



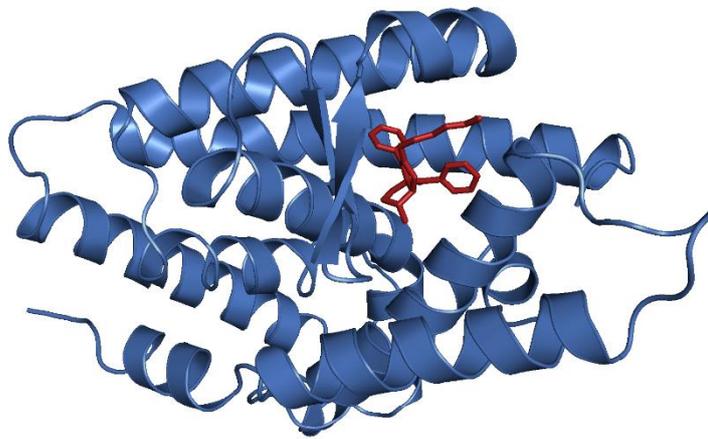
UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE  
Máster Universitario en Biotecnología y Bioingeniería  
Curso académico 2023-2024  
Facultad de Ciencias Experimentales  
Departamento de Bioquímica y Biología Molecular  
Área de Bioquímica y Biología Molecular

## Trabajo de Fin de Máster

---

# Búsqueda de compuestos moduladores del receptor LRH-1 con fines terapéuticos: una aproximación *in silico*

---



### Autor

Guillermo Gil Martín

### Tutores

Gregorio Fernández Ballester

María Clara Blanes Mira

## Declaración del tutor

D. GREGORIO FERNÁNDEZ BALLESTER, Doctor y Profesor Titular del área de Bioquímica y Biología molecular del Departamento de Bioquímica y Biología molecular, de la Universidad Miguel Hernández de Elche.

### CERTIFICA

Que el presente trabajo titulado:

“Búsqueda de compuestos moduladores del receptor LRH-1 con fines terapéuticos: una aproximación in silico”

y que constituye la Memoria del Trabajo Fin de Máster en Biotecnología y Bioingeniería, que presenta:

D. GUILLERMO GIL MARTÍN

ha sido realizado bajo su supervisión en el IDiBE, cumpliendo todos los requisitos necesarios.

Y para que así conste, se expide y firma el presente certificado en Elche a 29 de julio de 2024

Fdo.: Prof. Gregorio Fernández Ballester

## Índice

1. Resumen.....	1
2. Introducción .....	2
2.1. Receptor LRH-1: descripción y funciones fisiológicas .....	2
2.2. Receptores nucleares .....	2
2.3. Sitio de unión del receptor LRH-1 .....	4
2.4. Implicaciones terapéuticas del receptor LRH-1 .....	4
2.4.1. Implicación de LRH-1 en cáncer .....	4
2.3.2. Implicación de LRH-1 en diabetes .....	6
2.5. Descubrimiento de fármacos mediante cribado virtual. ....	7
3. Antecedentes y objetivos .....	7
4. Procedimiento experimental.....	8
4.1. Flujo de trabajo en cribado virtual.....	8
4.2. Selección de la estructura del receptor LRH-1 .....	10
4.3. Librería de compuestos etiquetada.....	11
4.4. Validación metodológica.....	12
4.5. Librería de compuestos naturales.....	14
4.6. Acoplamiento virtual con AutoDock .....	14
4.7. Acoplamiento virtual con Yasara.....	16
4.8. Otros programas de acoplamiento. ....	17
4.9. Métodos de acoplamiento consenso .....	18
4.11. Evaluación de propiedades farmacocinéticas .....	20
4.12. Programas accesorios de edición molecular .....	21
5. Resultados .....	22
5.1. Selección de la estructura tridimensional del receptor LRH-1.....	22
5.2. Análisis de los resultados de acoplamiento molecular. Curvas ROC y EP .....	24
5.3. Acoplamiento consenso .....	27
5.4. Ligandos seleccionados. Propiedades farmacocinéticas.....	28
5.5. Ligandos seleccionados. Interacciones proteína-ligando.....	30
5.6. Ligandos seleccionados. Disponibilidad comercial.....	35
6. Discusión.....	36
7. Conclusiones.....	38
Bibliografía.....	39

## Abreviaturas

AASS – Promedio de puntuación autoajustada (Average of auto-scaled scores)

AUC – Área bajo la curva (Area under the curve)

ADN – Ácido desoxirribonucleico

AF – Factor activador (Activator factor)

BBB – Barrera hematoencefálica (Blood-brain barrier)

CYP – Citocromo P450

DBD – Dominio de unión al ADN (DNA-binding domain)

DM1 – Diabetes mellitus tipo 1

EP – Gráficas de enriquecimiento (Enrichment plots)

ESOL – Solubilidad estimada (Estimated solubility)

ER – Receptor de estrógenos (Estrogen receptor)

FN – Falsos negativos (False negatives)

FP – Falsos positivos (False positives)

FPR – Tasa de falsos positivos (False positives rate)

GIA – Absorción gastrointestinal (Gastrointestinal absorption)

HBA – Aceptores de hidrógeno (Hydrogen bond acceptors)

HBD – Donadores de hidrógeno (Hydrogen bond donors)

HCC – Carcinoma hepatocelular

HTS – Cribado de alto rendimiento (High-throughput screening)

IL - Interleucina

LBD – Dominio de unión al ligando (Ligand-binding domain)

LRH-1 – Homólogo del receptor hepático 1 (Liver receptor homolog-1)

MMP – Metaloproteínasa de matriz

NSR – Relación de puntuación normalizada (Normalized score ratio)

NR – Receptor nuclear (Nuclear receptor)

PGP – Glicoproteína-P (P-glycoprotein)

PM – Peso molecular

PDB – Protein Data Bank

RBN – Clasificación por número (Rank by number)

RBR – Clasificación por rango (Rank by rank)

RBV – Clasificación por voto (Rank by vote)

ROC – Característica operativa del receptor (Receiver operating characteristic)

RBN – Número de enlaces rotables (Rotatable bonds number)

TN – Verdaderos negativos (True negatives)

TP – Verdaderos positivos (True positives)

TPR – Tasa de verdaderos positivos (True positives rate)

VS – Cribado virtual (Virtual Screening)

## 1. Resumen

El homólogo del receptor hepático 1 o LRH-1 (Liver Receptor Homologue 1), también conocido como NR5A2, es un receptor nuclear que actúa como factor de transcripción, promoviendo la transcripción de genes implicados en varios procesos como el metabolismo, el crecimiento y la diferenciación celular. Su potencial uso como diana terapéutica ha sido demostrado en múltiples estudios, por lo que el desarrollo de fármacos que tengan como objetivo LRH-1 podría representar una nueva oportunidad para el tratamiento de algunas enfermedades cancerosas, metabólicas e inmunitarias.

