT-SVM (Nugent y Jones 2009) y MEMSAT3 (Jones 2007).

10.5. Construcción de modelos comparativos

10.5.1. Construcción del modelo



Los modelos de las proteínas son construidos siguiendo el orden del árbol filogenético usando el programa Modeller (Sali y Blundell 1993), partiendo desde la ubicación del molde tridimensional. Esto se justifica mediante la evidencia presente en literatura de que la conservación de estructura tridimensional es mayor para secuencias relacionadas evolutivamente (Smith 2008). Para cada una de las secuencias en el árbol se construirá una serie de modelos candidatos y se seleccionará el mejor, según la evaluación posterior y las características estructurales y de secuencia descritas en la literatura.

Para cada una de las proteínas, se construirá un complejo con el número de subunidades según lo indicado en literatura o lo deducido para cada estructura, lo que puede lograrse mediante el programa Prody.

10.5.2. Evaluación de los modelos

A fin de evaluar y seleccionar los modelos construidos se utilizarán dos criterios, uno estructural y otro funcional:

1. Criterio estructural: Evaluación estereoquímica y de interacción proteína-ligando inferida por los alineamientos y la literatura.
2. Reproducción de resultados experimentales reportados en literatura, experimentos *in vitro* y otras simulaciones en las simulaciones a realizar.

10.6. Dinámica molecular

10.6.1. Preparación de los modelos para las simulaciones

Los modelos de hSGLT2 obtenidos, en formato PDB, se procesan con lenguaje Tk/Tcl en el visualizador VMD (Humphrey *et al.* 1996), con el cual se generan los archivos PSF (Protein Structure File), que contiene especificaciones estructurales para átomos, enlaces, ángulos, tanto entre átomos, diedros e impropios, y términos cruzados. Se utilizan los archivos de topología y de parámetros de CHARMM36 para proteínas (top_all36_prot.rtf y par_all36_prot), lípidos (top_all36_lipid y par_all36_lipid), carbohidratos (top_all36_carb y par_all36_carb) y agua e iones (toppar_water_ions). Una vez hecho esto se construye la membrana haciendo uso del constructor de membrana disponible en el sitio de CHARMM-GUI (<http://www.charmm-gui.org/?doc=input/membrane.bilayer>), la cual estará conformada de fosfolípidos POPC y se inserta en ella la proteína. La proteína debe estar orientada respecto al eje z de la membrana, y su orientación se obtiene a partir de las estructuras cristalizadas

de vSGLT de la base de datos OPM (*Orientation of Proteins in Membranes*) (Lomize *et al.* 2006) y los modelos son alineados con esta mediante un alineamiento estructural con cealign (en Pymol).

Una vez montado el sistema proteína-membrana, este debe solvatarse. Para esto se usa el programa Solvate (Heller *et al.* 1993), el cual solvata el sistema de forma explícita. Con ambos se crea una caja de agua alrededor del sistema proteína-membrana de 200x200x200.

Luego, se ioniza la caja de agua añadiendo concentraciones fisiológicas de cloruro de sodio (0.4 nmol/L) a cada lado de la membrana usando el plugin autoionize de VMD.

10.6.2. Dinámica de minimización



Se realizará una dinámica molecular de minimización del ensamble generado anteriormente usando el software NAMD. Esta dinámica usará un termostato de Langevin a 310 K y un baróstato de Langenvin a 1 atm a volumen constante, los cuales permiten tomar en cuenta fuerzas e interacciones dispersivas. Además, contará con condiciones de borde periódico. Cada paso se realizará en 2 femtosegundos hasta 0.5 nanosegundos. La dinámica se realizará mediante supercálculo en el GPU Southern Cluster.

10.6.3. Dinámica de equilibración

Se realizarán dos dinámicas moleculares de equilibración del ensamble minimizado usando el campo de fuerza Amber. Esta dinámica usará un termostato de Langevin a 310 K y un baróstato de Langenvin a 1 atm a volumen constante. Además, contará con condiciones de borde periódico. Cada paso se realizará en 2 femtosegundos hasta 0.5 nanosegundos. La dinámica se realizará mediante supercálculo en el GPU Southern Cluster.

10.6.4. Dinámica de producción

Finalmente, se realizará la dinámica de producción, la cual permitirá estudiar cómo interacciona el ensamble con las moléculas candidatas como posibles inhibidores. Estas dinámicas poseerán las mismas características descritas anteriormente, con la diferencia de que se realizarán pasos de 2 femtosegundos hasta 50 nanosegundos de simulación. La dinámica se realizará mediante supercálculo en el GPU Southern Cluster.

10.6.5. Análisis de las dinámicas moleculares

Se analizarán los resultados de las dinámicas de producción con los programas Prody (Bakan *et al.* 2011) y MDAnalysis (Michaud-Agrawal *et al.* 2011), a fin de caracterizar y validar los modelos generados ya equilibrados respecto a la literatura, tomando en cuenta interacciones proteína-ligando, formación de las cavidades y poro central, transporte de agua y movimiento de dominios.

El modelo de hSGLT2 que se usará para docking se seleccionará a partir de un agrupamiento jerárquico de las conformaciones proteicas y el cálculo de los centroides de los clusters, de los que se elegirá el centroide del cluster más poblado, que es teóricamente el más probable y por tanto, la conformación más estable.

10.7. Creación de biblioteca de inhibidores

Se construirá una biblioteca de inhibidores conocidos reportados en literatura y almacenados en bases de datos para evaluar las interacciones de los modelos con estos basados en docking molecular. Los inhibidores se obtendrán desde PubChem Compound, a partir de los ensayos *in vivo* o *in vitro* sobre hSGLT2 reportados en literatura y clasificados como “Activos”.



10.8. Docking Molecular

10.8.1. Detección de cavidades en las estructuras de los modelos

Se usará fpocket (Le Guilloux *et al.* 2009, Schmidtke *et al.* 2011) para detectar las posibles cavidades que pueden interactuar con los compuestos de las bibliotecas en las estructuras de los modelos.

10.8.2. Docking molecular

El docking molecular entre las estructuras de los modelos de hSGLT2 y la biblioteca de búsqueda se realiza usando el programa DOCK6 (Allen *et al.* 2015). Desde los resultados de score del docking se seleccionarán los mejores compuestos candidatos a inhibidores.

10.9. Identificación de inhibidores desde búsqueda en base de datos

10.9.1. Docking masivo con Dock 6

Usando la colección STK/STL, perteneciente a Laboratorios Vitas-M (<https://www.vitasmlab.biz/compound-libraries/stock-libraries>), con 1,371,136 compuestos disponibles comercialmente para screening biológico y optimización de leads disponibles, se realizará un docking masivo con Dock6, apoyado en supercálculo, contra el modelo generado y se analizarán estadísticamente los puntajes resultantes respecto a la distribución de puntajes obtenida en la caracterización de la base de datos de inhibidores.

10.9.2. Evaluación estadística de moléculas similares y con mejores puntajes

Luego de esto, se separarán por similitud, usando el índice de similitud de Dice, aquellos compuestos que presenten una similitud estructural superior a 0.6 respecto a los centroides obtenidos desde la base de datos de inhibidores conocidos, comparación hecha tras conversión a *fingerprints* circulares de Morgan.

Así, resultarán 3 grupos de valores: los puntajes de la base de datos STK/STL, los de similitud superior a 0.6 separados desde estos y los de inhibidores conocidos (como control). A estos tres conjuntos de observaciones se les aplicarán pruebas estadísticas para determinar si es posible distinguir significativamente entre estos grupos de puntajes de docking.

Para cuantificar el tamaño de efecto estadístico entre estas mediciones de puntajes de docking, usaremos la superposición, la tasa de clasificación errónea y el estadístico d de Cohen.

Para calcular superposición, en primer lugar calculamos el valor en que ambas distribuciones se intersectan usando la fórmula:

$$Intersecto = \frac{(DesvEst_1 Media_2 + DesvEst_2 Media_1)}{DesvEst_1 + DesvEst_2}$$

Ya que este valor es válido para distribuciones normalmente distribuidas, en casos donde no se cumpla la normalidad y no se pueda obtener mediante la fórmula, se obtendrá desde el valor en común de densidad de probabilidad. Con el valor de intersección como límite, calculamos la superposición tomando en cuenta el número de valores que son superiores o inferiores a este límite, según sea el valor de sus medias:

$$Superposición = \frac{N^\circ \text{ valores post - límite}}{N^\circ \text{ valores totales}}$$

La tasa de clasificación errónea, o tasa de error, entrega la probabilidad de que un valor de una distribución sea incorrectamente atribuida a la otra distribución en estudio, y es obtenida por el simple promedio entre ambas superposiciones:

$$Tasa\ de\ error = \frac{Superposición_1 + Superposición_2}{2}$$

En tanto, el tamaño de efecto de Cohen, o d de Cohen, es un índice adimensional que entrega la diferencia en desviaciones estándar entre dos grupos:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

donde:

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$$

es la desviación estándar agrupada entre los dos grupos.

Los intervalos de confianza se calcularán mediante el método de remuestreo *bootstrap* de 1000 muestras de 100 mediciones cada una sobre las distribuciones resultantes.

Finalmente, haremos test de hipótesis para evaluar significancia estadística; sin embargo, como se mencionará en la discusión, debido a la naturaleza de la adquisición de las muestras, estos resultados no contemplan una importancia fundamental.

11. RESULTADOS

11.1. Secuencias homólogas obtenidas desde SwissProt y Protein Data Bank

La búsqueda con los métodos y especificaciones detalladas en 10.1.2. entregó como resultados el número de secuencias que se resume en la Tabla III, a las que se añadieron los moldes estructurales 2XQ2 y 3DH4, para un total de 82 secuencias.

Tabla III: Métodos de búsqueda y especificaciones para la búsqueda en la base de datos SwissProt

Método de búsqueda	Homólogos hSGLT1	Homólogos hSGLT2	Consenso familia SSS	Filtro CD-HIT 90
PSI-BLAST (BLOSUM62, -11/-1, E=0.01)	113	114	273	242
PSI-BLAST (BLOSUM62, -11/-1, E=0.001)	114	114	309	275
PSI-BLAST (BLOSUM45, -14/-2, E=0.001)	110	112	112	90
PSI-SEARCH (SSEARCH, BLOSUM45, -11/-1, E=0.0001)	148	100	152	91
Jackhmmer (E=0.0001)	113	113	674	574
SSEARCH (SLIM, -11/-1, E=0.01)	112	103	115	92
SSEARCH (SLIM, -15/-1, E=0.1)	135	124	161	134
TOTAL CONSENSO			113	76

11.2. Alineamiento múltiple de secuencias y conservación

El alineamiento de secuencias obtenidas por consenso entregó los resultados que se resumen en la Tabla IV.

Tabla IV: Resultados de alineamiento y análisis de conservación

Método	Longitud (Número de sitios)	Número de gaps (Porcentaje)	Número de sitios con >90% conservación	Número de sitios con >80% conservación
ClustalO	1093	493 (45.11)%	11	22
ClustalW	902	302 (33.48)%	11	21
Dialign-TX	1432	832 (58.1)%	12	23
Expresso	1463	863 (58.99)%	12	24
Kalign	1415	815 (57.6)%	12	22
Mafft E-INS-i	1187	587 (49.45)%	12	24
Mafft E-INS-I (SLIM)	1215	615 (50.62)%	12	22
Mafft G-INS-i	1104	504(45.65)%	12	23
Mafft G-INS-i (SLIM)	1142	542 (47.46)%	13	25
Mafft L-INS-i	1151	551 (47.87)%	12	23
Mafft L-INS-i (SLIM)	1255	655 (52.19)%	12	23
M-Coffee	1321	721 (54.58)%	12	22
MSAPROBS	1250	650 (52.0)%	13	23
MUSCLE	1027	427 (41.58)%	13	24
Pnp-Probs	1436	836 (58.22)%	13	23
POA	2123	1523 (71.74)%	11	21
PRALINE	992	392 (39.52)%	12	22
PRANK	2337	1737 (74.33)%	8	18
ProbCons	1520	920 (60.53)%	13	26
Promals	1135	535 (47.14)%	12	22
PSI/TM-COFFEE	1429	829 (58.01)%	12	23
T-COFFEE	1507	907.0 (60.19)%	12	24
TM-ALIGNER	1124	525.0 (46.71)%	12	22

El alineamiento que se seleccionó para realizar el modelado guiado por filogenia fue el de MAFFT G-INS-i con matriz SLIM, debido a la cantidad de sitios con sobre 90% y 80% de conservación, su longitud y su menor número de gaps. Si bien ProbCons mostró similar cantidad de sitios conservados, a causa de las diferencias en su longitud y porcentaje de gaps y al alineamiento apropiado de los segmentos transmembrana (ver Figuras 15, 16 y 17) se seleccionó finalmente el alineamiento realizado con MAFFT G-INS-i (SLIM).

La elección del alineamiento a usar en la etapa de modelado siguió varios criterios. El principal fue la correspondencia de los segmentos transmembrana predichos por TOPCONS y MEMSAT-SVM con bloques reconocibles en el alineamiento (Figura 21).

Los segmentos transmembrana poseen mayor conservación que los lazos intra- y extracelulares y los programas de alineamiento múltiple que son capaces de reconocer esta tendencia suelen agrupar los segmentos transmembrana en bloques mejor definidos y con menos aperturas

Otras observaciones adicionales para validar este alineamiento incluyen hallar patrones en la secuencia, como secuencias consenso, puede ser útil para descubrir si el alineamiento replica lo descubierto y reportado en la literatura. En la Figura 17 se muestra un logo generado desde el alineamiento que muestra la secuencia consenso de la familia SSS.

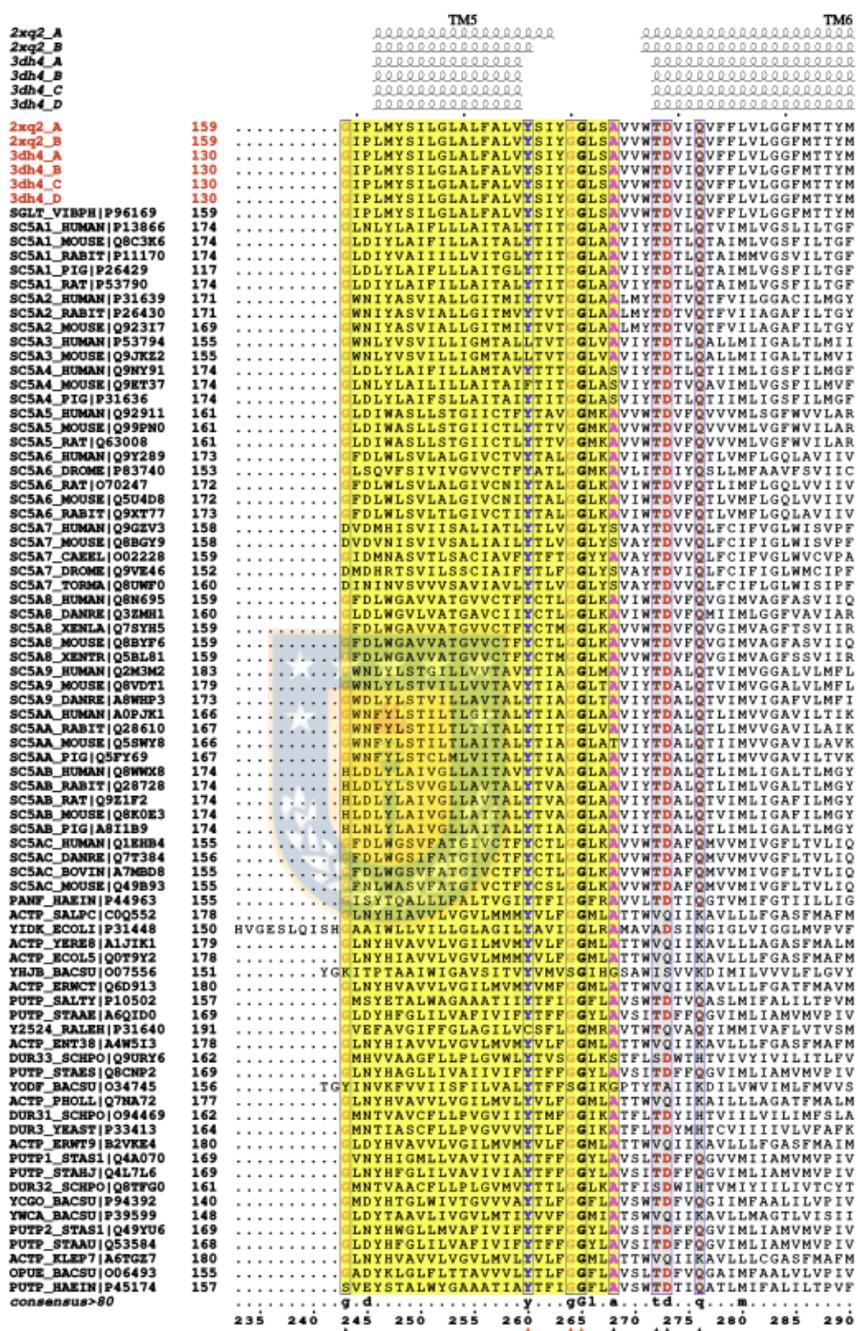


Figura 15: Sección del alineamiento de las secuencias obtenidas con MAFFT G-INS-i y matriz SLIM. Las flechas negras en la parte inferior indican sitios con >80% de conservación, mientras que las flechas rojas indican sitios con >90% de conservación y las estrellas rojas indican conservación completa. En amarillo se destaca la secuencia consenso de la familia SSS ([GS]-2(2)-[LIY]-x(3)-[LIVMFYWSTAG](7)-x(3)-[LIY]-[STAV]-x(2)-G-G-[LMF]-x-[SAP])

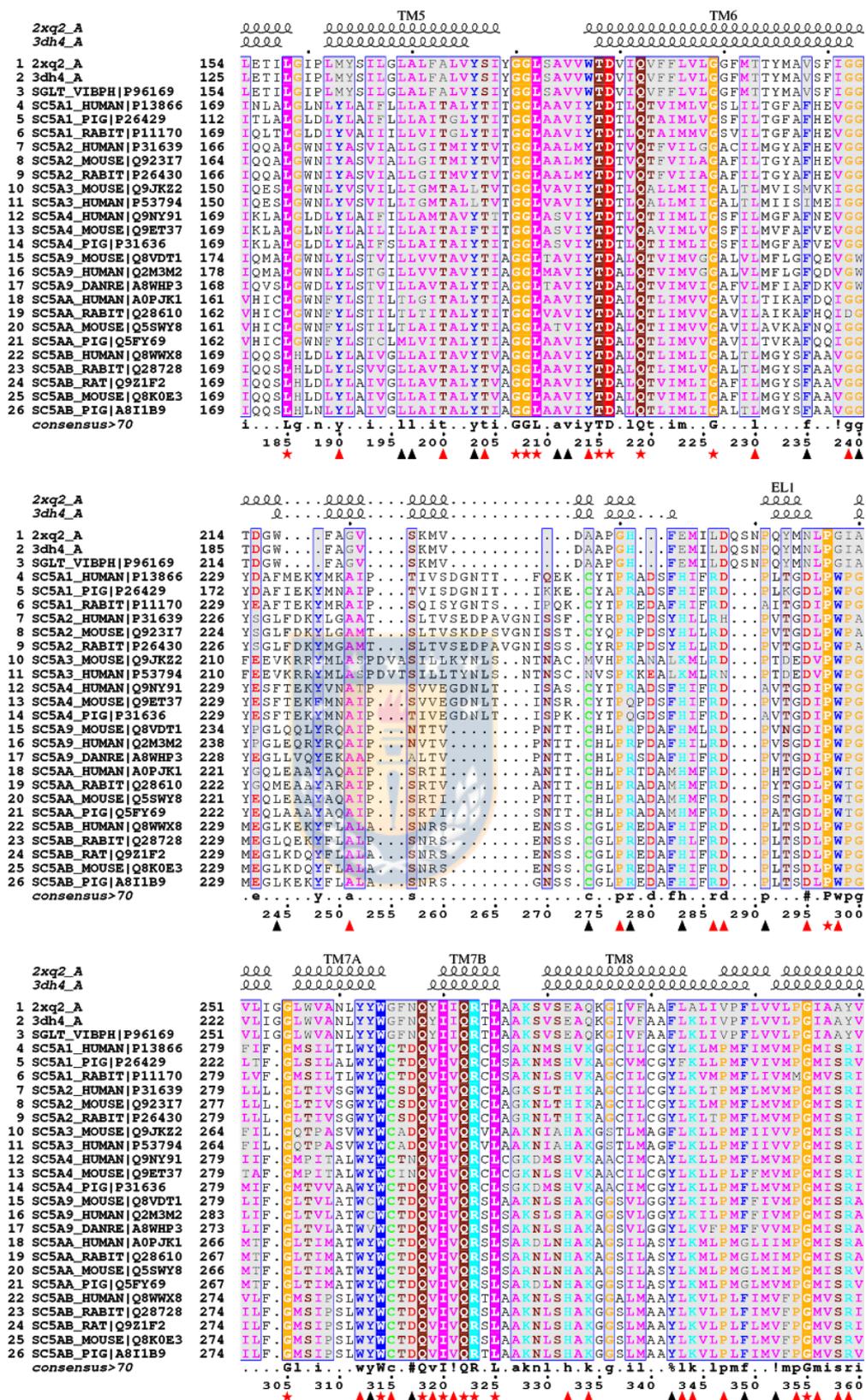


Figura 16: Sección del alineamiento de las secuencias de la familia SSS con MAFFT G-INS-i y matriz SLIM. Las flechas negras en la parte inferior indican sitios con >80% de conservación, mientras que las flechas rojas indican sitios con >90% de conservación y las estrellas rojas indican conservación completa.

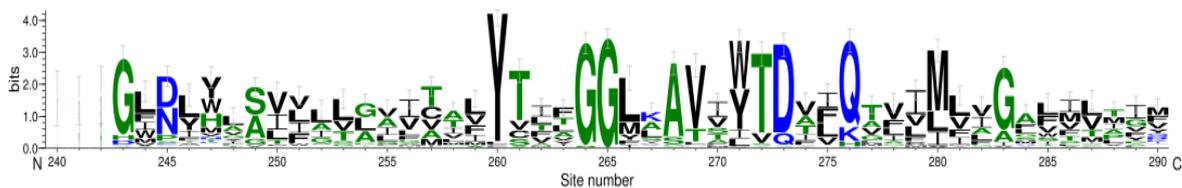


Figura 17: Logo de los sitios 240-290 del alineamiento con MAFFT G-INS-i. En esta sección se encuentra la secuencia consenso de la familia SSS, [GS]-2(2)-[LIY]-x(3)-[LIVMFYWSTAG](7)-x(3)-[LIY]-[STAV]-x(2)-G-G-[LMF]-x-[SAP]

En la Tabla V se muestran los sitios conservados con sobre 90% de conservación dentro del alineamiento seleccionado.