En este trabajo se han buscado nuevas moléculas que sean capaces de unirse a este receptor mediante una estrategia de cribado virtual, utilizando varios programas de acoplamiento molecular y métodos de acoplamiento consenso que optimizan el rendimiento de esta búsqueda. El objetivo es proponer potenciales activadores o inhibidores del receptor LRH-1 a partir de una librería de compuestos naturales disponibles comercialmente y poder trasladar estas moléculas a ensayos experimentales.

Del análisis de los resultados computacionales se proponen 5 ligandos cuyas propiedades farmacocinéticas resultan favorables, sugiriendo así que en un futuro podrían llevarse a cabo bioensayos *in vitro* e *in vivo* que verifiquen su potencial terapéutico en el tratamiento de las enfermedades anteriormente mencionadas.

Palabras clave: LRH-1, cáncer, diabetes, cribado virtual, biodisponibilidad, descubrimiento de fármacos.

### **Abstract**

*The liver receptor homologue 1 or LRH-1, also known as NR5A2, is a nuclear receptor that acts as a transcription factor, promoting the transcription of genes involved in various processes such as metabolism, growth, and cell differentiation. Its therapeutic potential has been demonstrated in multiple studies, thus, the development of drugs that target LRH-1 could be a new opportunity for the treatment of cancerous, metabolic, and immune diseases.*

*In this work we have searched for new molecules that are capable of binding to this receptor through a virtual screening strategy, using several molecular docking programs and consensus methods that improve search performance. All with the aim of finding potential activators or inhibitors of the LRH-1 receptor from a library of natural compounds commercially available and to be able to bring these molecules to experimental assays.*

*Analyzing the computational results, five candidate ligands whose pharmacokinetic properties are favorable have been proposed, thus suggesting that in the future, *in vitro* and *in vivo* bioassays could be carried out to verify their therapeutic potential for the treatment of the aforementioned diseases.*

*Keywords: LRH-1, cancer, diabetes, virtual screening, bioavailability, drug discovery.*

## 2. Introducción

### 2.1. Receptor LRH-1: descripción y funciones fisiológicas

LRH-1 es un receptor nuclear que actúa como un factor de transcripción dependiente de ligando. Se expresa en múltiples órganos y tejidos y regula la expresión de genes clave en procesos como la diferenciación celular, la proliferación y el metabolismo (Merk et al., 2021). Sus funciones específicas incluyen:

-En el páncreas, LRH-1 tiene roles tanto exocrinos como endocrinos. Es crucial para la producción y secreción del jugo pancreático, ya que está involucrado en la expresión de genes que codifican enzimas digestivas. En el páncreas endocrino, favorece la supervivencia de los islotes de Langerhans, aunque los mecanismos moleculares precisos aún no se han aclarado completamente (Gauthier et al., 2022). Durante el desarrollo embrionario, LRH-1 regula positivamente los genes de las ciclinas D1/E1 y c-Myc a través de la vía Wnt/ $\beta$ -catenina.

-En el hígado, LRH-1 desempeña un papel vital en la homeostasis de lípidos al incrementar la expresión de genes relacionados con el transporte inverso de colesterol, como ApoA1, componente principal de las lipoproteínas de alta densidad. También regula la biosíntesis de ácidos biliares mediante la inducción de los citocromos Cyp7a1 (colesterol 7 $\alpha$ -hidroxilasa) y Cyp8b1 (esterol 12 $\alpha$ -hidroxilasa), que convierten el colesterol en ácidos biliares. Además, participa en la homeostasis de glucosa al controlar la expresión de la enzima glucoquinasa. LRH-1 es un mediador importante en procesos inflamatorios y en la respuesta al estrés celular (Schoonjans et al., 2021), reprimiendo la expresión de proteínas de fase aguda hepática y promoviendo la síntesis de citocinas antiinflamatorias como el antagonista del receptor de interleucina 1 (IL-1RA).

-En el tejido mamario, LRH-1 promueve la biosíntesis de estrógenos mediante la inducción de la enzima aromatasa (CYP19). Los estrógenos, a su vez, activan los receptores de estrógenos (ER- $\alpha$  y ER- $\beta$ ), manteniendo el crecimiento mamario (Merk et al., 2021).

-En el intestino, LRH-1 facilita la renovación epitelial al aumentar la proliferación celular, a través de la señalización de la  $\beta$ -catenina, que promueve la expresión de ciclinas D1 y E1, entre otras proteínas que facilitan la progresión del ciclo celular. Además, regula la respuesta inmunitaria local al inducir la expresión de enzimas esteroidogénicas implicadas en la producción de glucocorticoides (Merk et al., 2021).

-En el ovario, LRH-1 participa en la biosíntesis de estrógenos, de forma similar a su función en el tejido mamario, regulando la ovulación y la maduración del folículo (Merk et al., 2021).

### 2.2. Receptores nucleares

LRH-1 es un miembro de la familia de receptores nucleares (NR) que se divide en siete subfamilias (NR0 a NR6). Dentro de la subfamilia 5, grupo A, se encuentra también el factor esteroidogénico 1, conocido como NR5A1. Los NR comparten una estructura global bastante conservada (Ortlund et al., 2018), compuesta por cinco dominios (ver Figura 1A):

**-A/B:** Este es el dominio N-terminal, caracterizado por su alta variabilidad y desorden. Incluye una región conocida como AF-1 (Activator Function-1), que interactúa con diversas proteínas reguladoras, tales como coactivadores y correpresores.