Tabla V: Sitios del alineamiento sobre 90% de conservación, con sus respectivas entropías de Shannon, residuos equivalentes en hSGLT1 y hSGLT2

Sitio >90% conservación	Entropía de Shannon (bits)	Residuo en hSGLT2	Residuo en hSGLT1	Ubicación
128	0.288	GLY83	GLY86	TM2B
139	0.066	GLY92	GLY95	TM2-TM3
188	0.359	ARG132	ARG135	IL1-TM4
260	0.246	TYR191	TYR188	TM5
264	0.115	GLY195	GLY192	TM5-TM6
265	0.0	GLY196	GLY193	TM5-TM6
393	0.36	PRO269	PRO269	EL1
457	0.288	ARG300	ARG300	TM7B
489	0.115	GLY332	GLY332	TM8
595	0.356	ASP408	ASP408	TM9
679	0.26	GLY484	GLY484	TM12
683	0.312	GLY488	GLY488	TM12
788	0.336	SER545	SER545	TM13

11.3. Predicción de segmentos transmembrana

La predicción de los segmentos transmembrana con MEMSAT-SVM entregó los resultados esquematizados en la Figura 18.

En tanto, TOPCONS entregó los resultados que se muestran en la Figura 19, donde se muestra el consenso realizado y la referencia entregada del molde PDBID: 2xq2.

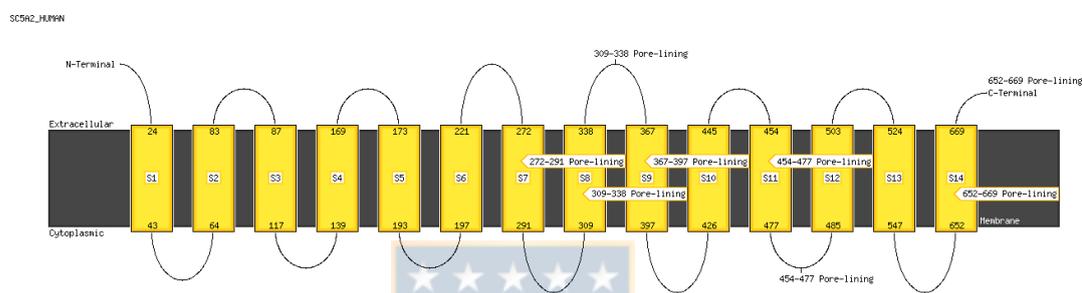


Figura 18: Esquema de los resultados de predicción de segmentos transmembrana entregados por MEMSAT-SVM para hSGLT2

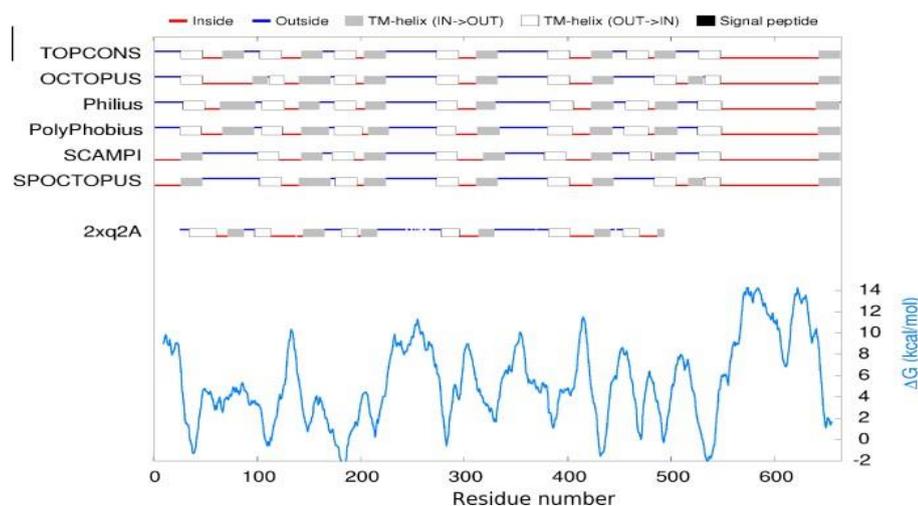


Figura 19: Resultados de predicción de segmentos transmembrana entregados por TOPCONS para hSGLT2

11.4. Reconstrucción filogenética

En la Tabla VI se detallan los resultados de la reconstrucción filogenética para los 20 programas de alineamiento múltiple, mientras que en la Figura 20 se muestra el resultado de la reconstrucción filogenética para MAFFT G-INS-i con matriz SLIM.

Tabla VI: Resultados de reconstrucción filogenética con RAxML para los alineamientos

Método de alineamiento	Modelo evolutivo escogido por RAxML	Puntaje de Máxima Verosimilitud de mejor árbol	Alfa	Longitud de mejor árbol
ClustalO	LG+F*	-53468.810	1.405090	58.596
ClustalW	LG+F	-54315.613	1.299671	72.385
Dialign-TX	LG+F	-53813.647	1.3497707	60.103
Expresso	LG+F	-53610.273	1.314094	65.733
Kalign	LG+F	-53064.858	1.378283	61.402
MAFFT E-INS-i	LG+F	-53046.655	1.428919	58.045
MAFFT E-INS-i (SLIM)	LG+F	-52990.143	1.375133	60.868
MAFFT G-INS-i	LG+F	-52980.419	1.486693	54.253
MAFFT G-INS-i (SLIM)	LG+F	-52994.982	1.407791	59.248
MAFFT L-INS-i	LG+F	-53050.822	1.449399	55.930
MAFFT L-INS-i (SLIM)	LG+F	-53021.268	1.412841	57.936
M-COFFEE	LG+F	-53894.449	1.254585	71.853
MSAProbs	LG+F	-57151.919	0.908500	115.791
MUSCLE	LG+F	-56391.631	1.118972	79.884
PnpProbs	LG+F	-55398.404	1.054389	92.695
POA	LG+F	-54168.733	1.327480	63.334
Praline	LG+F	-55178.880	1.206007	74.092
PRANK	LG+F	-52691.302	1.59731	44.750
ProbCons	LG+F	-56124.655	0.981862	100.694
Promals	LG+F	-55736.501	0.984705	104.121
PSI/TM-COFFEE	LG+F	-54029.076	1.194737	75.814
T-COFFEE	LG+F	-53563.899	1.350734	62.642
TM-ALIGNER	LG+F	-53887.893	1.351202	65.421

*F: Frecuencias empíricas

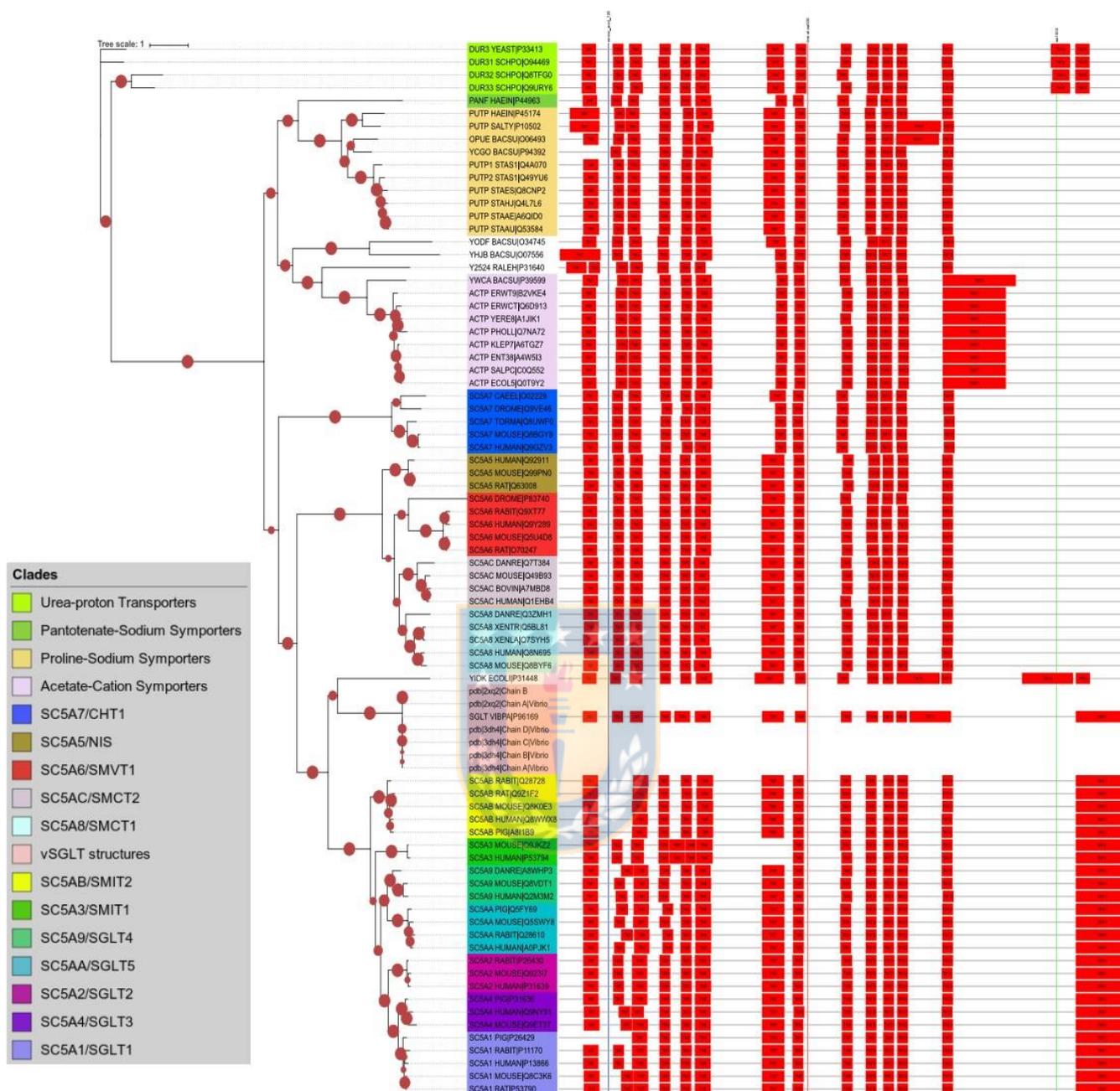


Figura 20: Árbol filogenético realizado mediante método de máxima verosimilitud con RAxML para el alineamiento de secuencias homólogas de la familia SSS con MAFFT G-INS-i y matriz SLIM. Los segmentos transmembrana predichos por TOPCONS se encuentran esquematizados a la izquierda en los rectángulos rojos, como se encuentran dentro del alineamiento. Los círculos indican el nivel de bootstrap de cada clado y nodo del árbol.

11.5. Resultados de modelado comparativo

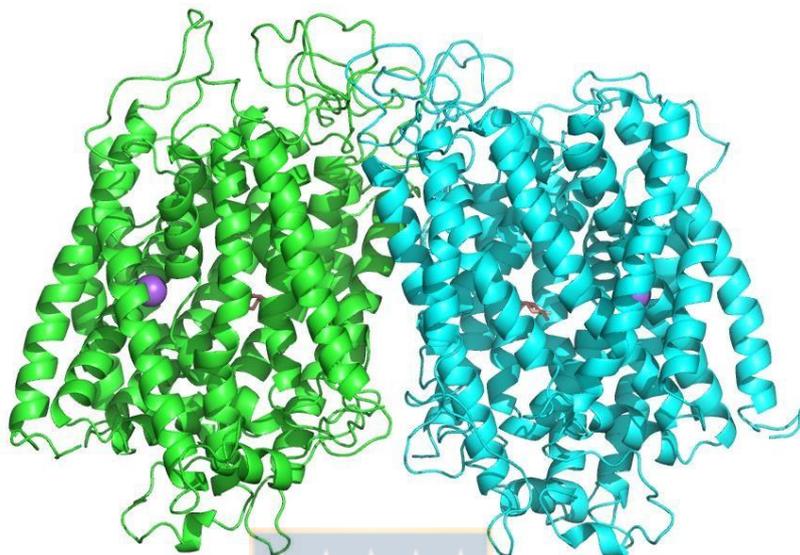
Los 30 modelos diméricos producidos para hSGLT2 desde el molde de 3DH4 fueron evaluados por MODELLER con el puntaje de potencial estadístico DOPE y dieron como resultado los puntajes que se resumen en la Tabla VII, donde se organizan desde el mejor al peor puntaje.

Tabla VII: Resultados ordenados según puntaje DOPE para modelado de hSGLT2

Modelo	Puntaje DOPE
3	-167663.953125
17	-167495.359375
8	-166555.78125
13	-166317.578125
1	-166275.125
7	-166139.875
19	-166086.390625
23	-165936.265625
16	-165477.375
14	-165418.28125
4	-165390.890625
15	-165311.65625

24	-165197.28125
9	-164862.25
27	-164758.421875
30	-164713.828125
26	-164515.203125
6	-164455.828125
22	-164434.40625
11	-164189.265625
21	-164178.234375
25	-164110.953125
28	-164014.703125
10	-164000.0
2	-163853.171875
29	-163208.015625
20	-162968.859375
12	-162764.34375
18	-162232.734375
5	-161518.78125

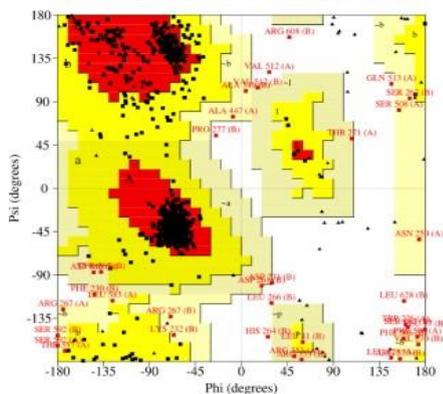
En la Figura 21 se muestra el resultado para el modelo 3 de hSGLT2.



**Figura 21: Modelo de hSGLT2 (modelo número 3)
producido con MODELLER**

En la Figura 22 se muestra los análisis estructurales realizados sobre el modelo, generados con Procheck, previo al refinamiento y equilibración posterior con dinámica molecular. En la Figura 23 se muestran los residuos que interactúan con el residuo de Galactosa dentro del modelo de hSGLT2

Gráfico de Ramachandran, todos los residuos



Análisis de Ramachandran de modelo número 3 de hSGT2 sin refinamiento

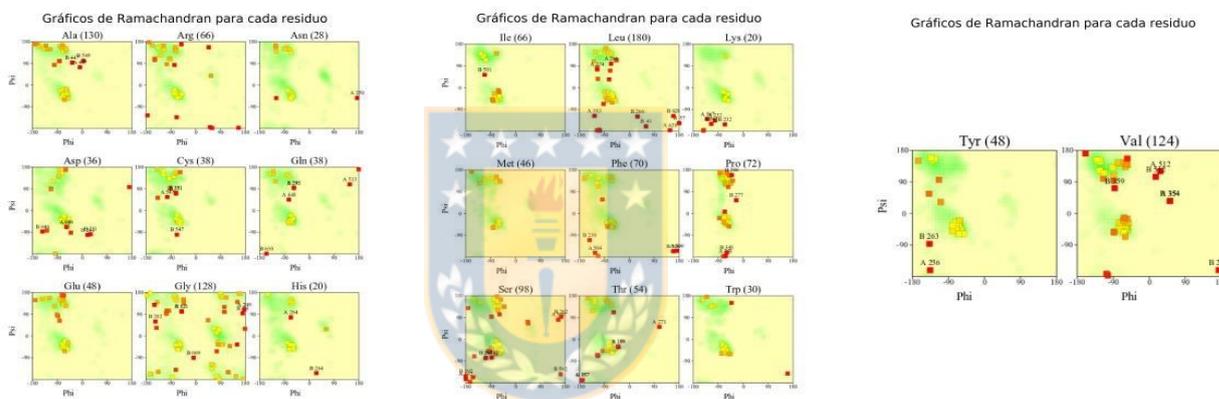


Figura 22: Análisis de Ramachandran de modelo 3 de hSGLT2 previo a la minimización

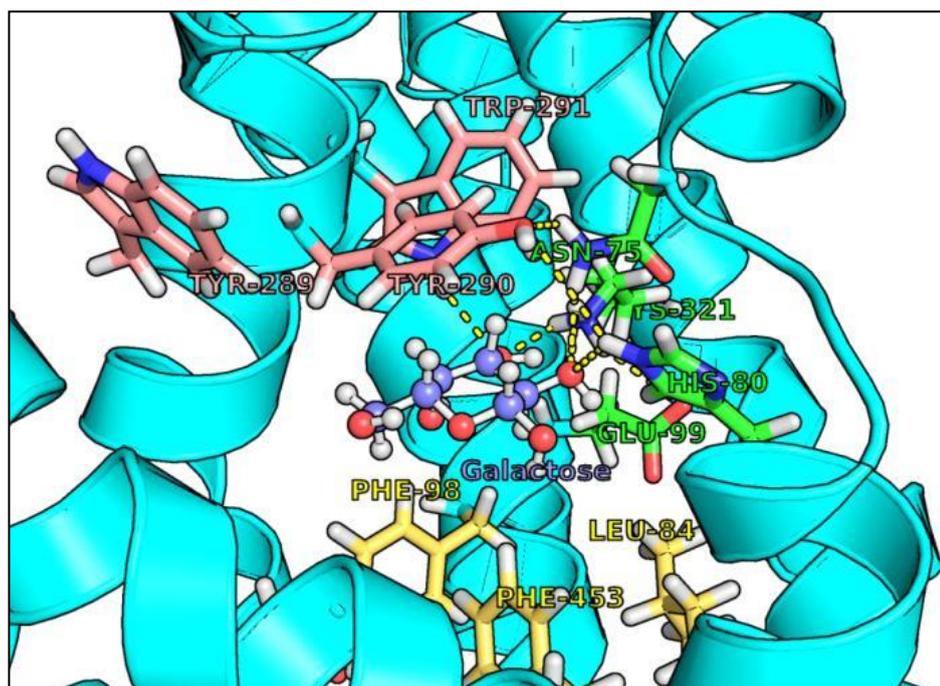


Figura 23: Residuos que interactúan con la Galactosa en el modelo de hSGLT2. Los residuos de la puerta extracelular, obtenidos por literatura, se colorearon con rosado, los del sitio de unión a azúcares con verde y los de la puerta intracelular en amarillos

11.6. Dinámica molecular para refinamiento

El sistema resultante de los procesamientos del modelo con CHARMM-GUI y Solvate para el modelo de hSGLT2 contiene 208091 moléculas de agua, 999 moléculas de POPC, 1584 iones cloruro, 1584 iones sodio (2 de los cuales se encuentran en el sitio de unión a sodio de la proteína, co-cristalizados desde el modelo) y 2 moléculas de galactosa situadas en el sitio de unión a azúcares de la proteína. Las dimensiones totales del sistema son 200x200x200 Angstrom, y se muestra en la Figura 24.

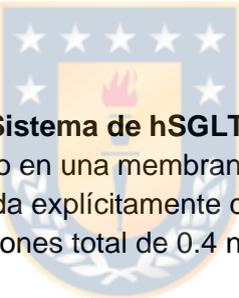
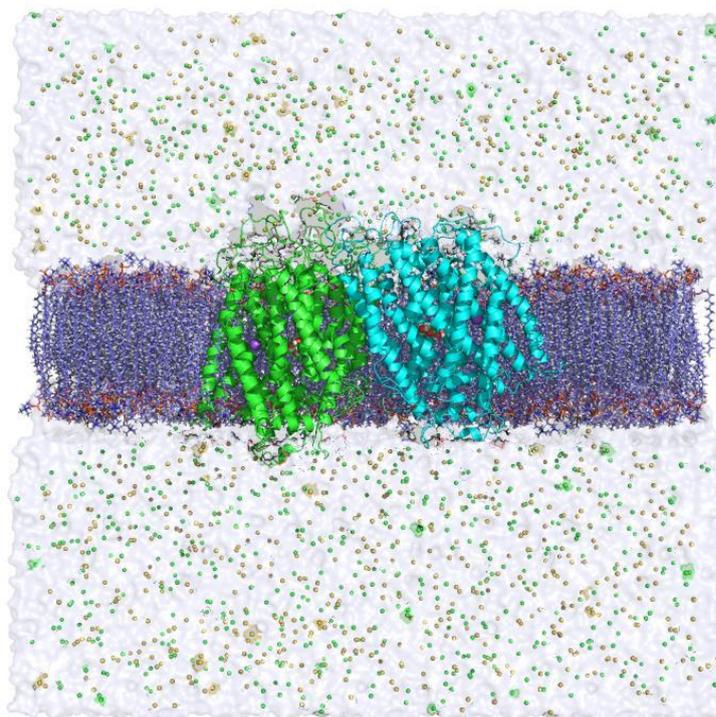


Figura 24: Sistema de hSGLT2-membrana.
hSGLT2 inserto en una membrana de 200x200 de
POPC solvatada explícitamente con concentración
de iones total de 0.4 nmol/L

El análisis de RMSD y RMSF de la dinámica molecular de hSGLT2 se muestran en la Figura 25.

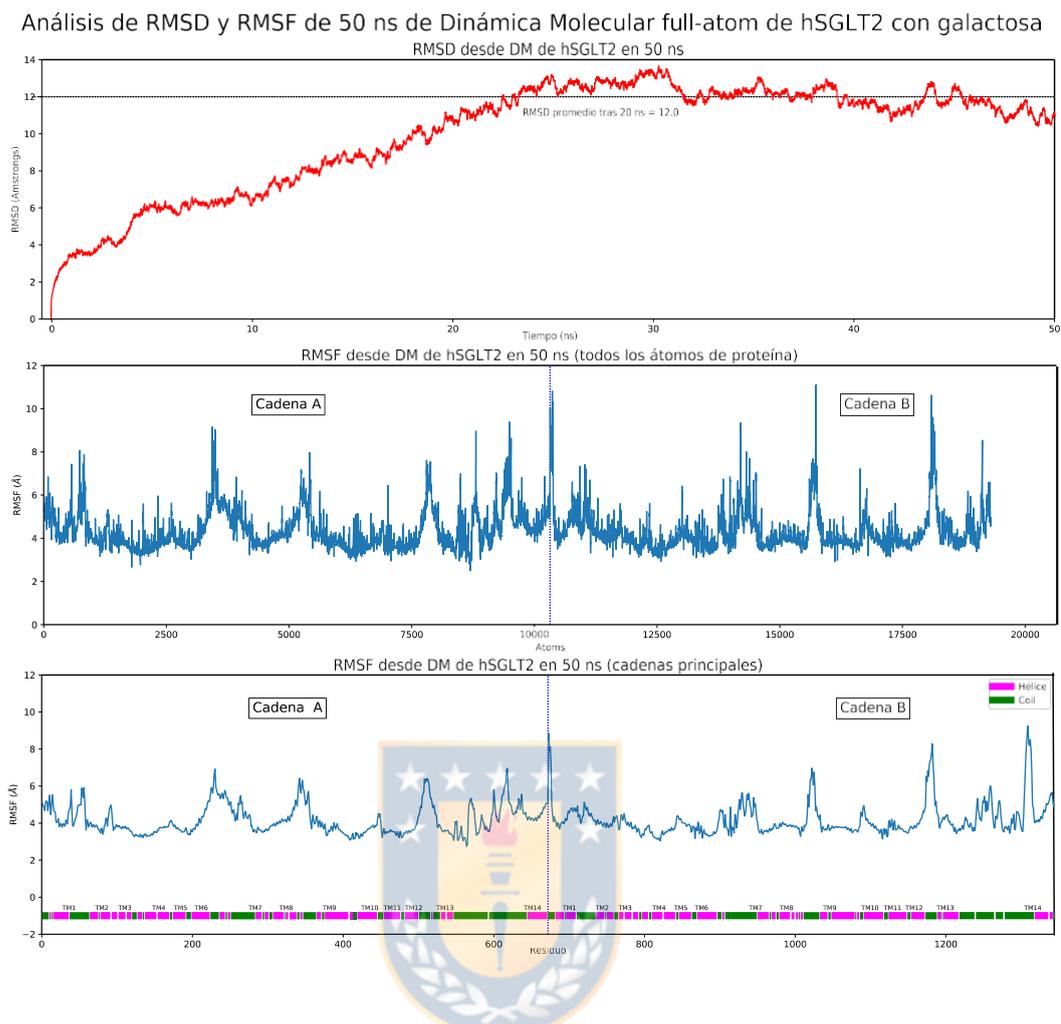


Figura 25: Análisis de RMSD y RMSF para 50 ns dinámica molecular de hSGLT2. Bajo la figura de RMSF se muestran los segmentos transmembrana de la proteína

En la figura 26 se muestra la distribución conjunta de RMSD y Radio de giro para una muestra de los pasos de las Dinámicas Moleculares de ambos transportadores. Mientras el RMSD permite conocer la variación de posición promedio de los átomos de la proteína, el radio de giro ofrece inferencias sobre su apertura radial durante la simulación.

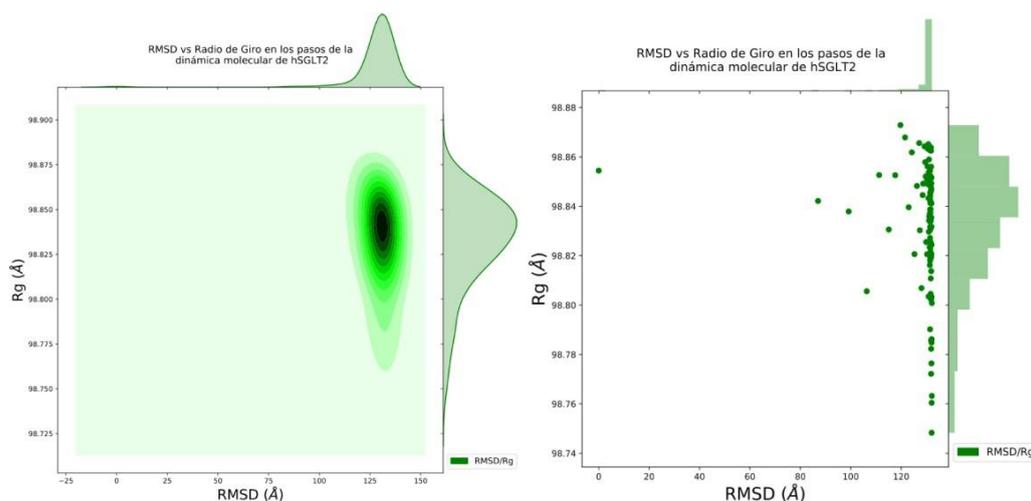


Figura 26: Distribución conjunta de valores de RMSD y Radio de giro durante las Dinámicas Moleculares de hSGLT2

En la Figura 27 se muestra el análisis del poro realizado con el software HOLE (Smart *et al.* 1993) y en la Figura 28 una representación del poro obtenida con HOLLOW (Bosco *et al.* 2008)

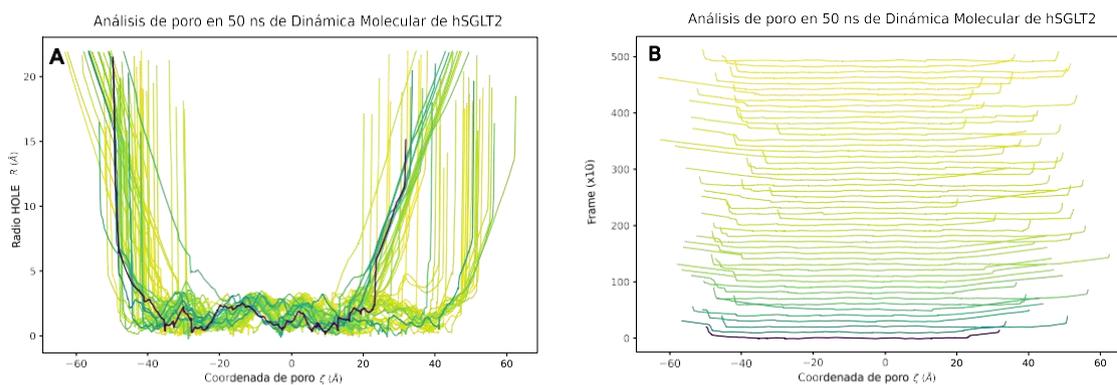


Figura 27: Análisis de poro en 50 ns de Dinámica Molecular de hSGLT2. A) Dinámica del poro de transporte, generado con HOLE para hSGLT2. B) Igual que A), con una separación entre pasos de la dinámica para mejor visualización

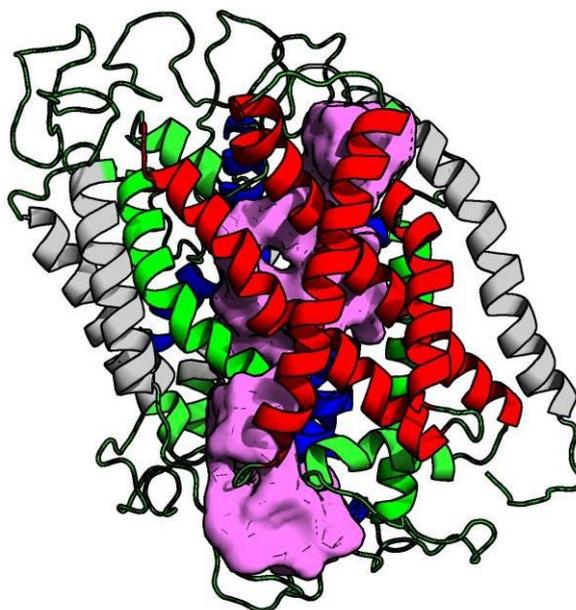


Figura 28: Determinante de poro de agua en hSGLT2. En esta imagen generada con el software HOLLOW, se observa la trayectoria del canal acuoso en hSGLT2. La abertura central es resultado de la bifurcación del canal hacia los dos sitios de unión de sodio, los que impulsan el transporte del sustrato.

11.7. Agrupamiento jerárquico y extracción de estructuras centroides

Los pasos de la trayectoria resultante de la dinámica molecular fueron agrupados mediante agrupamiento jerárquico de RMSD con los métodos UPGMA, WPGMA y de centroide. En la Figura 29 se muestra el resultado del agrupamiento de conformaciones para la dinámica molecular de hSGLT2 con WPGMA debido a que este método propuso una distinción de un mayor número de clusters, del que se seleccionó el centroide E, el más poblado, ya que representa teóricamente la conformación más probable y por ende, más estable.

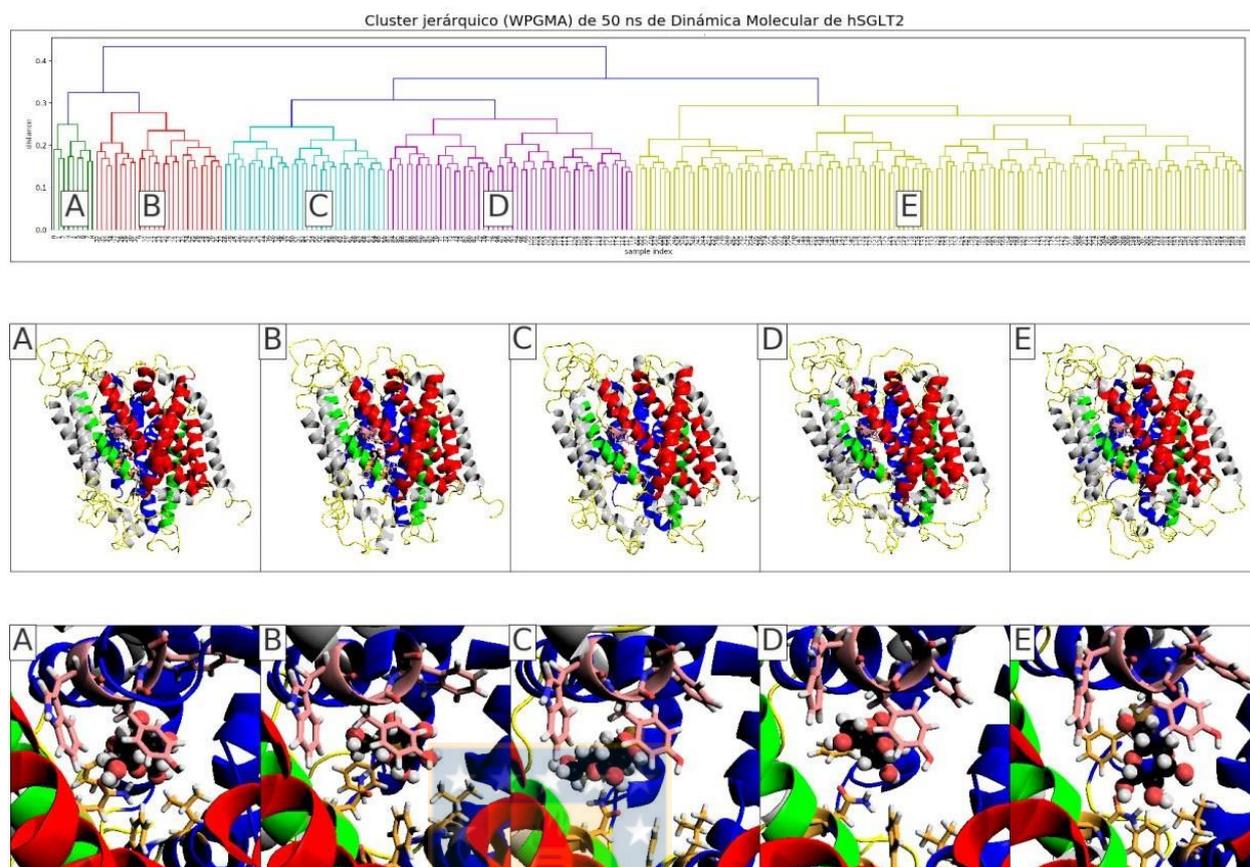


Figura 29: Agrupamiento de conformaciones de hSGLT1 y hSGLT2. En las dos fila superiores se muestran los agrupamientos jerárquicos de RMSD mediante WPGMA de 5000 pasos de dinámica molecular (50 ns) de hSGLT1 y hSGLT2 con galactosa y sodio. En la tercera fila, se muestran los resultados de los centroides de hSGLT2 con los dominios en distintos colores (dominio haz: azul; dominio hash: rojo, hélices auxiliares: verde; lazos: amarillo). En la fila inferior se muestra un acercamiento al sitio de unión a glúcidos y se muestran los residuos de la puerta extracelular (rosado) e intracelular (amarillo) interactuando con la molécula de galactosa en distintas conformaciones. Centroide E es, en teoría, la conformación más estable.

A los centroides seleccionados se les realizó nuevamente análisis estereoquímico con Procheck, resultados que se muestran en la figura 30 para hSGLT2.

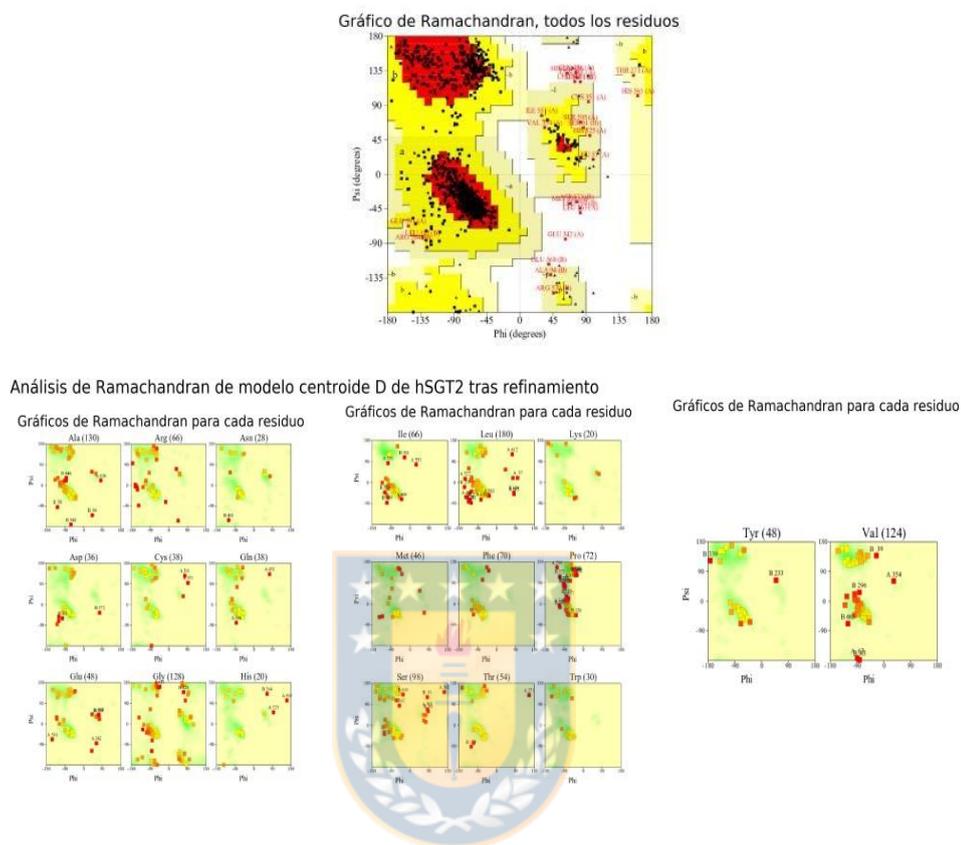


Figura 30: Análisis de Ramachandran de centroide E de hSGLT2 posterior a la equilibración y dinámica

11.8. Creación y caracterización de la biblioteca de inhibidores

La búsqueda de inhibidores es PubChem Compound resultó en 973 moléculas ensayadas *in vivo* en hSGLT2 que califican como “Activas”. Estas moléculas provienen de la optimización de *leads* a partir de los diversos *scaffolds*, incluyendo aquellos fármacos aprobados para su comercialización.

En la Figura 31 se muestran las distribuciones de puntajes de docking asociados a las mejores poses evaluadas por DOCK6 para hSGLT2, junto a una estimación de densidad de kernel (KDE, *Kernel Density Estimation*) y un encaje normal.

Datos de muestreo

- n=973
- I.C. (95%, n=100, iter.=1000) = [-33.697, -30.010]
- Error estándar= 0.922

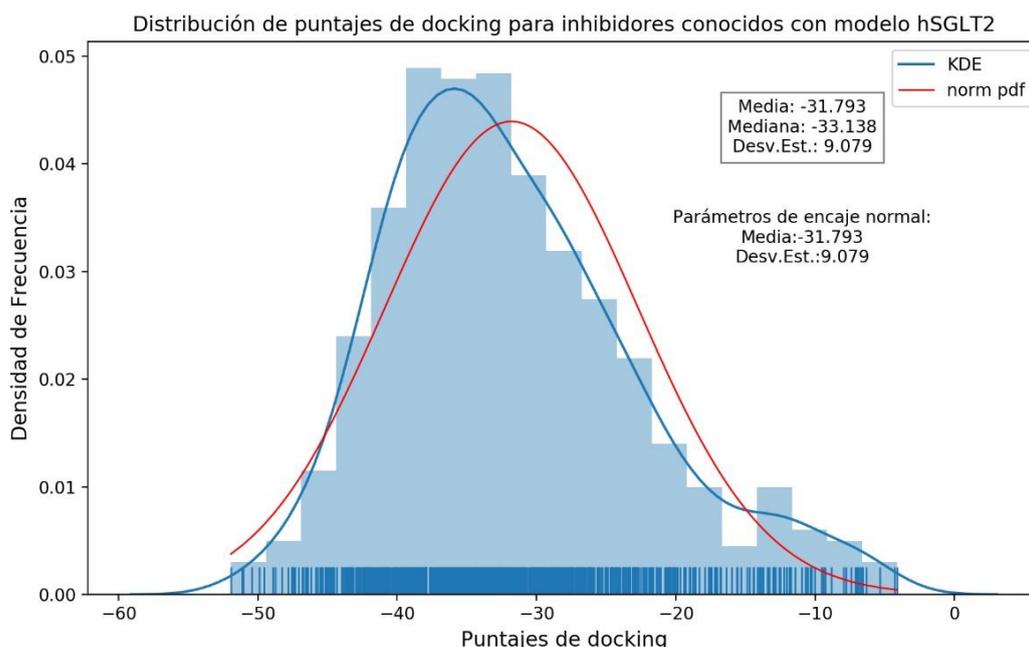


Figura 31: Distribución de puntajes de docking de inhibidores conocidos en docking con hSGLT2

En las Figura 32 y 33 se muestran dos Análisis de Componentes Principales, uno con reducción dimensional a 2D y otro con reducción a 3D, respectivamente, realizados sobre el espacio químico de la base de datos respecto a fingerprints de Morgan de los compuestos. El agrupamiento se realizó usando un índice de Tanimoto de 0.7 como valor de corte y se usó el algoritmo de Butina, específicamente diseñado para que los centroides encontrados sean similares, con el valor de Tanimoto dado, a cada molécula dentro del cluster de modo consistente y automático (Butina 1998).



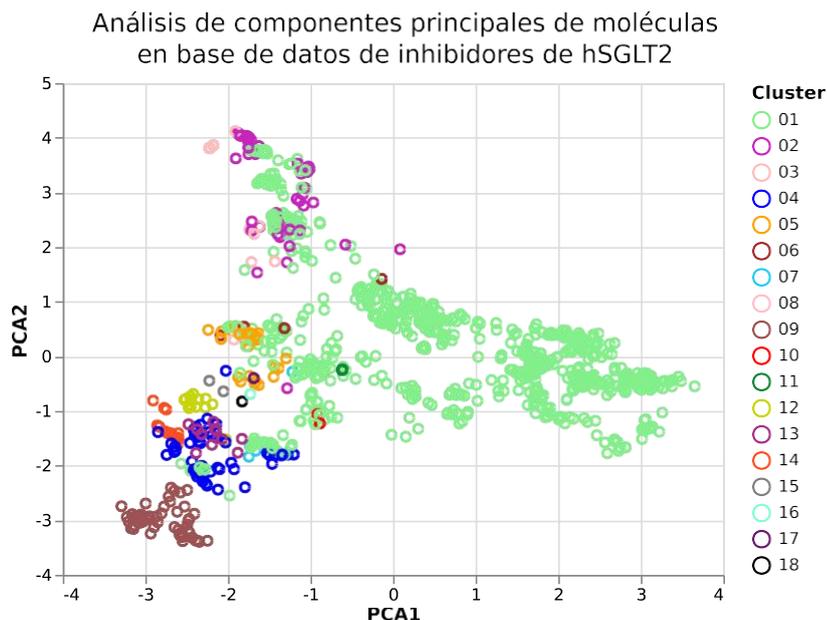


Figura 32: Análisis de componentes principales con reducción dimensional a 2 dimensiones del espacio químico de biblioteca de inhibidores de hSGLT2 a partir de Fingerprints de Morgan. Leyenda corresponde a clusters de índices de similaridad (Tanimoto) generados con agrupamiento de Butina y un valor de corte de 0.7.

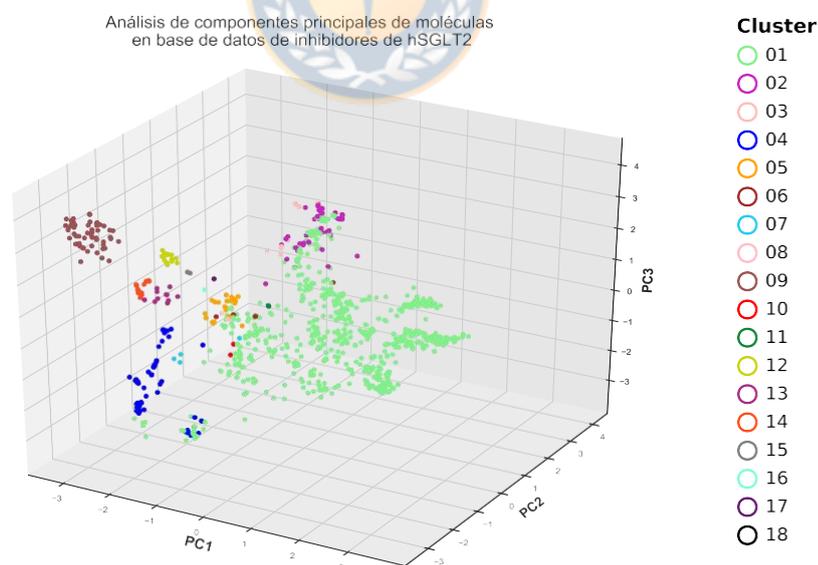


Figura 33: Análisis de componentes principales, con reducción dimensional a 3 dimensiones, del espacio químico de biblioteca de inhibidores de hSGLT2 a partir de Fingerprints de Morgan. Leyenda corresponde a clusters de índices de similaridad (Tanimoto) generados con agrupamiento de Butina y un valor de corte de 0.7.

Las moléculas centroides de los clusters resultantes se muestran en la Figura 34.