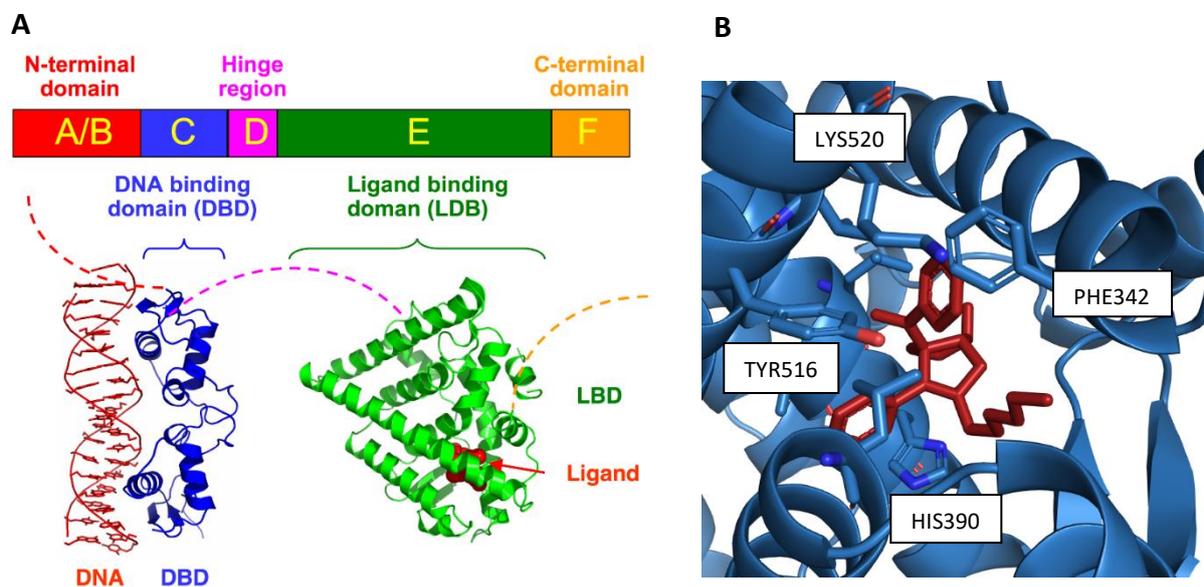
**-C:** Este dominio, denominado dominio de unión al ADN o DBD (DNA Binding Domain), está ubicado en la región N-terminal de la proteína. Su función es reconocer y unirse a secuencias específicas del ADN (mediante un motivo

de dedo de zinc) en los promotores de los genes que regula el receptor. Además, en este dominio se encuentran algunos residuos que facilitan la dimerización de los receptores nucleares.

-D: La región bisagra (hinge region) conecta el dominio de unión al ADN con el dominio de unión al ligando. Proporciona flexibilidad a la proteína para cambiar de conformación.

-E: Este es el dominio de unión al ligando o LBD (Ligand Binding Domain), ubicado en la región central de la proteína. Su principal función es unirse a los ligandos específicos (como hormonas o moléculas lipídicas) en el sitio de unión (ver Figura 1B). También contiene una región AF-2 (Activator Function-2) que facilita la interacción con proteínas reguladoras, de manera similar a lo que ocurre en el dominio A/B.

-F: El dominio C-terminal, que es altamente variable y desordenado.



**Figura 1. Organización estructural de los receptores nucleares (NR).** **A)** Arquitectura de los receptores nucleares por dominios (Hoffmeier, 2007). Se representa linealmente la secuencia de aminoácidos de un receptor nuclear y la estructura en 3D de los dominios DBD y LBD. Las estructuras del dominio N-terminal, la región bisagra y el dominio C-terminal no se han podido determinar con precisión, por lo tanto, están representadas con líneas discontinuas de color rojo, púrpura y naranja, respectivamente. **B)** Ampliación del dominio LBD. Se muestran los residuos del centro activo (en color azul) interaccionando con el ligando (rojo). Se etiquetan algunos de los residuos más relevantes.

En general, los receptores nucleares (NR) se encuentran como monómeros cuando no están ligados y forman homo o heterodímeros al unirse al ADN. Sin embargo, el receptor LHR-1 es una excepción, ya que se une al ADN como monómero. La mayoría de los NR actúan sobre la transcripción cuando la unión de un ligando provoca un cambio conformacional que incrementa la afinidad del receptor por proteínas coactivadoras o correpresoras (que se unen a los sitios AF-1 y AF-2), facilitando así el reclutamiento de estas para promover o inhibir la

transcripción de sus genes objetivo. En contraste, LRH-1 tiene una actividad transcripcional constitutiva incluso en ausencia de un ligando (Scott Jr., 2017). Además, modificaciones postraduccionales como la fosforilación y la sumoilación también pueden alterar la actividad de los NR.

El bolsillo de unión de los NR está compuesto principalmente por residuos hidrofóbicos, lo que permite la unión de ligandos de naturaleza lipídica. Sin embargo, algunos residuos polares en el bolsillo permiten la formación de puentes de hidrógeno, que orientan el ligando adecuadamente y confieren especificidad al receptor.

### 2.3. Sitio de unión del receptor LRH-1

El dominio de unión al ligando (LBD) de LRH-1 tiene una estructura típica de los receptores nucleares, formada por varias hélices alfa que crean una cavidad hidrofóbica donde se aloja el ligando (de naturaleza lipofílica). Entre los residuos importantes en el LBD de LRH-1 se encuentran (Figura 1B), por una parte residuos hidrofóbicos enterrados en el bolsillo de unión como la valina 417, la isoleucina 387 o la leucina 386, aunque también hay residuos polares importantes en esta región, como la histidina 390. Y por otra parte residuos polares cercanos a la superficie del sitio de unión como la tirosina 516, la treonina 341 y la lisina 520, existiendo también algunos residuos hidrofóbicos en esta zona, como la fenilalanina 342. Los residuos hidrofóbicos crean un entorno no polar que permite la unión eficaz de ligandos lipofílicos, estabilizando el complejo ligando-receptor y determinando la afinidad, mientras que los residuos polares de la superficie están involucrados en la especificidad y el mecanismo de unión inicial del ligando. Además, estos residuos también son importantes en la comunicación alostérica dentro de la proteína, transmitiendo los cambios conformacionales hacia los dominios que interactúan con proteínas coactivadoras, y permitiendo la activación transcripcional. Por otra parte, existen moléculas de agua conservadas en el sitio de unión que también juegan un papel relevante en la estabilización del ligando, mediando interacciones indirectas. (Cato et al., 2022).

### 2.4. Implicaciones terapéuticas del receptor LRH-1

Por los roles que desempeña este receptor, no es de extrañar que su disfunción esté asociada con enfermedades como cáncer y diabetes. Así pues, en este apartado se describen cuáles son los mecanismos que desencadenan estas patologías a través de LRH-1 y se plantean los posibles abordajes terapéuticos de tipo farmacológico.

#### 2.4.1. Implicación de LRH-1 en cáncer

El receptor LRH-1 puede contribuir al desarrollo de diversos tipos de cáncer, debido a su papel en la coordinación del desarrollo, la diferenciación y la proliferación celular. Se ha asociado específicamente con la tumorigénesis en cáncer de páncreas, hígado, mama, ovario y tumores gastrointestinales (Dong et al., 2015).

##### Cáncer de páncreas

El cáncer de páncreas tiene el peor pronóstico entre todos los tipos de cáncer, con una tasa de supervivencia a 5 años de entre el 5% y el 6%, debido a que aproximadamente el 90% de los casos se diagnostican en estadios avanzados. Estudios recientes han mostrado una mayor expresión de LRH-1 en líneas celulares de cáncer de

páncreas y en muestras de tejido de cáncer de páncreas humano en comparación con células normales del epitelio ductal pancreático. LRH-1 regula al alza los genes de ciclinas D1/E1 y c-Myc, lo que puede llevar a una mayor proliferación celular. También se ha observado que la sobreexpresión de LRH-1 estimula la expresión de las metaloproteinasas de matriz-2 (MMP-2) y de matriz-9 (MMP-9), aumentando la migración celular y la invasividad del tumor. Además, se sospecha de interacciones entre las vías de LRH-1 y Kras, aunque se requiere un análisis más detallado. Existen varios compuestos en fases tempranas de desarrollo que actúan como inhibidores específicos de LRH-1 y que han demostrado reducir el crecimiento y la proliferación de células cancerosas in vitro. Sin embargo, los mecanismos moleculares de estos antagonistas aún no se han comprendido completamente (Dong et al., 2015).

#### Carcinoma hepatocelular

El carcinoma hepatocelular (HCC), la forma más común de cáncer de hígado, se origina en los hepatocitos y suele presentarse en personas con enfermedades hepáticas crónicas como la cirrosis. Se ha observado que la expresión de LRH-1 aumenta significativamente en el HCC, contribuyendo a la proliferación de células cancerosas al regular positivamente la ciclina E1 y c-Myc. Además, LRH-1 puede estar involucrado en la resistencia a la apoptosis al inhibir la expresión de p21, permitiendo que las células cancerosas sobrevivan y se propaguen. Aunque el papel de LRH-1 en el desarrollo del HCC es complejo y no completamente comprendido, se ha demostrado que pequeñas moléculas que actúan como antagonistas o agonistas inversos de este receptor pueden suprimir la proliferación de células de cáncer de hígado. Por lo tanto, el desarrollo de inhibidores farmacológicos de LRH-1 tiene potencial terapéutico (Dong et al., 2015).

#### Cáncer de mama

El cáncer de mama, que se origina en las células de las mamas, es el segundo cáncer más común en mujeres en Estados Unidos y el más frecuente en España (Red Española de Registros de Cáncer, REDECAN). Aunque puede afectar tanto a hombres como a mujeres, es mucho más común en mujeres. Los estrógenos juegan un papel central en el crecimiento del cáncer de mama, y la mayoría de estos cánceres sobreexpresan el receptor de estrógeno  $\alpha$  (ER $\alpha$ ), un factor de transcripción que impulsa la proliferación de células cancerosas. Hasta el 50% de las pacientes con cáncer de mama ER $\alpha$  positivo no responden a las terapias endocrinas y muchas recaen. LRH-1 se expresa anormalmente en el 45% de los carcinomas de mama y se correlaciona con el estado de ER del tumor. LRH-1 se relaciona con genes involucrados en la síntesis de esteroides, sugiriendo que influye en la esteroidogénesis en el cáncer de mama, promoviendo la progresión del cáncer a través de la inducción de la enzima aromatasa y la biosíntesis local de estrógenos. Además, se ha demostrado que la sobreexpresión de LRH-1 estimula la motilidad, migración e invasividad celular, mediante la remodelación de la actina, la regulación positiva de MMP9 y la inactivación de la E-cadherina, lo que indica que LRH-1 está involucrado en la transición epitelio-mesénquima (EMT). Dado que muchas pacientes desarrollan resistencia a las terapias hormonales, bloquear la acción de LRH-1 en tumores ER positivos podría ser una nueva vía para el desarrollo de fármacos (Dong et al., 2015).

### Cáncer de ovario

El cáncer de ovario es el tipo de cáncer ginecológico más letal entre las mujeres. Su tasa de supervivencia a cinco años para las etapas avanzadas es del 30-40%. El receptor LRH-1 podría jugar un papel crucial en este tipo de cáncer. Se ha demostrado que promueve la proliferación y migración de células tumorales, y su sobreexpresión se correlaciona con un pronóstico desfavorable en el cáncer de ovario. Los niveles elevados de LRH-1 están relacionados con metástasis a los ganglios linfáticos y a la cavidad peritoneal, así como resistencia al platino (Libo Yu et al., 2020).

La inhibición de LRH-1 podría reducir la proliferación y migración de las células cancerosas, así como revertir la resistencia a los agentes quimioterapéuticos basados en platino. Esto sugiere que LRH-1 podría ser una diana terapéutica potencial, permitiendo el desarrollo de terapias que mejoren el pronóstico de los pacientes.

### Tumores gastrointestinales

Los tumores gastrointestinales se desarrollan en el tracto gastrointestinal, incluyendo el esófago, estómago, intestino delgado, colon y recto. Ocurren cuando las vías de señalización relacionadas con la proliferación celular se alteran y los factores de transcripción desregulados conducen a una regulación génica anormal. LRH-1 coordina la renovación de células intestinales mediante el control de la proliferación celular a través de la vía Wnt/ $\beta$ -catenina, que induce las ciclinas D1 y E1. Esta vía está mutada en un 80% de los cánceres colorrectales. Basándose en la sinergia entre LRH-1 y la señalización de la  $\beta$ -catenina, se sugiere que LRH-1 contribuye a la tumorigénesis gastrointestinal al influir en el ciclo celular. Aunque se necesita más investigación para entender completamente el papel de LRH-1 en estos tumores, los estudios actuales sugieren que los inhibidores farmacológicos de esta proteína podrían ser una vía prometedora para el tratamiento (Dong et al., 2015).

### 2.3.2. Implicación de LRH-1 en diabetes

La diabetes mellitus tipo 1 (DM1) es una enfermedad autoinmune caracterizada por la destrucción selectiva de las células beta productoras de insulina en los islotes pancreáticos, debido a la acción del sistema inmunitario. Esto resulta en la incapacidad del organismo para producir insulina y la consecuente aparición de hiperglucemia. A pesar de los avances en tecnología de dispositivos médicos, análogos de insulina y los esfuerzos por generar células productoras de insulina *in vitro*, todavía no existe una terapia efectiva para reemplazar y proteger las células beta perdidas en la DM1.

En los últimos años, LRH-1 ha sido validado como una nueva diana terapéutica potencial para la DM1, proponiendo un cambio de paradigma en su tratamiento. Se ha descubierto que la activación de LRH-1 protege a las células beta pancreáticas de la apoptosis y fomenta un entorno antiinflamatorio local mediante la síntesis de IL-10 en macrófagos. Investigaciones recientes también han demostrado que el agonismo de LRH-1 puede restaurar el control glucémico en varios modelos de ratones con DM1 y mejorar la supervivencia y función de los islotes humanos tanto *in vitro* como *in vivo*. Se plantea la hipótesis de que la regeneración de células beta mediada por esta activación podría estar facilitada por la transdiferenciación de células alfa en células beta (Gauthier et al., 2018). A partir de estos hallazgos, se considera que pequeñas moléculas químicas agonistas de LRH-1 podrían ser candidatas prometedoras para el tratamiento de la DM1.