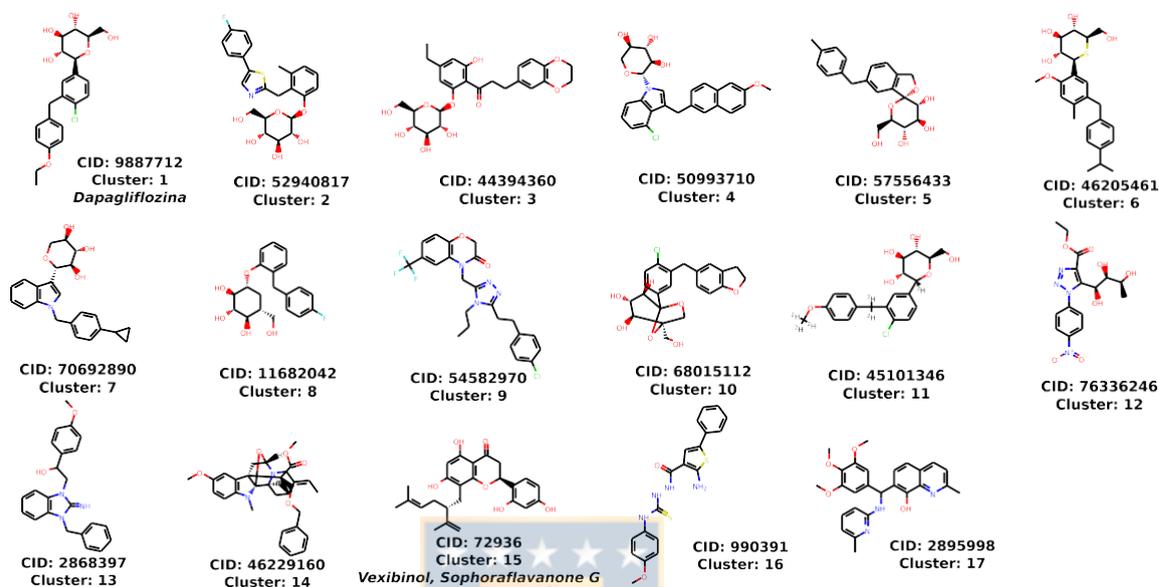


Figura 34: Moléculas centroides resultantes del agrupamiento de Butina de la base de datos de inhibidores de hSGLT2

11.9. Docking masivo

En la Figura 35 se muestra la distribución de los puntajes de docking para la base de datos STK/STL obtenidos con Dock6. A esta muestra se la llamará en adelante 'STK/STL'.

Datos de muestreo

- $n=1,163,733$
- I.C. (95%, $n=100$,iter.=1000) = [4.686, 13.674]
- Error estándar= 2.316

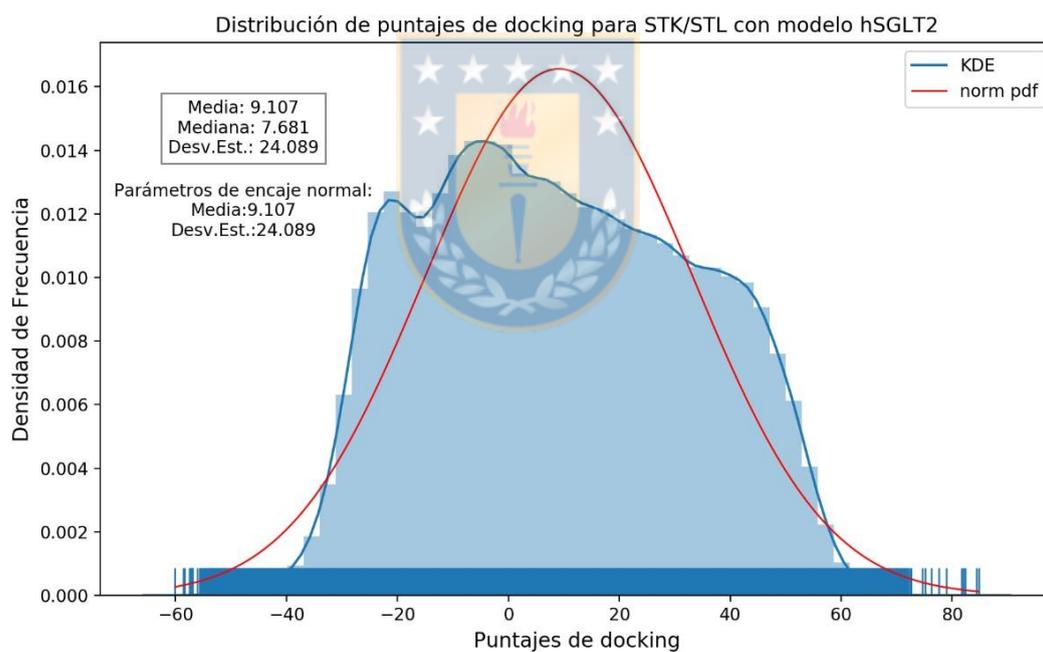


Figura 35: Distribución de puntajes de docking para base de datos STK/STL con Dock6 y modelo de hSGLT2

Desde esta distribución se extrajo el conjunto de moléculas que presentaron un índice de similitud molecular de Dice superior a 0.6 a alguna de las moléculas centroides de la base de datos de inhibidores conocidos de hSGLT2 (a la que se llamará 'Similares', con $n=267,185$). La distribución de puntajes del conjunto de puntajes de 'Similares' se presenta en la Figura 36.

Datos de muestreo

- $n=267,185$
- I.C. (95%, $n=100$, iter.=1000) = [-22.630, -20.728]
- Error estándar = 0.477

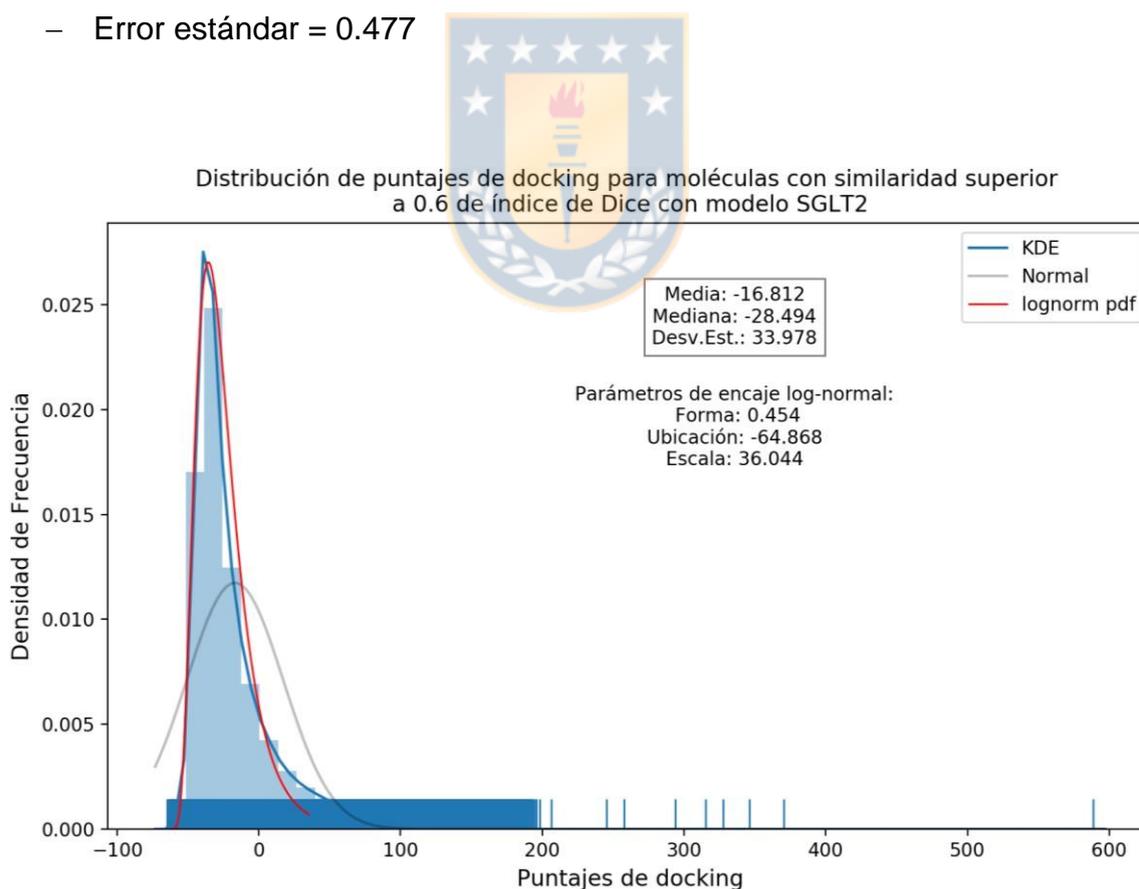


Figura 36: Distribución de puntajes de docking con modelos de hSGLT2 para moléculas con índice de similitud de Dice superior a 0.6

En la Figura 37 se muestran los boxplots para estos grupos de datos sin remover outliers. En tanto, en la Figura 38 se muestran los boxplots sin outliers.

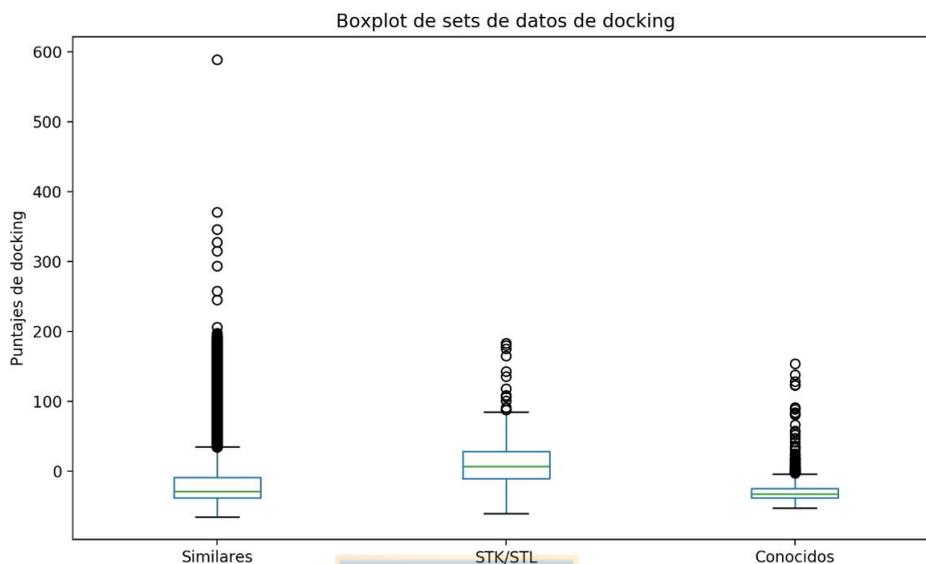


Figura 37: Bloxpot de los conjuntos de puntajes de docking 'Similares', 'STK/STL' y 'Conocidos' con outliers.

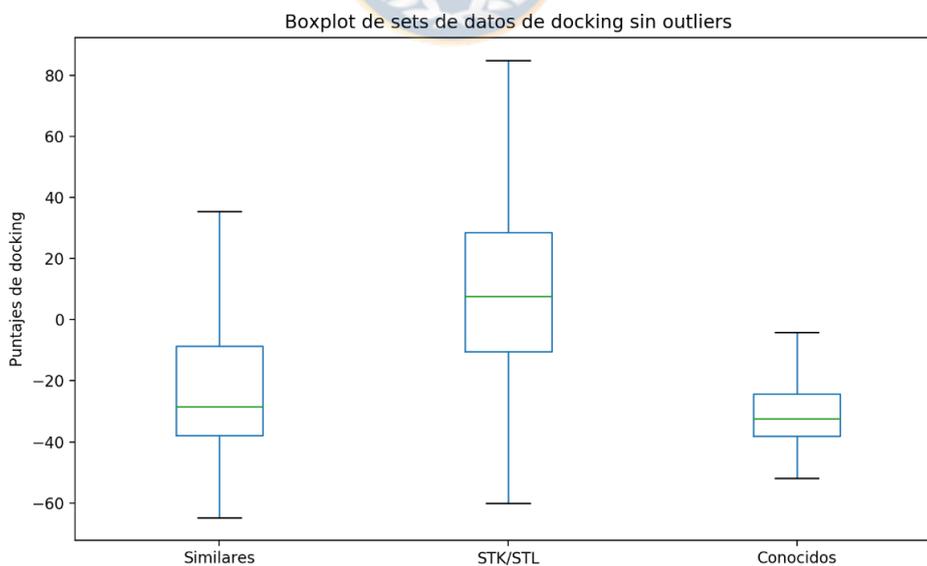


Figura 38: Bloxpot de los conjuntos de puntajes de docking 'Similares', 'STK/STL' y 'Conocidos' sin outliers.

En la Figura 39 se muestran las distribuciones de los 3 conjuntos de datos, a modo de comparación, mientras que en la Figura 40 se presenta una muestra de moléculas idénticas presentes tanto en 'STK/STL' como en 'Conocidos' en la distribución de 'Similares'.

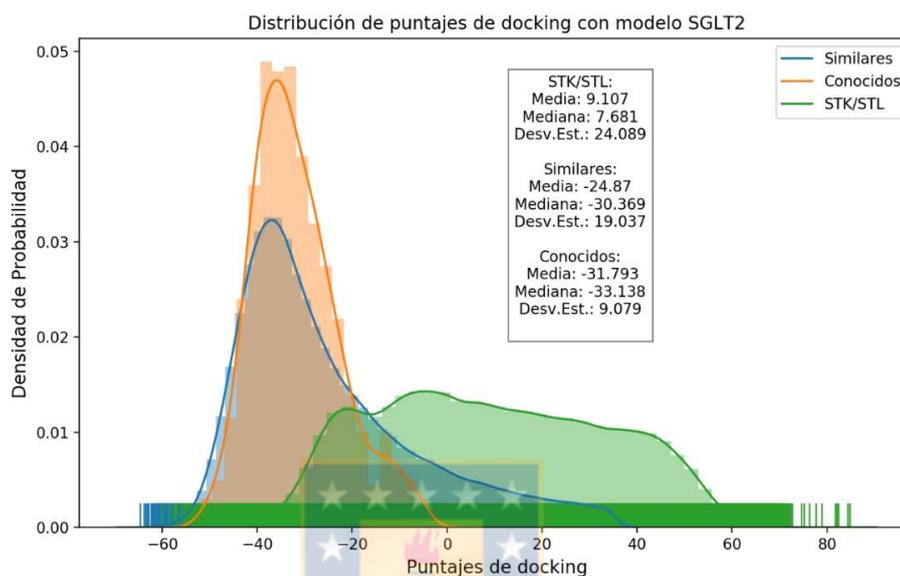


Figura 39: Distribuciones de puntajes de docking de los conjuntos de moléculas 'Similares', 'Conocidos' y 'STK/STL'

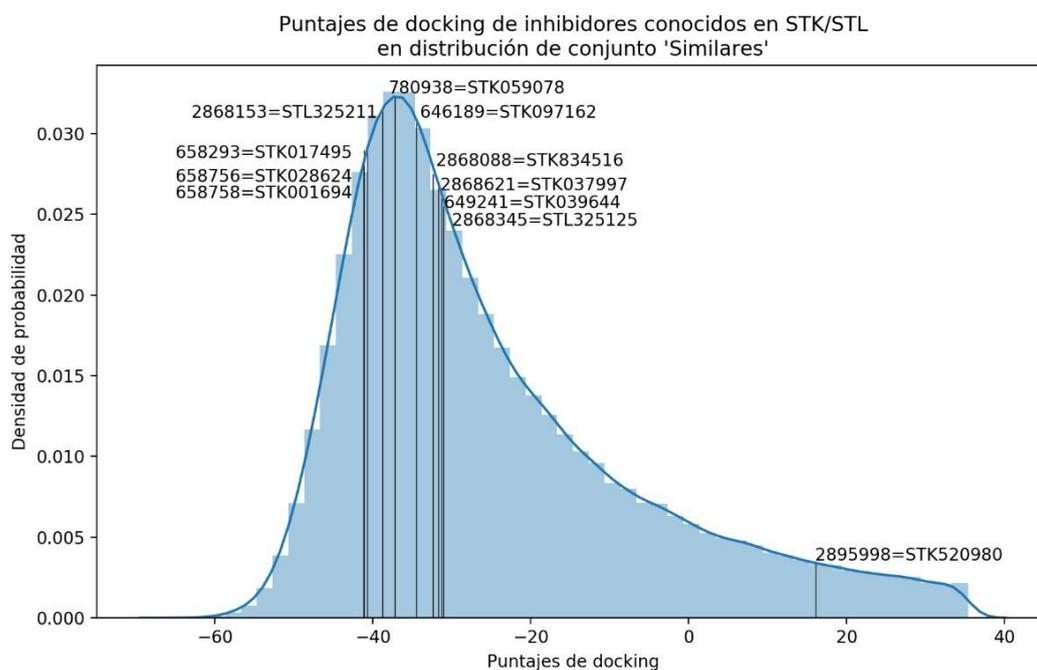


Figura 40: Muestra de moléculas idénticas del conjunto 'Conocidos' en distribución de puntajes de conjunto 'Similares'. Se presentan las moléculas idénticas como "Código PubMed = Código VitasM"



Las 1000 moléculas con mejores puntajes de docking de la base de datos VitasM resultaron formar parte del conjunto 'Similares'. En la Figura 41 se muestran las 16 mejores moléculas y sus puntajes de docking.

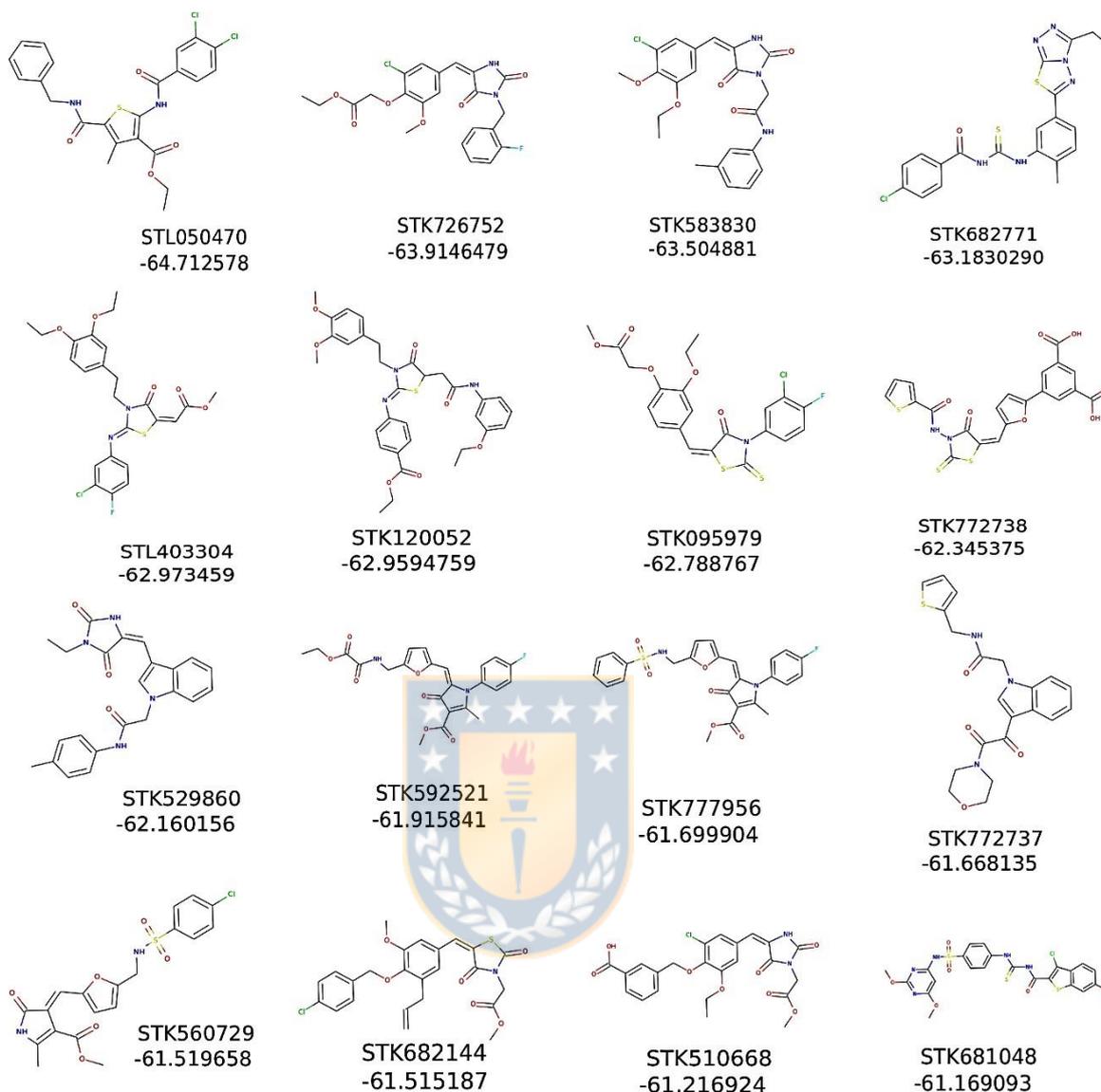


Figura 41: Mejores 16 moléculas respecto a puntaje de docking de VitasM. Se muestra el código de VitasM de la molécula y su puntaje de docking. Ver Figura 34 como comparación.

Debido al número de mediciones de cada muestra, y a causa de la ley de los grandes números, no podemos asegurar que estas muestras provengan de distribuciones poblacionales normales. Al momento de elegir un estadístico apropiado para un test de hipótesis, se debe considerar la normalidad de las distribuciones poblacionales

subyacentes. En las Figuras 42, 43 y 44 se muestran las distribuciones, boxplot y gráficos QQ para los 3 conjuntos de datos. Si bien podemos considerar la normalidad de la población de 'STK/STL' como normal, tal como indica el histograma, boxplot y gráfico QQ de esta, no podemos asegurar lo mismo para los datos 'Similares' y 'Conocidos', que presentan una asimetría negativa (hacia la izquierda). Si, como queremos comprobar en la hipótesis alternativa, estos dos sets de datos presentan diferencias por causas no aleatorias, no podemos asegurar que esta población hipotética sea normal. En consecuencia, las posteriores pruebas de hipótesis elegidas serán pruebas no paramétricas: en primer lugar, una prueba de Kruskal-Wallis para evaluar los 3 sets de datos en conjunto, seguido de la prueba U de Mann-Whitney para los conjuntos de datos individuales.