## 2.5. Descubrimiento de fármacos mediante cribado virtual.

El proceso de descubrimiento y desarrollo de medicamentos es extremadamente complejo, prolongado y costoso, por lo que es esencial utilizar todas las herramientas disponibles para facilitar esta tarea y reducir la probabilidad de fracaso en las etapas avanzadas del proceso. El cribado virtual (Virtual Screening, VS) es un método computacional que permite evaluar y clasificar grandes colecciones de compuestos según su capacidad para unirse a un receptor objetivo. Hay dos enfoques principales para realizar un cribado virtual:

**Cribado virtual basado en el ligando:** Este método se basa en los ligandos conocidos existentes. Utiliza modelos de relación cuantitativa estructura-actividad (quantitative structure-activity relationship, QSAR), farmacóforos o similitudes químicas para predecir nuevos compuestos (Devarakonda et al., 2023). Si se dispone de suficientes moléculas activas, también se pueden emplear métodos de aprendizaje automático o machine learning.

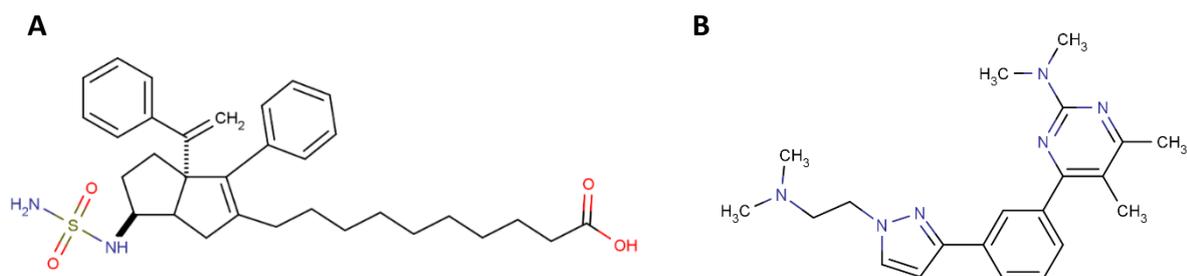
**Cribado virtual basado en la estructura:** Este enfoque implica el acoplamiento de la proteína con una gran variedad de pequeñas moléculas. Requiere la estructura tridimensional (3D) tanto del receptor como de los ligandos. En este método, el descubrimiento de nuevos compuestos activos se fundamenta en la evaluación de las interacciones fisicoquímicas entre el receptor y los ligandos mediante algoritmos matemáticos (Keseru et al., 2017). Esta estrategia será abordada en detalle en el punto 4.1.

El análogo experimental del cribado virtual es el cribado de alto rendimiento (High-throughput screening, HTS), que involucra la realización de experimentos de laboratorio o bioensayos para probar miles o incluso millones de compuestos químicos. Aunque el cribado virtual puede acelerar y reducir significativamente los costos del proceso de descubrimiento de fármacos al disminuir el número de compuestos que necesitan ser evaluados en el laboratorio, es importante tener en cuenta que no es una solución completa y siempre requiere una validación experimental rigurosa de los compuestos identificados durante el proceso de cribado virtual. En resumen, el VS y el HTS son complementarios.

## 3. Antecedentes y objetivos

Inicialmente, se pensaba que LRH-1 era un receptor huérfano, es decir, sin un ligando endógeno conocido. No obstante, la mayoría de los estudios recientes sugieren que los fosfolípidos podrían actuar como ligandos endógenos para este receptor, incluyendo el fosfatidilglicerol, así como derivados de la fosfatidilcolina y la fosfatidiletanolamina (Merk et al., 2021).

Además, se han identificado varios ligandos sintéticos para LRH-1 que no son lipídicos (Fletcher et al., 2013). Estos incluyen tanto agonistas como RJW100, BL001 y 6N-10A, como antagonistas y agonistas inversos como SR1848 y 3d2. Aunque ninguno de estos compuestos ha llegado aún a ensayos clínicos, representan avances significativos en la investigación básica. La figura 2 ilustra la estructura química de dos de estos compuestos.



**Figura 2. Moduladores sintéticos del receptor LRH-1. A)** 6N-10CA. Agonista de LRH-1 más potente sintetizado hasta la fecha, con una EC<sub>50</sub> de 43 nM. **B)** 3d2 (ChemBridge 16728690). Agonista inverso de LRH-1 con una IC<sub>50</sub> de 25 μM.

Los ligandos sintéticos de LRH-1 pueden ser valiosos para regular su actividad, pero presentan varios problemas, como baja estabilidad, falta de especificidad al interactuar con otros receptores hormonales, potencial toxicidad y efectos no deseados en otros sistemas biológicos que pueden interferir con la actividad de LRH-1. Estos inconvenientes pueden limitar la efectividad de estos ligandos en ciertas situaciones. Por ello, es importante desarrollar ligandos alternativos que, con un diseño y optimización adecuados, puedan convertirse en fármacos específicos y seguros. Existe amplia evidencia que sugiere que LRH-1 juega un papel crucial en diversos tipos de cáncer y en la diabetes, lo que hace de este receptor un objetivo prometedor para terapias contra estas enfermedades.

#### **Objetivo general:**

El objetivo general de este trabajo es encontrar nuevas moléculas con afinidad y especificidad mejoradas, a partir de librerías de compuestos naturales de APExBIO (<https://www.apexbt.com/>), que puedan funcionar como inhibidores o activadores del receptor LRH-1.

#### **Objetivos específicos:**

- Revisar y seleccionar las mejores estructuras tridimensionales de LRH-1 para ser utilizadas con cribado virtual.
- Seleccionar las librerías químicas y ordenar los compuestos según su capacidad de unión al receptor, mediante diferentes métodos de acoplamiento.
- Analizar los resultados mediante métodos de acoplamiento consenso.
- Seleccionar las mejores moléculas en base a sus propiedades farmacocinéticas y disponibilidad.

Se prevé poder evaluar posteriormente el potencial terapéutico de las moléculas seleccionadas mediante diferentes bioensayos.

## 4. Procedimiento experimental

### 4.1. Flujo de trabajo en cribado virtual

El proceso de cribado virtual (ver figura 3) comienza con la selección de un modelo estructural tridimensional de la diana objetivo (como un canal, una enzima o un receptor) involucrada en una enfermedad que en el caso de este trabajo fue el receptor LRH-1.

El siguiente paso es determinar el sitio de unión del receptor. En ocasiones, la estructura cristalina contiene un ligando cocrystalizado, lo que proporciona una información muy valiosa sobre los sitios de unión ortostéricos y alostéricos.

A continuación, se seleccionan las librerías químicas (o quimiotecas) para realizar el cribado:

- En primer lugar se utiliza una librería etiquetada formada por un conjunto de compuestos activos y no activos conocidos de LRH1, que se usará con el objetivo de determinar cuáles son los mejores métodos de acoplamiento para realizar posteriormente el cribado definitivo sobre una librería formada por compuestos naturales disponibles comercialmente.

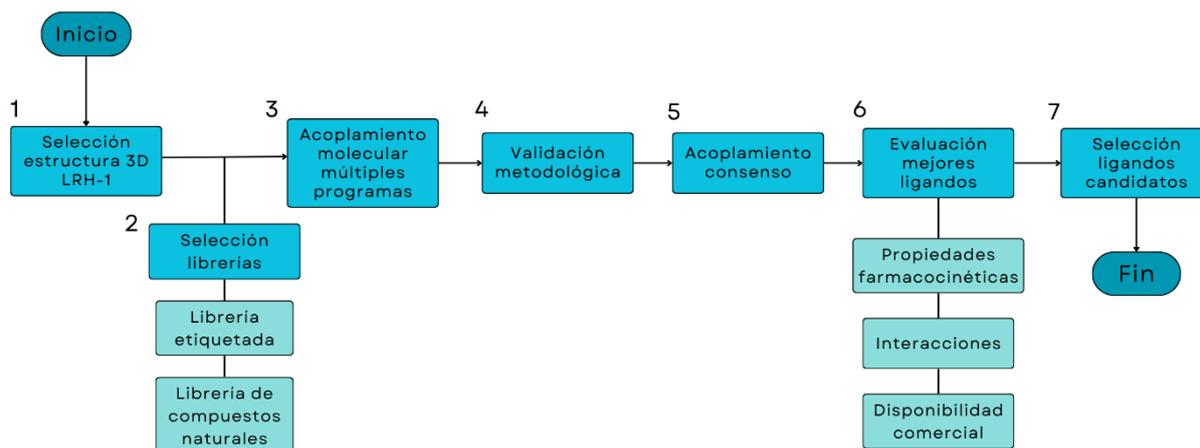
Seguidamente, se realiza el acoplamiento molecular con diferentes programas informáticos, que usan algoritmos para simular la interacción entre la diana y cada uno de los compuestos de una librería química, para encontrar la conformación óptima de cada ligando dentro del bolsillo de unión y poder evaluarla en función de un sistema de puntuación específico. Los programas de acoplamiento usados en este trabajo son: Dock6, AutoDock, Yasara-Vina, PLANTS y RxDock (ver apartados 4.6 a 4.8). Cada uno de estos programas utiliza un algoritmo de búsqueda y un sistema de puntuación distinto. Además, se usan X-Score y DSX como algoritmos que reevalúan los resultados obtenidos con AutoDock.

Tras el acoplamiento de la librería con los datos etiquetados, se realiza una validación metodológica para evaluar los programas y determinar cuáles son los más fiables para usar con la librería elegida, de la que desconocemos si los compuestos poseen actividad sobre el receptor.

Tras el acoplamiento de la librería seleccionada, se aplica el uso de métodos de acoplamiento consenso, que aumentan la fiabilidad de la búsqueda y ayudan a obtener una mayor tasa de éxito en el cribado virtual, evitando el sesgo de los programas individuales.

Por último, se evalúan los mejores ligandos obtenidos en los métodos consenso, analizando sus propiedades farmacocinéticas, su diversidad química, sus interacciones con el sitio de unión y su disponibilidad comercial. Los ligandos seleccionados se propondrán como candidatos para ser validados en ensayos experimentales *in vitro*.

En la figura 3, se muestra una representación esquemática del flujo de trabajo llevado a cabo según los pasos descritos anteriormente.



**Figura 3. Diagrama de flujo de cribado virtual en este trabajo.** 1) Selección y preparación del modelo estructural de LRH-1. 2) Selección y preparación de las librerías. 3) Acoplamiento con múltiples programas. 4) Validación metodológica y análisis estadístico para determinar los mejores métodos de acoplamiento. 5) Acoplamiento consenso con los mejores métodos de acoplamiento usando la librería de compuestos naturales. 6) Evaluación de los mejores ligandos en base a sus propiedades farmacocinéticas, interacciones y disponibilidad. 7) Selección de los ligandos candidatos.

#### 4.2. Selección de la estructura del receptor LRH-1

Al seleccionar la estructura más adecuada del receptor para el cribado virtual, es esencial considerar varios parámetros clave que determinan la calidad del modelo:

- Resolución: Mide la capacidad de detectar pequeñas diferencias en la densidad electrónica en la muestra cristalina. Se expresa como la distancia mínima entre dos átomos que pueden distinguirse claramente en la estructura cristalina. Una menor resolución indica una estructura de mejor calidad.

- Factor R: Evalúa la calidad del ajuste entre los datos experimentales y la estructura cristalina refinada. En cristalografía, se utilizan dos valores R: el factor R y el factor R libre (Rfree). El factor R se calcula comparando los datos observados con los datos del modelo refinado, mientras que el Rfree utiliza un conjunto de datos independiente excluido del refinamiento, sirviendo como medida imparcial del modelo. Idealmente, el valor de R debe estar alrededor del 20% y el Rfree no más del 5% por encima de R.

- Mapa de Ramachandran: Es una representación gráfica que muestra todas las combinaciones posibles de ángulos diédricos  $\Psi$  (psi) contra  $\Phi$  (phi) en los aminoácidos de un polipéptido, determinando la conformación de la proteína y restringida por limitaciones estéricas y electrostáticas.

- Factor B: También conocido como factor de temperatura, mide la flexibilidad de los átomos en distintas regiones de la estructura cristalina. Las regiones bien definidas tienen valores inferiores a 40.

RSRZ (real-space R-value Z-score): Es una normalización del valor R para residuos específicos a una resolución dada, con valores ideales menores que 2.

- Choques estéricos (Clashscore): Indica el número de pares de átomos en el modelo que están demasiado cerca entre sí, expresado como el número de choques por cada 1000 átomos.

Considerando estos criterios, se realizará una búsqueda en la base de datos Protein Data Bank (PDB) para seleccionar el modelo estructural más adecuado.

### 4.3. Librería de compuestos etiquetada

Una librería etiquetada consiste en una colección de compuestos activos cuya actividad frente al receptor es conocida.