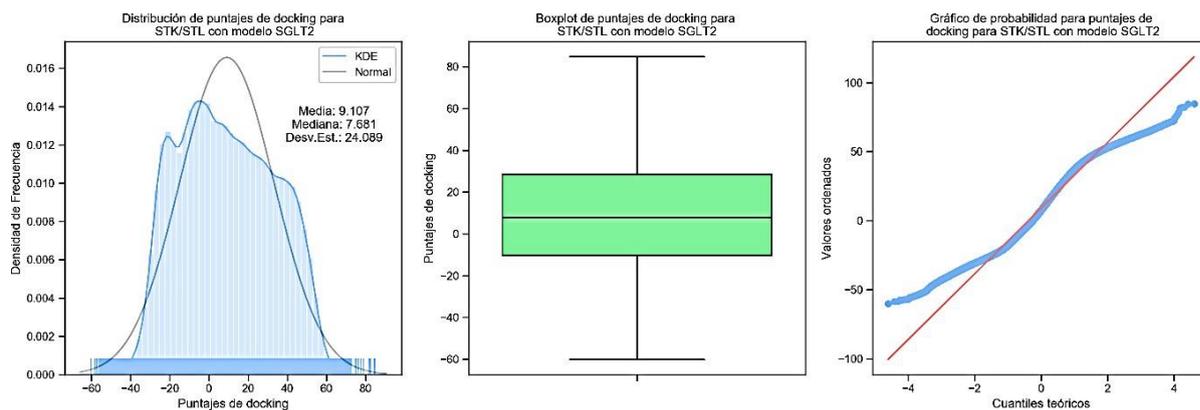


Figura 42: Distribución, boxplot y gráfico QQ de conjunto 'STK/STL'

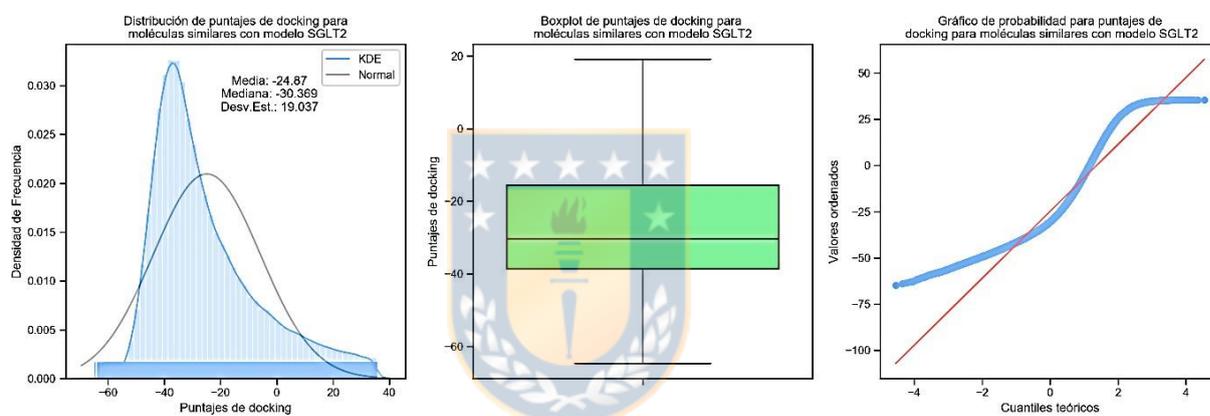


Figura 43: Distribución, boxplot y gráfico QQ de conjunto 'Similares'

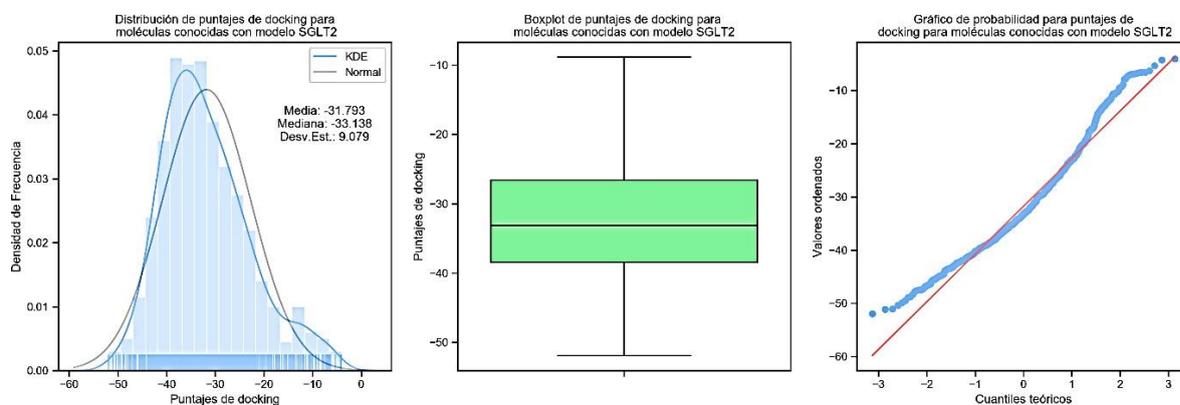


Figura 44: Distribución, boxplot y gráfico QQ de conjunto 'Conocidos'

Para determinar si los datos apoyan la hipótesis de que es posible identificar moléculas inhibidoras con métodos bioinformáticos y estadísticos, realizamos las pruebas estadísticas especificadas en Materiales y Métodos sobre los datos tabulados. En vez de usar la estadística inferencial, usamos conceptos utilizados en la estadística de estimación (informalmente llamada “nueva estadística”), cuya preocupación no es la de derivar conclusiones sobre la población subyacente, sino en comparar tamaños de efecto entre observaciones (esto es, la magnitud cuantitativa de un fenómeno observado) y su precisión mediante un intervalo de confianza (Cumming y Calin-Jageman 2016). Para nuestro caso, usamos un estimador adimensional conocido como *d de Cohen*, el cual reporta los tamaños de efecto de modo comparable entre estudios, y es usado como alternativa a los test de hipótesis, entregando información más relevante sobre las observaciones. Los valores resultantes pueden ser interpretados siguiendo la Tabla VIII (Cohen 1988, Sawilowsky 2009)

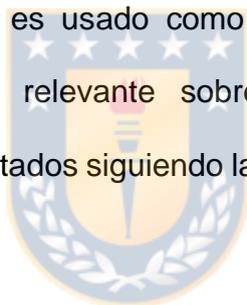


Tabla VIII: Interpretación de tamaño del efecto de *d* de Cohen

Tamaño del efecto	<i>d</i>
Muy pequeño	0.01
Pequeño	0.20
Medio	0.50
Grande	0.80
Muy grande	1.20
Enorme	2.00

STK/STL y Conocidos

- Superposición:

$$Superposición_{STK/STL} = 12.58 \%$$

$$Superposición_{Conocidos} = 11.81\%$$

- Tasa de error:

$$Tasa de error = 12.19\%$$

- Tamaño de efecto de Cohen:

$$d = 1.699$$

STK/STL y Similares

- Superposición:

$$Superposición_{STK/STL} = 15.8\%$$

$$Superposición_{Similares} = 27.99 \%$$

- Tasa de error:

$$Tasa de error = 21.9\%$$

- Tamaño de efecto de Cohen:

$$d = 1.536$$

Conocidos y similares

- Superposición:

$$Superposición_{Similares} = 99.95\%$$

$$Superposición_{Conocidos} = 93.97 \%$$

- Tasa de error:

$$Tasa\ de\ error = 96.96\%$$

- Tamaño de efecto de Cohen:

$$d = 0.3641$$

Test de hipótesis

1. Establecer hipótesis

H_0 : Las muestras provienen de poblaciones con la misma distribución

H_a : Las muestras provienen de poblaciones diferentes

2. Tasa de error Tipo I: $\alpha = 0.05$

3. Calcular estadístico de prueba:

El resultado de la prueba de Kruskal-Wallis es el siguiente:

```
KruskalResult(statistic=162541.69441621192, pvalue=0.0)
```

4. Resultado

Se rechaza la hipótesis nula. Los datos apoyan la hipótesis de que las muestras provienen de poblaciones diferentes.

El estadístico de la prueba de Kruskal-Wallis, H , es sólo una distribución χ^2 . En este caso, para una distribución χ^2 con 2 grados de libertad, el valor crítico es de 5.9914645 al 0.95 de confianza.

STK/STL vs Similar

1. Establecer hipótesis

H_0 : $Me(STK/STL) \leq Me(Similar)$

H_a : $Me(STK/STL) > Me(Similar)$

2. Tasa de error Tipo I: $\alpha = 0.05$

3. Calcular estadístico de prueba:

El resultado de la prueba de U de Mann-Whitney de una cola es el siguiente:

```
MannwhitneyuResult(statistic=72800778893.5, pvalue=0.0)
```

4. Resultado

Se rechaza la hipótesis nula. Los datos apoyan la hipótesis de que la muestra 'STK/STL' proviene de una población de valores significativamente mayores a la muestra 'Similar'.

STK/STL vs Conocidos

1. Establecer hipótesis

$$H_0: Me(STK/STL) \leq Me(Conocidos)$$

$$H_a: Me(STK/STL) > Me(Conocidos)$$



2. Tasa de error Tipo I: $\alpha = 0.05$

3. Calcular estadístico de prueba:

El resultado de la prueba de U de Mann-Whitney de una cola es el siguiente:

```
MannwhitneyuResult(statistic=266185740.0, pvalue=0.0)
```

4. Resultado

Se rechaza la hipótesis nula. Los datos apoyan la hipótesis de que la muestra 'STK/STL' proviene de una población de valores significativamente mayores a la muestra 'Conocidos'.

Conocidos vs Similares

1. Establecer hipótesis

$$H_0: Me(Similares) = Me(Conocidos)$$

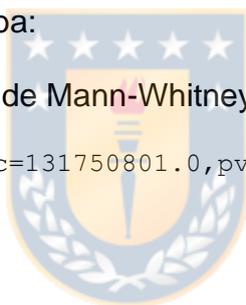
$$H_a: Me(Similares) \neq Me(Conocidos)$$

2. Tasa de error Tipo I: $\alpha = 0.05$

3. Calcular estadístico de prueba:

El resultado de la prueba de U de Mann-Whitney de dos colas es el siguiente:

MannwhitneyuResult (statistic=131750801.0, pvalue=9.877652661735354e-14)



4. Resultado

Se rechaza la hipótesis nula. Los datos no apoyan la hipótesis de que el conjunto de datos 'Similares' y los 'Conocidos' provengan de la misma distribución poblacional.

Más aun, con las hipótesis:

$$H_0: Me(Conocidos) \geq Me(Similares)$$

$$H_a: Me(Conocidos) < Me(Similares)$$

y $\alpha=0.05$. Resulta en:

MannwhitneyuResult (statistic=98028299.0, pvalue=4.938826330867677e-14)

12. DISCUSIÓN

12.1. Discusión preliminar

Nuestro trabajo se orienta hacia el desarrollo de un método robusto para la identificación estadística de moléculas con acción sobre el co-transportador de sodio:glucosa humano 2, hSGLT2, a partir de métodos de biología computacional. Este desarrollo lleva consigo una exploración y prueba de los métodos disponibles para realizar un barrido eficiente sobre los factores que puedan dificultar la identificación. La necesidad del desarrollo de una metodología de este tipo no se basa en la comprobación de la eficacia y eficiencia de los métodos *per se*, que han sido ampliamente discutidos y probados en la literatura científica (Xia 2017, Kontoyianni 2017, Cavasotto 2011, Rockey y Elcock 2006), sino en su uso eficaz sobre las características específicas de la proteína de estudio y de la historia investigativa de su inhibición. Nuestra tesis, explorando varias de las aproximaciones teóricas disponibles en cada paso de su metodología e implementando algunas propuestas metodológicas innovadoras, ha encontrado relaciones estadísticas que permiten concluir que los métodos actuales de biología computacional son capaces de reconocer los inhibidores conocidos y sus análogos estructurales.

Así, en esta tesis se exploran las técnicas disponibles para generar una metodología robusta y reproducible de reconocimiento de fármacos inhibidores de hSGLT2 como un marco general, que incluso puede extenderse potencialmente a otras proteínas de membrana si se aplican las consideraciones metodológicas, y en vistas de expandir los espacios químicos de inhibición de las proteínas transportadoras de membrana,

con la idea proyectada de abaratar los costes en tiempo y dinero de producción de análogos de inhibidores de transportadores transmembrana.

hSGLT2 se trata, en primera instancia, de una proteína de transporte transmembrana, las cuales han demostrado ser difíciles de dilucidar estructuralmente, como se mencionó en la introducción. En segundo lugar, la estructura misma predicha para los miembros de la familia APC, a la que hSGLT2 pertenece, posee características estructurales que vuelven su modelado más complejo respecto a otros transportadores de membrana, y hSGLT2 mismo presenta peculiaridades particulares en su forma tridimensional: la presencia de discontinuidades funcionalmente importantes en algunas de sus hélices transmembrana (Screpanti y Hunte 2007), los reportes experimentales de puentes disulfuro en la estructura tanto intrasubunitarios como intersubunitarios (Gagnon *et al.* 2007, Gagnon *et al.* 2006) y de un lazo re-entrante al plano de la membrana que puede tener un rol en inhibición (Sasseville *et al.* 2016, Gagnon *et al.* 2005), aparte de diferencias en estructura cuaternaria respecto a las estructuras homólogas conocidas (Sasseville *et al.* 2016), vuelven el correcto modelado de este transportador un desafío.

En vista de esto, las consideraciones estructurales de hSGLT2 mencionadas anteriormente no son tomadas en cuenta si no influyen en gran medida en la interacción del sitio de unión con su sustrato. Aunque se exploraron modelos con puentes de cisteína y lazo reentrante, la incertidumbre actual sobre la estructura cuaternaria de hSGLT2 respecto a los moldes de vSGLT cristalizados y a las investigaciones sobre hSGLT1, no permiten predecir la forma correcta en que el dímero debe ensamblarse y resultados de docking proteína-proteína realizados con el programa MegaDock no fueron apropiados en el sentido de no poner en cercanía suficiente los residuos de cisteína 355 de cada subunidad que producen el enlace

disulfuro intercadena, ni siquiera modificando la posición de los lazos en que se encuentran con el programa Modeller. Pese a que esto no tiene relevancia respecto a la inhibición del sitio de unión del sustrato, ya que se ha demostrado que la proteína monomérica posee actividad independiente (Wright *et al.* 2011) y como se discutirá luego, se reprodujeron características que demuestran funcionalidad en el modelo, un estudio refinado de la potencia de los inhibidores de hSGLT2 debe tomar en cuenta la cooperación entre subunidades y la influencia de restricciones estructurales como puentes disulfuro.

Estas consideraciones estructurales tampoco han sido tomadas en cuenta en estudios que también han realizado modelados estructurales de hSGLT2 similares a esta tesis, como en Dong *et al.* 2018, donde modelos monoméricos basados en homología y basados en análisis de campos moleculares comparativos (*comparative molecular field analysis*, CoMFA) y análisis de índices de similaridad molecular comparativos (*comparative molecular similarity indices analysis*, CoMSIA) fueron diseñados para evaluar las relaciones estructura-actividad cuantitativas (QSAR) entre hSGLT2 y sus mecanismos de interacción, dando resultados en que las interacciones electrostáticas e hidrofóbicas que mejoran la actividad coinciden con los resultados de los métodos de docking. Tampoco se tomaron en cuenta en Xu *et al.* 2015, donde modelados por homología de monómeros y evaluaciones de energía libre mediante dinámica molecular se realizaron sobre SGLT1 y SGLT2 para evaluar la selectividad de canagliflozina. Ejemplos del mismo tipo se hallan en Ng, Wai-Lung *et al.* 2017, Yu *et al.* 2017, Tamura *et al.* 2015 y Vorberg *et al.* 2012, entre otros estudios. Por tanto, determinamos que no es esencial para el desarrollo del modelo tomar esto en cuenta, aunque debido a los efectos reportados sobre la inhibición sobre el loop reentrante

(Sasseville *et al.* 2016), un modelo adecuado debería incluirlos, en especial para estudios de selectividad sobre SGLT1 y SGLT2.

12.2. Búsqueda de homólogos en bases de datos

Las secuencias de los 12 miembros de la familia SSS y sus homólogos se obtuvieron desde SwissProt, ya que se trata de una base de datos de secuencias curadas y se desean evitar las secuencias con errores o incompletas. Para la búsqueda se tomó en consideración que se reportan en literatura la ocurrencia frecuente de falsos positivos en los resultados producidos por programas de búsqueda en bases de datos de secuencias (Schäffer *et al.* 2001), en especial los que pueden ser causados por métodos iterativos que producen matrices de puntaje específicas de posición (*position-specific scoring matrix*, PSSM), como PSI-BLAST, o modelos de Markov (Jackhmmer), a causa de que la detección de una secuencia falsa positiva altera la composición de las matrices producidas, propagando el error a los siguientes ciclos. Se usaron los programas y parámetros listados en la Tabla III, eligiendo aquellos que tentativamente permitiesen una detección más amplia de secuencias. Para BLASTp y PSI-BLAST se usaron matrices BLOSUM62 y BLOSUM45 con distintos E-values, 0.01 y 0.001, para entregar distintos grados de estimación de distancia evolutiva entre secuencias. Con PSI-SEARCH, al ser un programa no heurístico como BLAST y ser iterativo (y por tanto, con menos probabilidad de entregar errores), se hizo más estricta la búsqueda, con una matriz BLOSUM45 y E-value 0.0001. Con Jackhmmer se usó un E-value de 0.001. Finalmente, para SSEARCH, se usó la matriz asimétrica SLIM (Müller *et al.* 2001), diseñada para proteínas de membrana, con E-values 0.1 y 0.01, a fin de volver la búsqueda de este método no iterativo más específico.

Como se mencionó, métodos iterativos a veces pueden causar errores que se propagan. Esto se hizo evidente en los resultados de Jackhmmer, donde la búsqueda en Swissprot para la familia SSS retornó 674 secuencias, debido a que una de las secuencias, hSGLT5, incorporó en una de sus iteraciones secuencias no relacionadas. Estas secuencias no relacionadas pertenecían a la subfamilia MraY de la familia de glicosiltransferasas 4 de bacterias, GT (4), necesaria para la biosíntesis de la pared celular. En TCDB, mientras la familia SSS se clasifica como 2.A, la familia GT (4) o MurG se clasifica como 9.B. Es decir, la relación entre las secuencias es decididamente inexistente y se trata de resultados falsos positivos.

A partir de esto, se realizó un criterio de selección de secuencias por consenso, en donde las secuencias comunes a tres o más de los resultados son incorporadas a los resultados finales, mientras que si otras secuencias aparecen dos o más veces en uno de los métodos iterativos también son seleccionadas. Esta diferenciación es para darle más relevancia a la mejor detección de homólogos distantes que ofrecen los métodos iterativos. Hasta donde se investigó en esta tesis, esta aproximación para evitar falsos positivos en resultados de búsqueda no se ha usado previamente y extiende la metodología actual para dar resultados de homología más exactos.

El resultado final fue de 113 secuencias que incluye a miembros representativos de la familia SSS en vertebrados y homólogos distantes bacterianos de transportadores de urea, pantotenato, prolina, acetato y el homólogo vSGLT de *Vibrio parahaemolyticus*. Un refinado posterior para eliminar redundancias de secuencias similares con CD-HIT90 resultó en 76 secuencias.

12.3. Alineamiento múltiple de secuencias

La etapa del alineamiento múltiple resulta crucial para los pasos posteriores de filogenia (Wang *et al.* 2011) y modelado tridimensional. Dado que nuestra aproximación al modelado de la estructura de hSGLT2 se realizará a partir de la hipótesis que las estructuras terciarias son altamente conservadas a través de la evolución, es decir, que se usarán métodos filogenéticos para guiar el modelado, los resultados de los alineamientos deben ser óptimos y deben reflejar la conservación de residuos críticos, fenómeno denominado consistencia. Esta aproximación ha sido usada con éxito en estudios no limitados a proteínas de membrana como, por ejemplo, en Satpathy *et al.* 2016, donde se usó para modelar enzimas deshalogenasas de origen bacteriano y fúngico.

En la Figura 45 se muestra parte del alineamiento y árbol filogenético realizado con POA, donde se observa que algunas de las hélices predichas por TOPCONS se extienden a través del alineamiento debido a la inserción de aperturas. Patrones similares se observaron en otros resultados de alineamiento para otros segmentos transmembrana (STM). El árbol filogenético con MAFFT G-INS-i y matriz SLIM de la Figura 20 muestra cómo este fue capaz de alinear los segmentos transmembrana de una forma adecuada y consistente con lo reportado en literatura (Wright *et al.* 2011, Wright y Turk 2004, Vastermark *et al.* 2014)

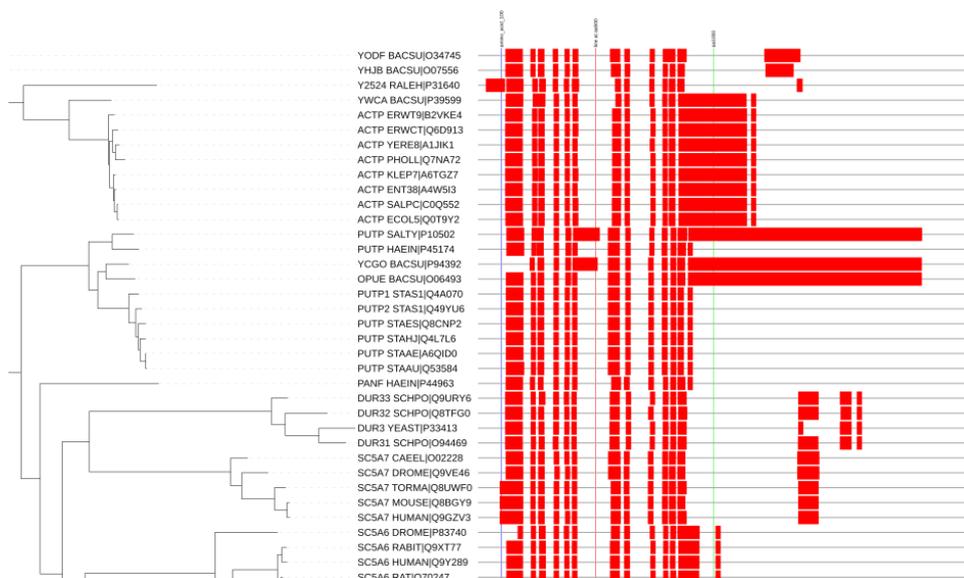


Figura 45: Alineamiento de secuencias con POA.

Se observa extensión de los segmentos transmembrana predichos por TOPCONS

Tanto en el caso de POA como de los otros alineamientos descartados, se observa que los STM que se hallan expandidos corresponden a las hélices adicionales particulares de cada proteína, mientras que los STM de los dominios de andamiaje (TM2, TM3, TM7 y TM8) y hash (TM4, TM5, TM9 y TM10) se hallan mejor alineados en general.

Otro criterio usado se guió por el número de sitios conservados que el programa de alineamiento pudo detectar, detallados en la Tabla IV. Como se observa en ésta, MAFFT G-INS-i con SLIM pudo detectar un sitio conservado adicional (sin contar primer residuo de metionina N-terminal común a todas las secuencias). Si bien ProbCons encontró 26 sitios sobre 80% de conservación, frente a los 25 de MAFFT G-INS-i con SLIM, la mayor longitud, porcentaje de aperturas y alineamiento de STM del primero lo vuelven menos favorable para su selección.

Si bien MAFFT G-INS-i con SLIM no es el óptimo en términos globales, se halla de igual modo entre los de valor inferior a la media, como muestra una simple visualización de los datos en la Figura 46, lo que sostiene la elección de los resultados de este método como el indicado para los pasos posteriores.