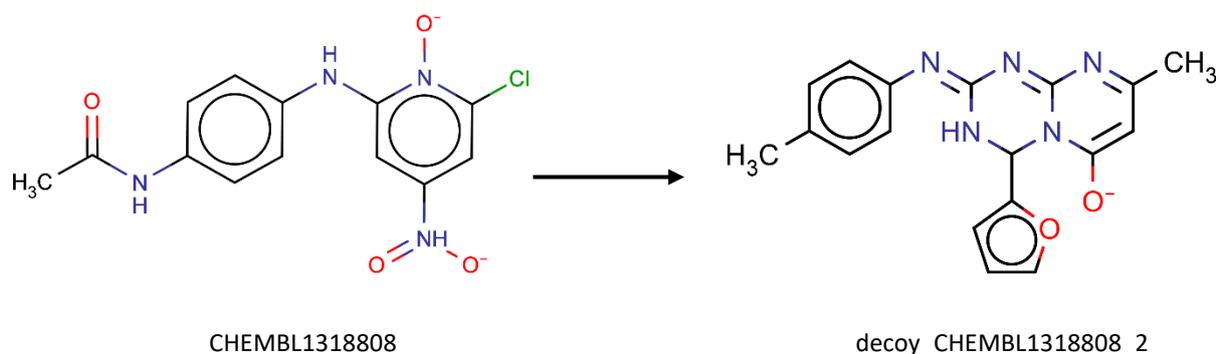
En este estudio, se han utilizado los compuestos activos experimentalmente conocidos de LRH-1 para crear una librería etiquetada como banco de pruebas para validar los ensayos de acoplamiento al receptor.

Para construir la librería, primero se accedió a la base de datos ChEMBL (<https://www.ebi.ac.uk/chembl/>) y se buscó LRH-1 como objetivo, donde se encontraron todas las moléculas con ensayos de actividad para este receptor. De este conjunto de 736 moléculas disponibles en ese momento (enero de 2024), se seleccionaron aquellas con ensayos de actividad biológica EC50 (activadores) e IC50 (inhibidores) y se eliminaron los duplicados, resultando en 100 activadores y 239 inhibidores. Del grupo de activadores, se descartaron 16 moléculas debido a la falta de actividad biológica reportada, quedando 84 activadores. Asimismo, se eliminaron 153 inhibidores con una IC50 excesivamente alta (superior a 10000 nM), quedando 86 inhibidores en total.

Para reducir el costo computacional del acoplamiento, las moléculas con mayor similitud química se agruparon mediante clustering utilizando el programa Data Warrior (<https://openmolecules.org/datawarrior/>), resultando en 22 activadores y 37 inhibidores, sumando un total de 59 compuestos activos.

La idea original de hacer la distinción entre activadores e inhibidores era poder entrenar un modelo de aprendizaje automático para el descubrimiento de nuevos fármacos, sin embargo, resultados preliminares demostraron que este método no era capaz de clasificar adecuadamente con los datos que se disponían, por lo que se abandonó la idea, pero se conservó el conjunto de moléculas.

Posteriormente, para cada una de estas moléculas, se generaron 10 señuelos (decoys). Estas moléculas señuelo, con agrupaciones químicas y propiedades fisicoquímicas similares a las moléculas activas pero con diferente topología, se consideran no activas y, teóricamente, no desarrollarían ninguna actividad frente al receptor. Los señuelos se generaron utilizando la herramienta en línea DUDE-decoys (<https://dude.docking.org>). En la figura 4 se muestra un ejemplo de un señuelo generado y su respectiva molécula original de la base de datos de ChEMBL. Después de este proceso, la librería etiquetada cuenta con un total de 649 compuestos (59 activos y 590 señuelos).



**Figura 4. Molécula señuelo generada a partir de una molécula activa.** El compuesto señuelo generado (decoy\_CHEMBL1318808\_2) posee propiedades fisicoquímicas similares a CHEMBL1318808, pero es topológicamente diferente.

#### 4.4. Validación metodológica.

La librería etiquetada se utiliza para evaluar la fiabilidad de los programas de acoplamiento en la identificación de compuestos activos frente a no activos. Esto permite determinar cuáles son los métodos de acoplamiento más efectivos para llevar a cabo, de manera confiable, el cribado virtual de la librería de compuestos naturales de la que se desconoce si los compuestos tienen actividad sobre el receptor. Al distinguir entre compuestos activos y no activos, los programas actúan como clasificadores, y las matrices de confusión se convierten en herramientas útiles para medir su rendimiento. En una matriz de confusión, los valores predichos por el modelo se comparan con los valores reales, presentados en una tabla (tabla 1), que muestra cuatro posibles resultados.

**Tabla 1. Esquema simple de una matriz de confusión.** Mide el desempeño del modelo (Predicción) con respecto a los valores conocidos (Observación).

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

VP, verdaderos positivos (TP, *true positives*): casos en los que el modelo predice correctamente un activo. FP (*false positives*), falsos positivos: casos en los que el modelo predice incorrectamente un activo. VN (TN, *true negatives*), verdaderos negativos: casos en los que se predice correctamente un no activo. FN (*false negatives*), falsos negativos: casos en los que se predice incorrectamente un no activo.

De las matrices de confusión derivan multitud de métricas como la tasa de verdaderos positivos (TPR, *true positive rate*), también llamada sensibilidad, o la tasa de falsos positivos (FPR, *false positive rate*), que se definen respectivamente como:

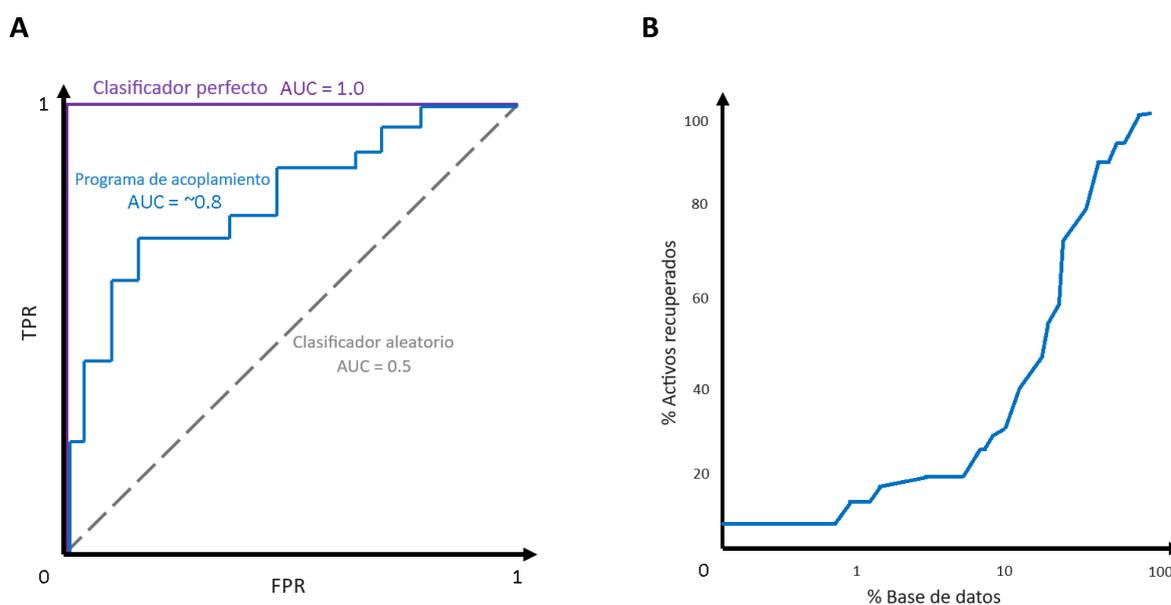
$$\text{TPR (sensibilidad)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad ; \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Utilizando estos dos parámetros, se puede evaluar la capacidad predictiva de un clasificador mediante la representación de la curva ROC (receiver operating characteristic). Esta curva muestra la tasa de verdaderos positivos (TPR) frente a la tasa de falsos positivos (FPR) para diferentes porcentajes del conjunto total de datos (porcentaje de ligandos de la librería), como se puede ver en el panel A de la figura 5.