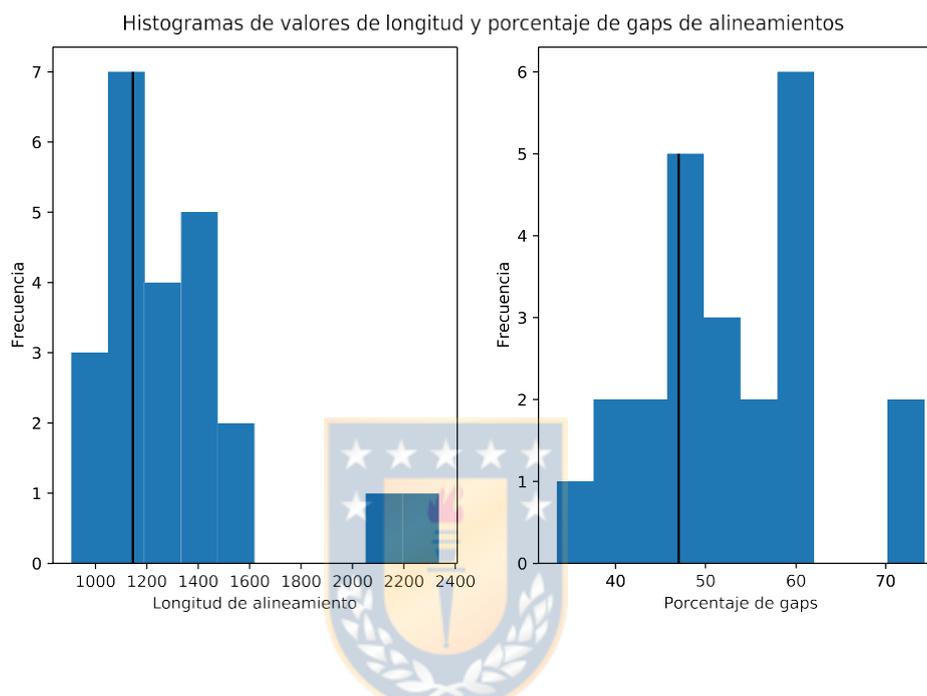


Figura 46: Histograma de longitud de alineamientos y porcentaje de gaps para los programas de alineamiento múltiple empleados. Con una línea negra se indican los valores obtenidos por MAFFT G-INS-i con SLIM

En resumen, la elección del método de alineamiento apropiado es un balance entre las características locales de la proteína misma, en este caso del adecuado alineamiento de los STM que la componen (o de otro dominio, si tal es el caso) y los resultados globales del alineamiento, su longitud y número de gaps, y otros indicadores como puntajes de consistencia o entropía. Se beneficia además del conocimiento de las características conservadas de la familia de la proteína de interés.

12.4. Filogenia

Los métodos de máxima verosimilitud intentan realizar un ajuste estadístico de los datos, en este caso los sitios de los alineamientos, como extraídos desde una función de probabilidad continua, usualmente una distribución gamma con iguales parámetros de escala α y de forma β , bajo un modelo evolutivo de sustitución, maximizando el valor de la función de verosimilitud asociada. A mayor valor de probabilidad de las secuencias, dado el árbol bajo cierto modelo, más el árbol es preferido (Yang 2014, Stamatakis 2014)

Esta aproximación, una extensión del método de parsimonia con incorporación de modelos probabilísticos basados en modelos de Markov para evaluación de los árboles, permite evitar errores comunes de los métodos de parsimonia como la conocida atracción de ramas largas que se da entre secuencias altamente divergentes; es decir, los métodos de máxima verosimilitud (junto con los métodos bayesianos) son los más apropiados para tratar con secuencias que se encuentran a mayores distancias evolutivas, como es el caso de las secuencias seleccionadas. Una desventaja de estos métodos se puede atribuir a su exhaustividad: un gran número de árboles debe ser evaluado antes de que el óptimo sea elegido, que pese a la existencia de métodos heurísticos como cambios de rama (*branch swapping*), sigue siendo computacionalmente caro en términos de tiempo (Yang 2014).

El uso de modelos evolutivos permite la elección del que sea más adecuado para las secuencias seleccionadas. La elección del modelo depende de la distancia evolutiva entre secuencias. Aquí, se permitió que RAxML eligiera el modelo apropiado para los alineamientos usando un criterio de información bayesiana (BIC, *Bayesian Information Criterion*), el cual penaliza más la introducción de nuevos parámetros que otros

criterios como el de Akaike (AIC, *Akaike Information Criterion*) o test de razón de verosimilitud (LRT, *Likelihood Ratio Test*), evitando sobreajuste de los datos (Yang 2004, Schwarz 1978).

Para todos los alineamientos, RAxML eligió el modelo LG con frecuencias empíricas de aminoácidos (modelo LG+F), que es el que tiene mayor número de parámetros para la matriz de tasas de sustitución del modelo de Markov subyacente. El modelo LG es una versión actualizada de la matriz WAG (Le y Gascuel 2008), la cual se produjo desde un modelo de máxima verosimilitud de alineamientos de entrenamiento. Esto refleja la divergencia evolutiva de las secuencias, dado que para alineamientos con secuencias evolutivamente próximas, se seleccionarían modelos más simples, con menor número de parámetros y menor tiempo de cálculo.

Los puntajes de verosimilitud resultantes varían entre los -5200 y -5400, indicando una consistencia general en los alineamientos, con sólo algunos de los programas, en particular los que usan métodos probabilísticos como ProbCons y MSAProbs, entregando peores valores que el promedio. Esta falla de los métodos probabilísticos también se observa en las diferencias en los valores de parámetro alfa de la función gamma (parámetro de forma) y de longitud de árbol, como se observa en la Tabla VI, así como en el alineamiento de los segmentos transmembrana, discutidos en la sección anterior. El alineamiento seleccionado, MAFFT G-INS-i con matriz SLIM, se encuentra entre los de mejor puntaje de verosimilitud, si bien no es el mejor. Su elección, como ya se discutió, se debe a su preferente alineamiento de segmentos transmembrana y también tomando en cuenta evitar el riesgo de sobreencaje de datos, que es un problema frecuente de todo método de máxima verosimilitud.

12.5. Modelado comparativo

Mientras que se eligió finalmente el modelo desarrollado con MODELLER, otros métodos fueron probados, con resultados insatisfactorios, muy posiblemente debido a estar diseñados para proteínas globulares. I-TASSER, por ejemplo, un método de modelado por hilado (*threading*), no logró reproducir la posición de los lazos y el correcto plegamiento de las hélices transmembrana. El mismo caso se produjo con SWISS-MODEL. Tampoco RAPTOR-X, un servidor de modelado mediante redes neuronales, entregó resultados satisfactorios, deformando la proteína resultante.

En tanto, se realizó un modelado comparativo con MODELLER basado en el árbol filogenético obtenido anteriormente, siguiendo la longitud relativa de las ramas respecto a la rama del molde 3DH4. Con esto, se partió modelando hSGLT1 desde 3DH4, seguido de hSGLT3 desde el modelo generado de hSGLT1 y finalmente hSGLT2 desde modelo de hSGLT3. Este orden se reafirma con los porcentajes de identidad indicados en la Introducción.

La evaluación de los modelos se realizó primero mediante una inspección visual, buscando que no hubiese errores en el modelado, como lazos cruzados o rupturas. La elección del modelo se basó en el puntaje DOPE entregado por MODELLER, que si bien no es el más adecuado para proteínas de membrana, entrega una evaluación general de su estabilidad energética según potenciales estadísticos (Shen y Sali 2006) y en el análisis de Ramachandran tanto antes como después de la dinámica molecular. Un punto importante en el modelado es encontrar la correspondencia de los residuos funcionalmente importantes, conocidos experimentalmente, con el modelo generado. En la Figura 23 se ilustra esta correspondencia, que fue usada para definir el posterior sitio sobre el que se realizaría el docking.

12.6. Dinámica Molecular

Durante los 50 ns de dinámica molecular, el modelo de hSGLT2 mostró transiciones conformacionales desde el estado exofacial-ocluído, original del molde 3DH4, hacia estados ocluidos. Esto se ve en la Figura 23, donde el residuo de galactosa en centroide E de hSGLT2 se encuentra formando interacciones de apilamiento entre los residuos aromáticos de la puerta extracelular y los de la puerta intracelular, configuración característica de esta conformación.

Estos cambios conformacionales permitieron el establecimiento de grupos de estructuras basados en RMSD, cada uno con distinto número de pasos, que representan distintos volúmenes del espacio de fase conformacional del sistema, incluido el espacio de fase de la proteína restringida. De estos, los más numerosos son considerados los más estables energéticamente tomando en consideración el principio ergódico, en que el tiempo que pasa el sistema en cierta región de su espacio de fase es equivalente al volumen de este; es decir, el cluster más probable es aquel que contiene aquellas conformaciones que representan el estado de equilibrio de la proteína, donde cada una de estas conformaciones es equiprobable dentro del espacio de fase (Tuckerman 2010).

En las Figura 30 se presentan los gráficos de Ramachandran para los centroides de hSGLT2 seleccionados. Se puede observar que la gran mayoría de los residuos se ubica dentro de las zonas conformacionales permitidas. Pese a que algunos de ellos se ubican fuera de éstas, es una observación normal dentro de las variaciones angulares que pueden producirse a causa de las interacciones con otros elementos del sistema. Es notoria la presencia de un grupo de cinco residuos -no glicina ni prolina- de la cadena A de hSGLT2 que se ubican fuera de las zonas permitidas. Esto

se produce a causa de defectos del modelado de la estructura cuaternaria que se discutieron anteriormente, que provoca una inserción indebida en la membrana. Por ello, y dado que no se ve afectada la función del transportador si se trata de una subunidad aislada, la cadena A fue descartada y se usó la cadena B para los procedimientos posteriores de docking.

Esta subunidad B, además, logró replicar los determinantes estructurales de poro de agua descritos en Adelman *et al.* 2014, ilustrados en la Figura 28. Se observa la formación de dos canales de agua desde el vestíbulo extracelular, los que conducen los átomos de sodio y la glucosa a sus sitios de unión de manera específica. Luego, el canal es obstruido por los residuos de la puerta intracelular en las conformaciones ocluidas y se conecta con el vestíbulo extracelular cuando el transportador cambia de conformación para permitir el paso de los solutos.

12.7. Biblioteca de diversidad de inhibidores de hSGLT2

Tres aproximaciones principales pueden tomarse para seleccionar y clasificar compuestos diversos con el objetivo de encontrar nuevos *leads*: Análisis de agrupamiento, métodos basados en disimilitud y métodos basados en partición (Leach 2001). Ya que nuestra aproximación se basó en usar los compuestos que ya han sido reconocidos como inhibidores, clasificados como “Activos” en PubChem, los cuales son ya de por sí diversos, el uso de descriptores individuales para la búsqueda de compuestos adicionales no reflejaría la diversidad de los subconjuntos de moléculas en el espacio, por lo que usamos un método de agrupamiento y análisis de componentes principales para confirmar que son capaces de diferenciar las distintas clases de moléculas seleccionadas y usar los centroides de cada uno de estos clusters

como modelos para buscar similaridad molecular en el espacio químico de interés. Esto se demuestra en el cluster de las Figuras 33 y 34, donde se observan distintos subconjuntos de moléculas, cada una con características estructurales y fisicoquímicas diferentes, derivadas desde los *fingerprints* de Morgan de cada molécula y la agrupación de estos en clusters. La mayoría de estas se corresponden con el paradigma farmacológico moderno de inhibición de hSGLT2, enfocada en C-arilglucósidos como canagliflozina y dapagliflozina (esta misma un centroide), que se ven representados en el grupo 01. Desde estos es posible estimar la correlación entre descriptores usando coeficiente correlación de Pearson, gráfico de Craig o Análisis de Componentes Principales (Leach 2001), la última de las cuales fue usada aquí. Desde los grupos se seleccionaron los centroides representativos de las especies químicas de cada uno de ellos y se utilizaron luego para buscar moléculas con similaridad molecular en la base de datos VitasM (STK/STL).

12.8. Docking masivo con Dock6 e identificación de moléculas similares

Las moléculas de inhibidores conocidos y las presentes en el conjunto VitasM no coinciden en su totalidad. De las 973 moléculas conocidas, sólo 11 están presentes en las 1,371,136 de moléculas de VitasM (Índice de Dice=1.0), las que se muestran en la Figura 40. Sin embargo, existen 267,185 análogos estructurales a los centroides, la mayoría de los cuales presentan puntajes de docking similares a la de los inhibidores conocidos, presentando una elevada superposición en sus distribuciones. Particularmente, todas las 1000 moléculas con mejores puntajes de docking de VitasM pertenecen al conjunto de moléculas estructuralmente similares, con puntajes que

rondan -64.712578 para STL050470 a -53.7027180 para STK873328, como se muestra en la Figura 41.

12.9. Evaluación estadística de observaciones de docking

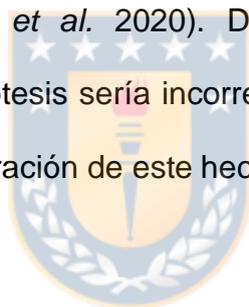
El espacio muestral de donde se obtienen las muestras 'STK/STL', 'Conocidos' y 'Similares' corresponde al espacio de puntajes de docking del espacio químico de moléculas orgánicas. En otras palabras, teniendo el espacio químico de todas las moléculas orgánicas, si a estas se les aplica una función que le asigna un puntaje de docking óptimo en el sitio de unión de hSGLT2, el espacio muestral es el rango de valores de docking del que se extraen las muestras, que es la imagen de esta función.

El rango de valores teórico puede extenderse desde $-\infty$ (por ejemplo, una molécula teórica con afinidad tan fuerte que la unión sea en todo término práctico irreversible) a $+\infty$ (p.ej. una molécula demasiado grande para encajar en el sitio de unión).

El muestreo sobre la base de datos de moléculas VitasM, o 'STK/STL', corresponde a un muestreo aleatorio, donde cada molécula de la muestra posee la misma probabilidad de ser escogida desde el espacio muestral. Este no es el caso, sin embargo, de las muestras 'Conocidas' y 'Similares'. En ambos casos, se introduce un sesgo en la recolección de la muestra. Para 'Conocidos', la muestra control, este sesgo se debe a un conocimiento previo de los puntajes de docking posibles (prevalentemente negativos). Para 'Similares', el sesgo se introduce tras aplicar el criterio de similaridad.

Pareciera tentador usar pruebas de hipótesis para determinar si estas muestras provienen de poblaciones disjuntas, pero esta aproximación sería errónea. Como dice en Hirschauer *et al.*: "Los p-values son a veces vistos como una pieza concluyente de

evidencia. Sin embargo, en años recientes, una abundante cantidad de publicaciones han advertido sobre errores inferenciales asociados con dichas visiones. (...) El incorrecto muestreo no-probabilístico descarta el uso de p-values ya que la estadística inferencial, preocupada en generalizar hacia una mayor población, está conceptualmente basada en la noción de muestreo aleatorio repetido (replicación estadística) y una distribución de muestreo resultante. Cuando no hay muestreo aleatorio, no hay distribución de muestreo y no hay error aleatorio de muestreo. La generación de datos aleatorios es por tanto una condición necesaria para una aplicación significativa de errores estándar y p-values. Cuando los datos no satisfacen este requerimiento probabilístico, los p-values son esencialmente ininterpretables” (traducido desde Hirschauer *et al.* 2020). Debido al sesgo explicado, intentar interpretar una prueba de hipótesis sería incorrecto en este contexto. La realizamos de todos modos como demostración de este hecho.



Como se observa en la Figura 39, las distribuciones ‘Conocidos’ y ‘Similares’ presentan una elevada superposición de sus valores de puntaje de docking (99.95% de superposición para ‘Similares’ y 93.97% para ‘Conocidos’), y sería altamente improbable, al elegir al azar una muestra entre ambas, ser capaces de distinguir desde cuál se originó (tasa de error de 96.96%). Más aun, el tamaño del efecto es de 0.3641, una diferencia pequeña a media según la Tabla VIII, lo que indica que la magnitud de la diferencia entre los valores de puntajes de docking entre ambas distribuciones es, para efectos prácticos, indistinguible. A pesar de eso, el test de hipótesis de U de Mann-Whitney entrega un p-value de $9.877652661735354e-14$, indicando una diferencia significativamente estadística. Esta aparente contradicción no es tal, como

ya se mencionó: ambas muestras no fueron obtenidas por muestreo aleatorio y este resultado no tiene valor interpretativo. Ha sido incluida sólo como ilustración de cómo una interpretación basada en pruebas de hipótesis, como es costumbre aún en el ambiente científico-biológico, hubiese llevado a conclusiones incorrectas sobre las observaciones presentadas, pese a que la estadística descriptiva presentada en la Figura 39 hace intuir una relación existente.

En tanto, las muestras 'Conocidos' y 'STK/STL' presentan un tamaño de efecto de 1.699, que corresponde a una magnitud entre muy grande a enorme. La superposición entre ambas es mínima (12.58% para 'STK/STL' y 11.81% para 'Conocidos'), y la probabilidad de muestreo erróneo es de 12.19%. Estos son indicadores fuertes de que el control y la muestra aleatoria son decididamente diferentes y aproximadamente disjuntos.

Finalmente, las muestras 'Similares' y 'STK/STL' presentan un patrón similar. Su tamaño de efecto es de 1.536, indicando una magnitud entre muy grande a enorme. La superposición entre ambas es de 15.8% para 'STK/STL' y 27.99% para 'Similares', con una tasa de error de 21.9%.

Estos resultados son consistentes con la hipótesis inicial de que los métodos de biología computacional son capaces de identificar, y más aun, expandir, el repertorio de moléculas estructuralmente similares con afinidad por el sitio de unión de hSGLT2.

13. CONCLUSIONES Y PROYECCIÓN DEL TRABAJO

Nuestra hipótesis inicial trataba sobre reconocer inhibidores conocidos y análogos estructurales de estos mediante métodos de biología computacional. Sin embargo, la similitud estructural no resulta necesariamente en actividad biológica, por más que una cierta correlación haya sido determinada en los resultados de docking. Esta aproximación, conocida como principio de propiedades similares (SPP, *similarity property principle*), que declara que compuestos similares deben tener propiedades similares, ha resultado ser muchas veces inadecuada, como lo ha demostrado la llamada “discontinuidad SAR (*Structure–activity Relationship*)”, en que pequeños cambios en estructura provocan grandes cambios en respuesta biológica, cuya forma más extrema son los “acantilados de actividad” (Maggiore *et al.* 2014). Si bien la similitud es y seguirá siendo importante en el proceso de desarrollo y descubrimiento de fármacos, resulta falso argumentar que cualquier compuesto similar inducirá similares respuestas en un determinado organismo. Va más allá de los objetivos de esta tesis evaluar o estimar la actividad biológica de los compuestos similares aquí descritos, pese a que puede realizarse mediante la familia de métodos computacionales conocidos como QSAR (*Quantitative Structure–activity Relationship*).

Una metodología más extensa debiera incluir no sólo el transportador en estudio, sino también explorar sus posibles interacciones. Por ejemplo, los inhibidores de SGLT2 tienen un grado variable de acción sobre SGLT1, y un fármaco eficaz debiese ser selectivo para SGLT2. Otros posibles transportadores que pueden sufrir interacción con la inhibición de SGLT2 son otros transportadores de azúcares, como la familia de

los GLUTs y SWEETs. Por tanto, los métodos aquí empleados deberían también aplicarse sobre transportadores adicionales y evaluar la selectividad de la respuesta. La metodología aquí empleada no se puede limitar sólo a la proteína de estudio, hSGLT2, sino que puede extenderse al estudio de cualquier otro transportador transmembrana. En particular, algunos procedimientos que se utilizaron aquí con utilidad para otros casos se enlistan a continuación:

- Uso de múltiples programas y parámetros a la hora de buscar homólogos, usando un consenso estimado para seleccionar secuencias. Aunque aquí se usó un criterio “crudo” de selección por consenso, este tiene la potencialidad de extenderse mediante la creación de una función objetivo que pondere las contribuciones de cada programa y seleccione secuencias siguiendo ésta, a fin de evitar con más precisión y rapidez la presencia de falsos positivos.
- La combinación de alineamientos múltiples, filogenia y predicción de segmentos transmembrana para seleccionar el mejor alineamiento y de este modo optimizar los pasos posteriores.
- El uso de criterios evolutivos para modelar proteínas de estructura desconocida, siguiendo el principio de máxima parsimonia.
- El uso de dinámica molecular para seleccionar estructuras estables para realizar docking molecular.
- Explotar, si es posible, el uso de computación de alto rendimiento para realizar una exploración más exhaustiva del espacio químico de actividad.
- Estimaciones estadísticas que reflejen los efectos del fenómeno observable y tomen en cuenta los tamaños de efecto por sobre la inferencia de poblaciones.

Es parte de la conclusión, por tanto, reconocer que los métodos de biología computacional continúan siendo una alternativa barata y eficiente a la exploración de nuevas alternativas farmacológicas si se sigue una metodología robusta y fundamentada, y que, como se ha comprobado en esta tesis, su eficacia para reconocer moléculas con actividad potencial puede ser utilizada para el caso de los transportadores transmembrana, los cuales son importantes blancos farmacológicos para una amplia variedad de patologías.



14. BIBLIOGRAFÍA

-Abramson, Jeff, Irina Smirnova, Vladimir Kasho, Gillian Verner, H. Ronald Kaback, and So Iwata (2003) Structure and Mechanism of the Lactose Permease of *Escherichia Coli*. *Science*, 301, no. 5633,610–15.

-Abramson, Jeff, and Ernest M. Wright. (2009) Structure and Function of Na(+)-Symporters with Inverted Repeats. *Current Opinion in Structural Biology*, 19, no. 4, 425–32.

-Adelman, Joshua L., Amy L. Dale, Matthew C. Zwier, Divesh Bhatt, Lillian T. Chong, Daniel M. Zuckerman, and Michael Grabe. (2011) Simulations of the Alternating Access Mechanism of the Sodium Symporter Mhp1. *Biophysical Journal*, 101, no. 10, 2399–2407

-Adelman, Joshua L., Ying Sheng, Seungho Choe, Jeff Abramson, Ernest M. Wright, John M. Rosenberg, and Michael Grabe (2014). Structural Determinants of Water Permeation through the Sodium-Galactose Transporter VSGLT. *Biophysical Journal* 106, no. 6, 1280–89.

-Al-Suyufi, Yasir, Khalid Al-Saleem, Ali Al-Mehaidib, Mohammed Banemai, Wajeeh M. Aldekhail, Abdulhakim Al-Muhandes, Manal Mohammed, *et al.* (2017) SLC5A1 Mutations in Saudi Arabian Patients with Congenital Glucose-Galactose Malabsorption. *Journal of Pediatric Gastroenterology and Nutrition*

-Allen, William J., Trent E. Balius, Sudipto Mukherjee, Scott R. Brozell, Demetri T. Moustakas, P. Therese Lang, David A. Case, Irwin D. Kuntz, y Robert C. Rizzo. (2015) DOCK 6: Impact of New Features and Current Docking Performance . *Journal of computational chemistry* 36, n.º 15, 1132-56.

- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, y D. J. Lipman.(1997) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 25, n.º 17, 3389-3402.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, y D. J. Lipman. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215, n.º 3, 403-10.
- Alvarado, F., y R. K. Crane. (1962) Phlorizin as a Competitive Inhibitor of the Active Transport of Sugars by Hamster Small Intestine, in Vitro. *Biochimica Et Biophysica Acta* 56 170-72.
- American Diabetes Association (2018). Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care*, 41(5), 917–928.
- Armougom, Fabrice, Sébastien Moretti, Olivier Poirot, Stéphane Audic, Pierre Dumas, Basile Schaeli, Vladimir Keduas, y Cedric Notredame. (2006) Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee . *Nucleic Acids Research* 34, n.o Web Server issue, W604-8.
- Bakan, Ahmet, Anindita Dutta, Wenzhi Mao, Ying Liu, Chakra Chennubhotla, Timothy R. Lezon, y Ivet Bahar. (2014) Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics. *Bioinformatics* 30, n.º 18, 2681-83.
- Bakan, Ahmet, Lidio M. Meireles, y Ivet Bahar.(2011) ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* 27, n.º 11 ,1575-77.
- Bernsel, Andreas, Håkan Viklund, Aron Hennerdal, y Arne Elofsson (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Research* 37, n.º Web Server issue W465-68.
- Bezanilla, F., E. Perozo, D. M. Papazian, and E. Stefani. (1991) Molecular Basis of Gating Charge Immobilization in Shaker Potassium Channels. *Science*. 254, no. 5032, 679–83.

- Bhat, Basharat, Nazir A. Ganai, Syed Mudasir Andrabi, Riaz A. Shah, y Ashutosh Singh (2017). TM-Aligner: Multiple Sequence Alignment Tool for Transmembrane Proteins with Reduced Time and Improved Accuracy. *Scientific Reports* 7, n.º 1, 12543
- Bill, Roslyn M., Peter J. F. Henderson, So Iwata, Edmund R. S. Kunji, Hartmut Michel, Richard Neutze, Simon Newstead, Bert Poolman, Christopher G. Tate, y Horst Vogel (2011). Overcoming Barriers to Membrane Protein Structure Determination. *Nature Biotechnology* 29, n.º 4, 335-40.
- Blaschek, Wolfgang. (2017) Natural Products as Lead Compounds for Sodium Glucose Cotransporter (SGLT) Inhibitors. *Planta Medica* 83, n.º 12-13: 985-93.
- Boehm, Markus, Matthew Crawford, Jamie E. Moscovitz, y Philip A. Carpino. (2018) Diabetes Area Patent Participation Analysis - Part II: Years 2011-2016. *Expert Opinion on Therapeutic Patents* 28, n.º 2, 111-22.
- Bosco K and Gruswitz Franz (2008) HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Structural Biology*
- Butina, Darko (1999) Unsupervised Data Base Agrupamiento Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *Journal of Chemical Information and Computer Sciences* 39(4), 747-750.
- Calado, Joaquim. (2009) Dapagliflozin, an Oral Sodium Glucose Cotransporter Type 2 Inhibitor for the Treatment of Type 2 Diabetes Mellitus. *IDrugs: The Investigational Drugs Journal* 12, n.o 12, 785-98.
- Cavasotto, Claudio N. (2011) Homology Models in Docking and High-Throughput Docking». *Current Topics in Medicinal Chemistry* 11, n.º 12. 1528-34.
- Cefalo, Chiara Maria Assunta, Francesca Cinti, Simona Moffa, Flavia Impronta, Gian Pio Sorice, Teresa Mezza, Alfredo Pontecorvi, y Andrea Giaccari. (2019) Sotagliflozin,

the first dual SGLT inhibitor: current outlook and perspectives. *Cardiovascular Diabetology* 18, n.º 1, 20.

-Chao, Edward C., y Robert R. Henry.(2010) SGLT2 Inhibition--a Novel Strategy for Diabetes Treatment . *Nature Reviews. Drug Discovery* 9, n.o 7, 551-59.

-Chasis, Herbert, Norman Jolliffe, y Homer W. Smith. (1933) «THE ACTION OF PHLORIZIN ON THE EXCRETION OF GLUCOSE, XYLOSE, SUCROSE, CREATININE AND UREA BY MAN». *Journal of Clinical Investigation* 12, n.º 6 1083-90.

- Chen, Yu-Chian. (2015) Beware of Docking! *Trends in Pharmacological Sciences* 36, n.º 2: 78-95.

-Choi, Chang-Ik (2016) Sodium-Glucose Cotransporter 2 (SGLT2) Inhibitors from Natural Products: Discovery of Next-Generation Antihyperglycemic Agents. *Molecules*. 21, n.o 9.

-Cock, Peter J. A., Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, *et al.* (2009) Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* 25, n.º 11, 1422-23

- Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. Ed. Routledge. 566 p.

-Crane, R.K., Miller, D. and Bihler, L., (1961) The restrictions on possible mechanisms of intestinal active transport of sugars in *Membrane Transport and Metabolism*. Eds., *Academic press*, 1961, 439-449

- Cumming, Geoff. Calin-Jageman Robert (2016) *Introduction to the New Statistics: Estimation, Open Science, and Beyond*. Ed. Routledge. 594 p.

- DeFronzo, Ralph A. (2009) From the Triumvirate to the Ominous Octet: A New Paradigm for the Treatment of Type 2 Diabetes Mellitus. *Diabetes* 58, n.º 4, 773-95.
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47, 20–33.
- Do, Chuong B., Mahathi S. P. Mahabhashyam, Michael Brudno, y Serafim Batzoglou. (2005) ProbCons: Probabilistic Consistency-Based Multiple Sequence Alignment. *Genome Research* 15, n.o 2 ,330-40.
- Dong, Lili, Ruirui Feng, Jiawei Bi, Shengqiang Shen, Huizhe Lu, y Jianjun Zhang.(2018) Insight into the Interaction Mechanism of Human SGLT2 with Its Inhibitors: 3D-QSAR Studies, Homology Modeling, and Molecular Docking and Molecular Dynamics Simulation. *Journal of Molecular Modeling* 24, n.º 4: 86.
- Drew, David, and Olga Boudker. (2016) Shared Molecular Mechanisms of Membrane Transporters. *Annual Review of Biochemistry* 85 , 543–72.
- Du, Xiaohui, Mike Lizarzaburu, Simon Turcotte, Taeweon Lee, Joanne Greenberg, Bei Shan, Peter Fan, Yun Ling, Julio C. Medina, y Jonathan Houze. (2011) Optimization of triazoles as novel and potent nonphlorizin SGLT2 inhibitors. *Bioorganic & Medicinal Chemistry Letters* 21, n.o 12 ,3774-79.
- Edgar, Robert C. (2004) MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinformatics* 5, 113.
- Edgar, Robert C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, n.o 5, 1792-97.
- Edwards, Noel, Catriona M. H. Anderson, Nichola J. Conlon, Andrew K. Watson, Rebecca J. Hall, Timothy R. Cheek, T. Martin Embley, y David T. Thwaites. (2017)

Resculpting the Binding Pocket of APC Superfamily LeuT-Fold Amino Acid Transporters. *Cellular and Molecular Life Sciences*, 1-18.

-Ehrenkranz, Joel R. L., Norman G. Lewis, C. Ronald Kahn, and Jesse Roth. (2005) Phlorizin: A Review. *Diabetes/Metabolism Research and Reviews* 21, no. 1 , 31–38.

-Eskandari, S., D. D. Loo, G. Dai, O. Levy, E. M. Wright, y N. Carrasco. (1997) Thyroid Na⁺/I⁻ Symporter. Mechanism, Stoichiometry, and Specificity. *The Journal of Biological Chemistry* 272, n.o 43, 27230-38.

-Eskandari, S., E. M. Wright, M. Kreman, D. M. Starace, y G. A. Zampighi. (1998) Structural Analysis of Cloned Plasma Membrane Proteins by Freeze-Fracture Electron Microscopy. *Proceedings of the National Academy of Sciences of the United States of America* 95, n.o 19, 11235-40.

-Faham, Salem, Akira Watanabe, Gabriel Mercado Besserer, Duilio Cascio, Alexandre Specht, Bruce A. Hirayama, Ernest M. Wright, y Jeff Abramson. (2008) The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na⁺/Sugar Symport. *Science* 321, n.o 5890, 810-14.

-Ferreira, Leonardo G., Ricardo N. Dos Santos, Glaucius Oliva, y Adriano D. Andricopulo. (2015) Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* 20, n.º 7: 13384-421.

-Floden, Evan W., Paolo D. Tommaso, Maria Chatzou, Cedrik Magis, Cedric Notredame, y Jia-Ming Chang. (2016) PSI/TM-Coffee: A Web Server for Fast and Accurate Multiple Sequence Alignments of Regular and Transmembrane Proteins Using Homology Extension on Reduced Databases. *Nucleic Acids Research* 44, n.o W1 , W339-343.

- Forrest, Lucy R., Reinhard Krämer, y Christine Ziegler. (2011) The structural basis of secondary active transport mechanisms. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1807, n.o 2, 167-88.
- Forrest, Lucy R., y Gary Rudnick. (2009) The Rocking Bundle: A Mechanism for Ion-Coupled Solute Flux by Symmetrical Transporters. *Physiology* 24 , 377-86.
- Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, y Weizhong Li. (2012) CD-HIT: accelerated for agrupamiento the next-generation sequencing data. *Bioinformatics* 28, n.º 23, 3150-52.
- Fuhrmann, G. F., S. Dervede, and G. Frenking. (1992) Phloretin Keto-Enol Tautomerism and Inhibition of Glucose Transport in Human Erythrocytes (Including Effects of Phloretin on Anion Transport). *Biochimica Et Biophysica Acta* 1110, no. 1 , 105–11.
- Fujimori, Yoshikazu, Kenji Katsuno, Ikumi Nakashima, Yukiko Ishikawa-Takemura, Hideki Fujikura, y Masayuki Isaji. (2008) Remogliflozin Etabonate, in a Novel Category of Selective Low-Affinity Sodium Glucose Cotransporter (SGLT2) Inhibitors, exhibits Antidiabetic Efficacy in Rodent Models. *The Journal of Pharmacology and Experimental Therapeutics* 327, n.o 1, 268-76.
- Gagnon, Dominique G., Andrea Holt, Francis Bourgeois, Bernadette Wallendorff, Michael J. Coady, and Jean-Yves Lapointe. (2005) Membrane Topology of Lazo 13-14 of the Na⁺/Glucose Cotransporter (SGLT1): A SCAM and Fluorescent Labelling Study. *Biochimica Et Biophysica Acta* 1712, no. 2, 173–84.
- Gagnon, Dominique G., Carole Frindel, and Jean-Yves Lapointe. (2007) Voltage-Clamp Fluorometry in the Local Environment of the C255-C511 Disulfide Bridge of the Na⁺/Glucose Cotransporter. *Biophysical Journal* 92, no. 7, 2403–11.

- Gagnon, Dominique G., Pierre Bissonnette, and Jean-Yves Lapointe. (2006) Identification of a Disulfide Bridge Linking the Fourth and the Seventh Extracellular Lazos of the Na⁺/Glucose Cotransporter. *The Journal of General Physiology* 127, no. 2, 145–58.
- Garnock-Jones, Karly P. (2017) Saxagliptin/Dapagliflozin: A Review in Type 2 Diabetes Mellitus *Drugs* 77, n.º 3 319-30.
- Goodwin, Nicole C., Ross Mabon, Bryce A. Harrison, Melanie K. Shadoan, Zheng Y. Almstead, Yiling Xie, Jason Healy, *et al.* (2009) Novel I-Xylose Derivatives as Selective Sodium-Dependent Glucose Cotransporter 2 (SGLT2) Inhibitors for the Treatment of Type 2 Diabetes. *Journal of Medicinal Chemistry* 52, n.o 20, 6201-4.
- Grempler, R., L. Thomas, M. Eckhardt, F. Himmelsbach, A. Sauer, D. E. Sharp, R. A. Bakker, M. Mark, T. Klein, y P. Eickelmann. (2012) Empagliflozin, a Novel Selective Sodium Glucose Cotransporter-2 (SGLT-2) Inhibitor: Characterisation and Comparison with Other SGLT-2 Inhibitors . *Diabetes, Obesity & Metabolism* 14, n.o 1 , 83-90.
- Hediger, Matthias A., Benjamin Cléménçon, Robert E. Burrier, and Elspeth A. Bruford. (2013) The ABCs of Membrane Transporters in Health and Disease (SLC Series): Introduction. *Molecular Aspects of Medicine* 34, no. 2–3, 95–107.
- Hediger, Matthias A., Michael F. Romero, Ji-Bin Peng, Andreas Rolfs, Hitomi Takanaga, and Elspeth A. Bruford. (2004) The ABCs of Solute Carriers: Physiological, Pathological and Therapeutic Implications of Human Membrane Transport proteinsIntroduction. *Pflugers Archiv: European Journal of Physiology* 447, no. 5, 465–68.
- Hediger MA, Coady MJ, Ikeda TS, Wright EM. (1987) Expression cloning and cDNA sequencing of the Na/glucose co-transporter. *Nature* 330: 379–381.

-Hediger MA, Ikeda T, Coady M, Gundersen CB, Wright EM. (1987) Expression of size-selected mRNA encoding the intestinal Na/glucose cotransporter in *Xenopus laevis* oocytes. *Proc Natl Acad Sci USA* 84, 2634–2637.

-Hediger MA, Mendlein J, Lee HS, Wright EM. (1991) Biosynthesis of the cloned intestinal Na/glucose cotransporter. *Biochim Biophys Acta* 1064, 360–364.

-Hediger MA, Turk E, Wright EM. (1989) Homology of the human intestinal Na/glucose and *Escherichia coli* Na/proline cotransporters. *Proc Natl Acad Sci USA* 86, 5748–5752.

-Hediger, M. A., J. Mendlein, H. S. Lee, y E. M. Wright. (1991) Biosynthesis of the Cloned Intestinal Na⁺/Glucose Cotransporter . *Biochimica Et Biophysica Acta* 1064, n.o 2, 360-64.

-Heller, Helmut, Michael Schaefer, y Klaus Schulten. (1993) Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *The Journal of Physical Chemistry* 97, n.º 31, 8343-60.

-Hirschauer, Norbert, Sven Grüner, Oliver Mußhoff, Claudia Becker, y Antje Jantsch. (2020) Can p-Values Be Meaningfully Interpreted without Random Sampling? *Statistics Surveys* 14: 71-91

-Huerta-Cepas, Jaime, François Serra, y Peer Bork. (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data . *Molecular Biology and Evolution* 33, n.º 6, 1635-38.

-Hummel, Charles S., Chuan Lu, Donald D. F. Loo, Bruce A. Hirayama, Andrew A. Voss, y Ernest M. Wright. (2011) Glucose Transport by Human Renal Na⁺/D-Glucose Cotransporters SGLT1 and SGLT2 . *American Journal of Physiology. Cell Physiology* 300, n.o 1, C14-21.

- Humphrey, William, Andrew Dalke, y Klaus Schulten. (1996) VMD: Visual molecular dynamics . *Journal of Molecular Graphics* 14, n.º 1, 33-38.
- Imamura, Masakazu, Keita Nakanishi, Takayuki Suzuki, Kazuhiro Ikegai, Ryota Shiraki, Takashi Ogiyama, Takeshi Murakami, et al. (2012) Discovery of Ipragliflozin (ASP1941): A Novel C-Glucoside with Benzothiophene Structure as a Potent and Selective Sodium Glucose Co-Transporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes Mellitus». *Bioorganic & Medicinal Chemistry* 20, n.º 10 3263-79.
- Ishchenko, Andrii, Enrique E. Abola, y Vadim Cherezov. (2017) Crystallization of Membrane Proteins: An Overview. *Methods in Molecular Biology* (Clifton, N.J.) 1607: 117-41.
- Jardetzky, O. (1966) Simple Allosteric Model for Membrane Pumps. *Nature* 211, no. 5052, 969–70.
- Jiang, Xuan, Donald D. F. Loo, Bruce A. Hirayama, y Ernest M. Wright. (2012) The Importance of Being Aromatic: π Interactions in Sodium Symporters . *Biochemistry* 51, n.o 47, 9480-87.
- Johnson, L. Steven, Sean R. Eddy, y Elon Portugaly. (2010) Hidden Markov model speed heuristic and iterative HMM search procedure . *BMC Bioinformatics* 11 , 431
- Jones, David T. (2007) Improving the Accuracy of Transmembrane Protein Topology Prediction Using Evolutionary Information. *Bioinformatics*. 23, n.º 5 , 538-44.
- Kakinuma, Hiroyuki, Takahiro Oi, Yuko Hashimoto-Tsuchiya, Masayuki Arai, Yasunori Kawakita, Yoshiki Fukasawa, Izumi Iida, et al. (2010) (1S)-1,5-Anhydro-1-[5-(4-Ethoxybenzyl)-2-Methoxy-4-Methylphenyl]-1-Thio-D-Glucitol (TS-071) Is a Potent, Selective Sodium-Dependent Glucose Cotransporter 2 (SGLT2) Inhibitor for Type 2 Diabetes Treatment . *Journal of Medicinal Chemistry* 53, n.o 8 , 3247-6.

- Kanwal, Abhinav, y Sanjay K. Banerjee. (2013) SGLT Inhibitors: A Novel Target for Diabetes. *Pharmaceutical Patent Analyst* 2, n.o 1, 77-91.
- Kato, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, y Takashi Miyata. (2002) MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Research* 30, n.o 14, 3059-66.
- Kato, Kazutaka, Kei-ichi Kuma, Hiroyuki Toh, y Takashi Miyata. (2005) MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment. *Nucleic Acids Research* 33, n.o 2, 511-18.
- Kato, Kazutaka, y Hiroyuki Toh. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9, n.o 4, 286-9.
- Katsuno, Kenji, Yoshikazu Fujimori, Yukiko Takemura, Masahiro Hiratochi, Fumiaki Itoh, Yoshimitsu Komatsu, Hideki Fujikura, y Masayuki Isaji. (2007) Sertigliflozin, a Novel Selective Inhibitor of Low-Affinity Sodium Glucose Cotransporter (SGLT2), Validates the Critical Role of SGLT2 in Renal Glucose Reabsorption and Modulates Plasma Glucose Level . *The Journal of Pharmacology and Experimental Therapeutics* 320, n.o 1, 323-30.
- Kazmier, Kelli, Derek P. Claxton, y Hassane S. Mchaourab. (2017) Alternating Access Mechanisms of LeuT-Fold Transporters: Trailblazing towards the Promised Energy Landscapes. *Current Opinion in Structural Biology* 45, 100-108.
- Keller, Rebecca, Christine Ziegler, y Dirk Schneider. (2014) When two turn into one: evolution of membrane transporters from half modules. *Biological Chemistry* 395, n.o 12, 1379–1388.
- Kinch, Lisa N, and Nick V Grishin. (2002) Evolution of Protein Structures and Functions. *Current Opinion in Structural Biology* 12, no. 3, 400–408.

- Kontoyianni, Maria. (2017) Docking and Virtual Screening in Drug Discovery». *Methods in Molecular Biology* (Clifton, N.J.) 1647: 255-66.
- Kovacich, Nikolas, y Benjamin Chavez. (2018) Ertugliflozin (Steglatro): A New Option for SGLT2 Inhibition. *Pharmacy and Therapeutics* 43, n.º 12,736-42.
- Krishnamurthy, Harini, y Eric Gouaux. (2012) X-Ray Structures of LeuT in Substrate-Free Outward-Open and Apo Inward-Open States. *Nature* 481, n.o 7382, 469-74.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, *et al.* (2007) Clustal W and Clustal X Version 2.0 . *Bioinformatics*. 23, n.o 21, 2947-48.
- Lassmann, Timo, y Erik LL Sonnhammer. (2005) Kalign – an accurate and fast multiple sequence alignment algorithm . *BMC Bioinformatics* 6, 298.
- Le, Si Quang, y Olivier Gascuel. (2008) An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution* 25, n.º 7 1307-20.
- Le Guilloux, Vincent, Peter Schmidtke, y Pierre Tuffery. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* 10, 168.
- Lee, Christopher, Catherine Grasso, y Mark F. Sharlow. (2012) Multiple sequence alignment using partial order graphs. *Bioinformatics* 18, n.o 3, 452-64.
- Leach, Andrew (2001) *Molecular Modelling: Principles and Applications*. 2º Edition. Pearson Education. 744 p.
- Letunic, Ivica, y Peer Bork. (2016) Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees . *Nucleic Acids Research* 44, n.º W1, W242-245.
- Li, An-Rong, Jian Zhang, Joanne Greenberg, TaeWeon Lee, y Jiwen Liu. (2011) Discovery of non-glucoside SGLT2 inhibitors. *Bioorganic & Medicinal Chemistry Letters* 21, n.o 8, 2472-75.

- Li, Jing, and Emad Tajkhorshid. (2012) A Gate-Free Pathway for Substrate Release from the Inward-Facing State of the Na⁺-Galactose Transporter. *Biochimica et Biophysica Acta (BBA) - Biomembranes, Membrane protein structure and function*, 1818, no. 2, 263–71.
- Li, Jing, y Emad Tajkhorshid. (2009) Ion-Releasing State of a Secondary Membrane Transporter. *Biophysical Journal* 97, n.o 11, L29-31.
- Li, Weizhong, Hamish McWilliam, Mickael Goujon, Andrew Cowley, Rodrigo Lopez, y William R. Pearson. (2012) PSI-Search: iterative HOE-reduced profile SSEARCH searching. *Bioinformatics* 28, n.º 12, 1650-51.
- Li, Zheng, Ashley S. E. Lee, Susanne Bracher, Heinrich Jung, Aviv Paz, Jay P. Kumar, Jeff Abramson, Matthias Quick, and Lei Shi. (2015) Identification of a Second Substrate-Binding Site in Solute-Sodium Symporters. *Journal of Biological Chemistry* 290, no. 1, 127–41.
- Lin, Lawrence, Sook Wah Yee, Richard B. Kim, and Kathleen M. Giacomini. (2015) SLC Transporters as Therapeutic Targets: Emerging Opportunities. *Nature Reviews. Drug Discovery* 14, no. 8, 543–60.
- Link, J. T, y Bryan K Sorensen. (2000) A Method for Preparing C-Glycosides Related to Phlorizin. *Tetrahedron Letters* 41, n.º 48, 9213-17.
- Liu, Yongchao, Bertil Schmidt, y Douglas L. Maskell. (2010) MSAProbs: Multiple Sequence Alignment Based on Pair Hidden Markov Models and Partition Function Posterior Probabilities. *Bioinformatics*. 26, n.o 16, 1958-64.
- Lomize, Mikhail A., Andrei L. Lomize, Irina D. Pogozheva, y Henry I. Mosberg. (2006) OPM: Orientations of Proteins in Membranes Database. *Bioinformatics*. 22, n.º 5 , 623-25.

- Löytynoja, Ari, y Nick Goldman. (2008) A model of evolution and structure for multiple sequence alignment. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, n.º 1512, 3913-19.
- Maggiora, Gerald, Martin Vogt, Dagmar Stumpfe, y Jürgen Bajorath. (2014) Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry* 57, n.º 8 (24 de abril de 2014): 3186-3204.
- Mahaffey, Kenneth W., Bruce Neal, Vlado Perkovic, Dick de Zeeuw, Greg Fulcher, Ngozi Erondu, Wayne Shaw, et al. (2018) Canagliflozin for Primary and Secondary Prevention of Cardiovascular Events: Results From the CANVAS Program (Canagliflozin Cardiovascular Assessment Study). *Circulation* 137, n.º 4 323-34.
- Mascitti, Vincent, Tristan S. Maurer, Ralph P. Robinson, Jianwei Bian, Carine M. Boustany-Kari, Thomas Brandt, Benjamin M. Collman, et al. (2011) Discovery of a Clinical Candidate from the Structurally Unique Dioxo-bicyclo[3.2.1]octane Class of Sodium-Dependent Glucose Cotransporter 2 Inhibitors. *Journal of Medicinal Chemistry* 54, n.o 8, 2952-60.
- Meng, Wei, Bruce A. Ellsworth, Alexandra A. Nirschl, Peggy J. McCann, Manorama Patel, Ravindar N. Girotra, Gang Wu, et al. (2008) Discovery of Dapagliflozin: A Potent, Selective Renal Sodium-Dependent Glucose Cotransporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes . *Journal of Medicinal Chemistry* 51, n.o 5, 1145-49.
- Meng, Xuan-Yu, Hong-Xing Zhang, Mihaly Mezei, y Meng Cui. (2011) Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer-Aided Drug Design* 7, n.º 2: 146-57.
- Michaud-Agrawal Naveen, Denning Elizabeth J., Woolf Thomas B., y Beckstein Oliver. (2011) MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* 32, n.º 10, 2319-27.

- Mishra, S., A. S. Singh, N. Mishra, H. Pandey, y V. K. Tiwari. (2017) Carbohydrate-Based Antidiabetic Agents From Nature, Chapter 5 . Elsevier. Discovery and Development of Antidiabetic Agents from Natural Products, editado por Goutam Brahmachari, 147-83.
- Mitchell, Peter. (1957) A General Theory of Membrane Transport From Studies of Bacteria. *Nature* 180, no. 4577, 134.
- Mittermayer, Friedrich, Erica Caveney, Claudia De Oliveira, Loukas Gourgiotis, Mala Puri, Li-Jung Tai, y J. Rick Turner. (2015) Addressing Unmet Medical Needs in Type 2 Diabetes: A Narrative Review of Drugs under Development. *Current Diabetes Reviews* 11, n.º 1 (2015): 17-31.
- Morgenstern, B. (1999) DIALIGN 2: Improvement of the Segment-to-Segment Approach to Multiple Sequence Alignment . *Bioinformatics*.15, n.o 3, 211-18.
- Müller, T., S. Rahmann, y M. Rehmsmeier. (2001) Non-Symmetric Score Matrices and the Detection of Homologous Transmembrane Proteins. *Bioinformatics*.17 Suppl 1, S182-189.
- Neal, Bruce, Vlado Perkovic, Dick de Zeeuw, Kenneth W. Mahaffey, Greg Fulcher, Peter Stein, Mehul Desai, *et al.* (2013) Rationale, Design, and Baseline Characteristics of the Canagliflozin Cardiovascular Assessment Study (CANVAS)--a Randomized Placebo-Controlled Trial . *American Heart Journal* 166, n.o 2, 217-223.e11.
- Nespoux, Josselin, y Volker Vallon (2020) Renal Effects of SGLT2 Inhibitors: An Update. *Current Opinion in Nephrology and Hypertension* 29, n.º 2 190-98.
- Ng, Wai-Lung, Ho-Chuen Li, Kit-Man Lau, Anthony K. N. Chan, Clara Bik-San Lau, y Tony K. M. Shing. (2017) Concise and Stereodivergent Synthesis of Carbasugars Reveals Unexpected Structure-Activity Relationship (SAR) of SGLT2 Inhibition. *Scientific Reports* 7, n.º 1 5581.

- Nisly, Sarah A., Denise M. Kolanczyk, y Alison M. Walton. (2012) Canagliflozin, a New Sodium-Glucose Cotransporter 2 Inhibitor, in the Treatment of Diabetes . *American Journal of Health-System Pharmacy: AJHP: Official Journal of the American Society of Health-System Pharmacists* 70, n.o 4, 311-19.
- Nissen, Steven E., y Kathy Wolski. (2007) Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *New England Journal of Medicine* 356, n.º 24: 2457-71.
- Nomura, Sumihiro, Shigeki Sakamaki, Mitsuya Hongu, Eiji Kawanishi, Yuichi Koga, Toshiaki Sakamoto, Yasuo Yamamoto, *et al.* (2010) Discovery of Canagliflozin, a Novel C-Glucoside with Thiophene Ring, as Sodium-Dependent Glucose Cotransporter 2 Inhibitor for the Treatment of Type 2 Diabetes Mellitus. *Journal of Medicinal Chemistry* 53, n.o 17, 6355-60.
- Notredame, C., D. G. Higgins, y J. Heringa. (2000) T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *Journal of Molecular Biology* 302, n.o 1, 205-17.
- Nugent, Timothy, y David T Jones. (2009) Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* 10, 159.
- Ohtake, Yoshihito, Tsutomu Sato, Takamitsu Kobayashi, Masahiro Nishimoto, Naoki Taka, Koji Takano, Keisuke Yamamoto, *et al.* (2012) Discovery of Tofogliflozin, a Novel C-Arylglucoside with an O-Spiroketal Ring System, as a Highly Selective Sodium Glucose Cotransporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes. *Journal of Medicinal Chemistry* 55, n.o 17, 7828-40.
- Oku, A., K. Ueta, K. Arakawa, T. Ishihara, M. Nawano, Y. Kuronuma, M. Matsumoto, *et al.* (2000) T-1095, an Inhibitor of Renal Na⁺-Glucose Cotransporters, May Provide a Novel Approach to Treating Diabetes. *Diabetes* 48, n.o 9, 1794-1800.

- Pajor, Ana M., Kathleen M. Randolph, Sandy A. Kerner, and Chari D. Smith. (2008) Inhibitor Binding in the Human Renal Low- and High-Affinity Na⁺/Glucose Cotransporters. *The Journal of Pharmacology and Experimental Therapeutics* 324, no. 3, 985–91.
- Palasz, A., Cież, D., Trzewik, B. et al. (2019) In the Search of Glycoside-Based Molecules as Antidiabetic Agents. *Topics in Current Chemistry (Z)* 377, 19.
- Pearson, William R. (1991) Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* 11, n.º 3, 635-50.
- Petersen C. (1835) Analyse des Phloridzins. *Annales Academie Science Francaise*. 15: 178.
- Poole, Raewyn M., y Rosselle T. Dungo. (2014) Ipragliflozin: First Global Approval . *Drugs* 74, n.o 5, 611-17.
- Posada, David, Thomas R. Buckley, y Jeffrey Thorne. (2004) Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology* 53, n.º 5, 793-808.
- Prasanna Kumar, K. M., Sujoy Ghosh, William Canovatchel, Nishant Garodia, y Sujith Rajashekar. (2017) A review of clinical efficacy and safety of canagliflozin 300 mg in the management of patients with type 2 diabetes mellitus. *Indian Journal of Endocrinology and Metabolism* 21, n.o 1, 196-209.
- Quick, M., D. D. Loo, y E. M. Wright. (2001) Neutralization of a Conserved Amino Acid Residue in the Human Na⁺/Glucose Transporter (HSGLT1) Generates a Glucose-Gated H⁺ Channel. *The Journal of Biological Chemistry*. 276, n.o 3, 1728-34.

- Rastogi, Ashu, y Anil Bhansali. (2017) SGLT2 Inhibitors Through the Windows of EMPA-REG and CANVAS Trials: A Review. *Diabetes Therapy*. 8, n.o 6, 1245-51.
- Ressl, Susanne, Anke C. Terwisscha van Scheltinga, Clemens Vonnheim, Vera Ott, and Christine Ziegler. (2009) Molecular Basis of Transport and Regulation in the Na(+)/Betaine Symporter BetP. *Nature*. 458, no. 7234, 47–52.
- Rockey, William M., y Adrian H. Elcock. «Structure Selection for Protein Kinase Docking and Virtual Screening: Homology Models or Crystal Structures?» *Current Protein & Peptide Science* 7, n.º 5 (octubre de 2006): 437-57. <https://doi.org/10.2174/138920306778559368>.
- Rudnick, Gary. (2013) How do transporters couple solute movements? *Molecular membrane biology*. 30, n.o 7, 355-59.
- Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*
- Sahraeian, Sayed Mohammad Ebrahim, y Byung-Jun Yoon. (2010) PicXAA: greedy probabilistic construction of maximum expected accuracy alignment of multiple sequences. *Nucleic Acids Research*. 38, n.o 15, 4917-28.
- Sali, A., y T. L. Blundell. (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*. 234, n.º 3, 779-815.
- Santer, René, y Joaquim Calado. (2010) Familial Renal Glucosuria and SGLT2: From a Mendelian Trait to a Therapeutic Target. *Clinical Journal of the American Society of Nephrology: CJASN* 5, n.o 1, 133-41.
- Sasseville, Louis J., Michael Morin, Michael J. Coady, Rikard Blunck, and Jean-Yves Lapointe. (2016) The Human Sodium-Glucose Cotransporter (HSGLT1) Is a Disulfide-

Bridged Homodimer with a Re-Entrant C-Terminal Lazo. *PLOS ONE* 11, no. 5, e0154589.

- Sawilowsky, Shlomo. (2009) New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods* 8, n.º 2

-Schmidtke, Peter, Axel Bidon-Chanal, F. Javier Luque, y Xavier Barril. MDpocket: (2011) Open-Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics*. 27, n.º 23, 3276-85.

-Schultz, S G, y P F Curran. (1970) Coupled transport of sodium and organic solutes. *Physiological Reviews*. 50, n.o 4, 637-718.

-Schwarz, Gideon. (1978) Estimating the Dimension of a Model. *Annals of Statistics* 6, n.º 2: 461-64

-Schweikhard, Eva S., y Christine M. Ziegler. (2012) Amino Acid Secondary Transporters: Toward a Common Transport Mechanism. *Current Topics in Membranes*. 70, 1-28.

-Screpanti, Emanuela, y Carola Hunte. (2007) Discontinuous Membrane Helices in Transport Proteins and Their Correlation with Function. *Journal of Structural Biology* 159, n.º 2, 261-67.

- Schäffer, Alejandro A., L. Aravind, Thomas L. Madden, Sergei Shavirin, John L. Spouge, Yuri I. Wolf, Eugene V. Koonin, y Stephen F. Altschul. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research* 29, n.º 14: 2994-3005.

- Shen, Min-yi, y Andrej Sali. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Science: A Publication of the Protein Society* 15, n.º 11. 2507-24.

- Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, *et al.* (2011) Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Molecular Systems Biology* 7, n.o 1, 539.
- Simossis, V. A., y J. Heringa. (2005) PRALINE: A Multiple Sequence Alignment Toolbox That Integrates Homology-Extended and Secondary Structure Information . *Nucleic Acids Research* 33, n.º suppl_2, W289-94.
- Singh, Satinder K., Atsuko Yamashita, y Eric Gouaux. (2007) Antidepressant Binding Site in a Bacterial Homologue of Neurotransmitter Transporters. *Nature* 448, n.o 7156, 952-56.
- Singh, Satinder K., Chayne L. Piscitelli, Atsuko Yamashita, y Eric Gouaux. (2008) A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* 322, n.o 5908, 1655-61.
- Smart, O.S., Goodfellow, J.M., Wallace, B.A. (1993). The pore dimensions of gramicidin A. *Biophysical Journal.*, 65, 2455-2460.
- Smith, Allen Watkins. (2008) Phylogenetics and Homology Modeling. Tesis doctoral. Rutgers University - Graduate School - New Brunswick.
- Song, Panai, Akira Onishi, Hermann Koepsell, and Volker Vallon. (2016) Sodium Glucose Cotransporter SGLT1 as a Therapeutic Target in Diabetes Mellitus. *Expert Opinion on Therapeutic Targets* 20, no. 9, 1109–25.
- Stamatakis, Alexandros. (2014) RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30, n.º 9, 1312-13.
- Stiles PG, Lusk G (1903) On the action of phlorizin. *American Journal of Physiology*; 10: 61-79.

- Subramanian, Amarendran R., Michael Kaufmann, y Burkhard Morgenstern. (2008) DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms for Molecular Biology* 3, 6.
- Sujatha, M. S., and Petety V. Balaji. (2004) Identification of Common Structural Features of Binding Sites in Galactose-Specific Proteins. *Proteins* 55, no. 1, 44–65.
- Söding, Johannes. (2005) Protein Homology Detection by HMM-HMM Comparison . *Bioinformatics* 21, n.o 7, 951-60.
- Takasu, Toshiyuki, Shoji Takakura, y Seiji Kaku. Pharmacological and clinical profile of ipragliflozin (Suglat®): a new therapeutic agent for type 2 diabetes. *Nihon Yakurigaku Zasshi. Folia Pharmacologica Japonica* 145, n.o 1, 36-42.
- Tamura, Yunoshin, Hiroh Miyagawa, Tatsusada Yoshida, y Hiroshi Chuman. (2015) Binding Interaction of SGLT with Sugar and Thiosugar by the Molecular Dynamics Simulation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1848, n.º 11, Part A : 2799-2804.
- Thompson, J. D., D. G. Higgins, y T. J. Gibson. (1994) CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research* 22, n.o 22, 4673-80.
- Trott, Oleg, y Arthur J. Olson. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of computational chemistry* 31, n.º 2, 455-61.
- Tuckerman, Mark. (2010) Statistical Mechanics: Theory and Molecular Simulation. Oxford University Press. 696 p.
- Turner, R. C., C. A. Cull, V. Frighi, y R. R. Holman. (1999) Glycemic Control with Diet, Sulfonylurea, Metformin, or Insulin in Patients with Type 2 Diabetes Mellitus:

Progressive Requirement for Multiple Therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group. *JAMA* 281, n.º 21,2005-12.

-Turk, E., M. G. Martín, and E. M. Wright. (1994) Structure of the Human Na⁺/Glucose Cotransporter Gene SGLT1. *The Journal of Biological Chemistry* 269, no. 21, 15204–9.

-Turk, E., y E. M. Wright. (1997) Membrane Topology Motifs in the SGLT Cotransporter Family. *The Journal of Membrane Biology* 159, n.o 1, 1-20.

- Vastermark, Åke, y Milton H. Saier. (2014) Evolutionary Relationship between 5+5 and 7+7 Inverted Repeat Folds within the Amino Acid-Polyamine-Organocation Superfamily. *Proteins* 82, n.º 2, 336-46.

- Vick, H., D. F. Diedrich, y K. Baumann (1973) Reevaluation of Renal Tubular Glucose Transport Inhibition by Phlorizin Analogs. *The American Journal of Physiology* 224, n.º 3: 552-57.

- Viereck, C., & Boudes, P. (2011). An analysis of the impact of FDA's guidelines for addressing cardiovascular risk of drugs for type 2 diabetes on clinical development. *Contemporary Clinical Trials*, 32(3), 324–332.

-Von Mering J (1886) Über künstlichen Diabetes; *Centralblatt für die medizinische Wissenschaft* 22:31

- Vorberg, Susann, Ina Koch, y Christian Buning. (2012) Sodium dependent glucose transporter (SGLT) 1 / 2 - elucidating inhibitor SAR and selectivity using homology modelling and 3D QSAR studies. *Journal of Cheminformatics* 4, n.º Suppl 1 P41.

-Wallace, Iain M., Orla O'Sullivan, Desmond G. Higgins, y Cedric Notredame. (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research* 34, n.o 6, 1692-99.

- Wang, L.-S.; Leebens-Mack, J.; Wall, P. K.; Beckmann, K.; Depamphilis, C. W. & Warnow, T. (2011). The impact of multiple protein sequence alignment on phylogenetic estimation. *IEEE/ACM Trans Comput Biol Bioinform* 8, 1108-1119.
- Watanabe, Akira, Seungho Choe, Vincent Chaptal, John M. Rosenberg, Ernest M. Wright, Michael Grabe, y Jeff Abramson. (2010) The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature* 468, n.o 7326, 988-91.
- Wells, R. G., A. M. Pajor, Y. Kanai, E. Turk, E. M. Wright, and M. A. Hediger. (1992) Cloning of a Human Kidney CDNA with Similarity to the Sodium-Glucose Cotransporter. *The American Journal of Physiology* 263, no. 3 Pt 2, F459-465.
- Weyand, S., T. Shimamura, O. Beckstein, M. S. P. Sansom, S. Iwata, P. J. F. Henderson, and A. D. Cameron. (2011) The Alternating Access Mechanism of Transport as Observed in the Sodium-Hydantoin Transporter Mhp1. *Journal of Synchrotron Radiation* 18, no. 1, 20–23.
- Wielert-Badt, S., J. T. Lin, M. Lorenz, S. Fritz, and R. K. Kinne. (2000) Probing the Conformation of the Sugar Transport Inhibitor Phlorizin by 2D-NMR, Molecular Dynamics Studies, and Pharmacophore Analysis. *Journal of Medicinal Chemistry* 43, no. 9, 1692–98.
- Wright, E. M., B. A. Hirayama, and D. F. Loo. (2007) Active Sugar Transport in Health and Disease. *Journal of Internal Medicine* 261, no. 1, 32–43.
- Wright, Ernest M. (2013) Glucose Transport Families SLC5 and SLC50. *Molecular Aspects of Medicine* 34, no. 2–3,183–96..
- Wright, Ernest M., Chiara Ghezzi, and Donald D. F. Loo. (2017) Novel and Unexpected Functions of SGLTs. *Physiology*, 32, no. 6, 435–43

- Wright, Ernest M., Donald D. F. Loo, Bruce A. Hirayama, y Eric Turk. (2004) Surprising Versatility of Na⁺-Glucose Cotransporters: SLC5. *Physiology*, 19, 370-76.
- Wright, Ernest M., Donald D. F. Loo, y Bruce A. Hirayama. (2011) Biology of Human Sodium Glucose Transporters. *Physiological Reviews* 91, n.o 2, 733-94.
- Wright, Ernest M., Eric Turk, and Martin G. Martin. (2002) Molecular Basis for Glucose-Galactose Malabsorption. *Cell Biochemistry and Biophysics* 36, no. 2–3, 115–21.
- Wright, Ernest M., y Eric Turk. (2004) The Sodium/Glucose Cotransport Family SLC5. *Pflugers Archiv: European Journal of Physiology* 447, n.o 5, 510-18.
- Wu, Jian-Sung, Yi-Hui Peng, Jiun-Ming Wu, Chieh-Jui Hsieh, Szu-Huei Wu, Mohane Selvaraj Coumar, Jen-Shin Song, *et al.* (2010) Discovery of Non-Glycoside Sodium-Dependent Glucose Co-Transporter 2 (SGLT2) Inhibitors by Ligand-Based Virtual Screening. *Journal of Medicinal Chemistry* 53, n.o 24, 8770-74.
- Xia, Xuhua. «Bioinformatics and Drug Discovery». *Current Topics in Medicinal Chemistry* 17, n.º 15 (2017): 1709-26.
<https://doi.org/10.2174/1568026617666161116143440>.
- Xie, Z., E. Turk, y E. M. Wright. (2000) Characterization of the *Vibrio Parahaemolyticus* Na⁺/Glucose Cotransporter. A Bacterial Member of the Sodium/Glucose Transporter (SGLT) Family. *The Journal of Biological Chemistry* 275, n.o 34, 25959-64.
- Xu, Ge, Binhua Lv, Jacques Y. Roberge, Baihua Xu, Jiyan Du, Jiajia Dong, Yuanwei Chen, *et al.* (2014) Design, Synthesis, and Biological Evaluation of Deuterated C-Aryl Glycoside as a Potent and Long-Acting Renal Sodium-Dependent Glucose Cotransporter 2 Inhibitor for the Treatment of Type 2 Diabetes. *Journal of Medicinal Chemistry* 57, n.o 4, 1236-51.

- Xu, Jinxing, Haoliang Yuan, Ting Ran, Yanmin Zhang, Haichun Liu, Shuai Lu, Xiao Xiong, et al. (2015) A Selectivity Study of Sodium-Dependent Glucose Cotransporter 2/Sodium-Dependent Glucose Cotransporter 1 Inhibitors by Molecular Modeling. *Journal of Molecular Recognition: JMR* 28, n.º 8: 467-79.
- Yamashita, Atsuko, Satinder K. Singh, Toshimitsu Kawate, Yan Jin, y Eric Gouaux. (2005) Crystal Structure of a Bacterial Homologue of Na⁺/Cl⁻-Dependent Neurotransmitter Transporters. *Nature* 437, n.o 7056, 215-23.
- Yang, Ziheng (2014) *Molecular Evolution: A Statistical Approach*. Oxford University Press. 449 p.
- Ye, Yongtao, Tak-Wah Lam, y Hing-Fung Ting. (2016) PnpProbs: a better multiple sequence alignment tool by better handling of guide trees. *BMC Bioinformatics* 17, n.o Suppl 8.
- Yu, Shuling, Jintao Yuan, Yi Zhang, Shufang Gao, Ying Gan, Meng Han, Yuewen Chen, Qiaoqiao Zhou, y Jiahua Shi. (2017) Combined HQSAR, Topomer CoMFA, Homology Modeling and Docking Studies on Triazole Derivatives as SGLT2 Inhibitors. *Future Medicinal Chemistry* 9, n.º 9 847-58.
- Zhao, Hongtao, y Amedeo Caflisch. (2015) Molecular Dynamics in Drug Design. *European Journal of Medicinal Chemistry* 91 : 4-14.
- Zhou, Lubing, Ellen V. Cryan, Michael R. D'Andrea, Stanley Belkowski, Bruce R. Conway, y Keith T. Demarest. (2003) Human Cardiomyocytes Express High Level of Na⁺/Glucose Cotransporter 1 (SGLT1). *Journal of Cellular Biochemistry* 90, n.o 2, 339-46.
- Zhou, Zheng, Juan Zhen, Nathan K. Karpowich, Christopher J. Law, Maarten E. A. Reith, y Da-Neng Wang. (2009) Antidepressant Specificity of Serotonin Transporter

Suggested by Three LeuT-SSRI Structures. *Nature Structural & Molecular Biology* 16, n.o 6, 652-57.

-Zinman, Bernard, Christoph Wanner, John M. Lachin, David Fitchett, Erich Bluhmki, Stefan Hantel, Michaela Mattheus, *et al.* (2015) Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *The New England Journal of Medicine* 373, n.o 22, 2117-28.