En un escenario ideal, un clasificador perfecto tendría un área bajo la curva (AUC, *area under the curve*) de 1, lo que significa que al recorrer la librería de manera ordenada, el modelo identificaría los compuestos activos con una tasa de verdaderos positivos del 100% y una tasa de falsos positivos del 0%. Por el contrario, un modelo completamente aleatorio tendría una AUC de 0.5, representada por una línea diagonal. Los programas de

acoplamiento utilizados en este estudio se evaluarán utilizando este parámetro estadístico, que generalmente tendrá un valor entre 0.5 y 1.

Además de las curvas ROC, también se emplean las curvas de enriquecimiento (enrichment plots, EP), que representan el porcentaje de compuestos activos recuperados por cada porcentaje dado de la librería ordenada. Un EP ideal trazaría una curva donde rápidamente se recuperen todos los activos en los primeros conjuntos de datos de la librería ordenada (como se muestra en el panel B de la figura 5), indicando un rendimiento perfecto en la clasificación. Normalmente, los métodos de acoplamiento mostrarán una línea más suave, tardando más en alcanzar el 100% de los activos recuperados a medida que su rendimiento disminuye.



**Figura 5. Curvas usadas para la validación metodológica. A) Curva ROC.** Se representa la tasa de verdaderos positivos frente a la tasa de falsos positivos por cada porcentaje de la librería ordenada. El área bajo la curva (AUC) toma valores superiores conforme más preciso sea el clasificador (1, clasificador perfecto; 0.8, Programa de acoplamiento de ejemplo; 0.5 clasificador aleatorio) **B) EP.** Se representa el porcentaje de activos recuperados por cada porcentaje de la librería ordenada.

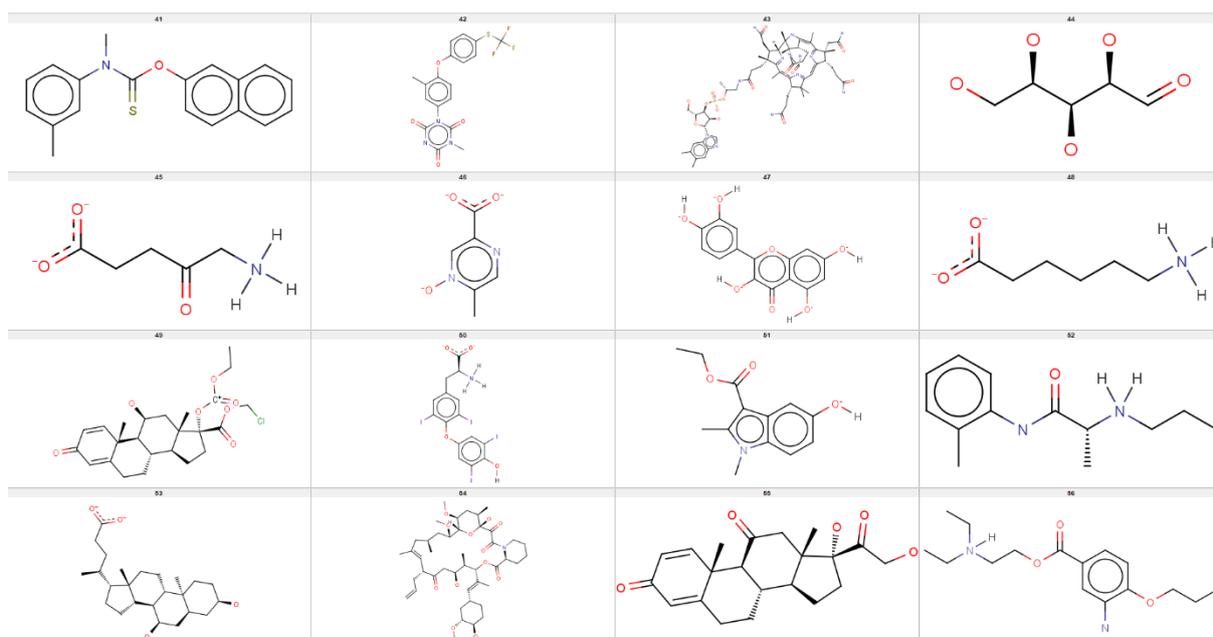
En la práctica, los resultados del cribado de la librería etiquetada se registrarán en una hoja de Excel, donde se calcularán las gráficas ROC y EP para evaluar cada método. Los resultados de esta validación metodológica con los programas de acoplamiento utilizados (Dock6, AutoDock, Yasara-Vina, X-Score, DSX, PLANTS y RxDock) se presentan en el apartado 5.2.

#### 4.5. Librería de compuestos naturales

Esta librería consiste en una colección de moléculas naturales de las cuales no se conoce si tienen actividad sobre el receptor. Para seleccionar la librería para el cribado definitivo, se parte del conocimiento de que el sitio de unión de LRH-1 es hidrofóbico y favorece la interacción con ligandos lipídicos. Sin embargo, como se mencionó anteriormente, existen ligandos que no son de naturaleza lipídica (principalmente compuestos aromáticos) y que pueden activar el receptor.

Teniendo esto en cuenta, se consideró que una elección adecuada e interesante sería una librería de compuestos naturales que incluya compuestos similares, como la librería comercial APEXBT-natural (<https://www.apexbt.com/nature-products.html>), que contiene 510 moléculas de naturaleza química muy diversa. En la figura 6 se muestran algunos de estos compuestos.

Es importante destacar que la estructura 5L11 tiene un bolsillo de unión relativamente pequeño, ya que fue cristalizada con una molécula efectora (el ligando RJW100) más pequeña que algunas de las moléculas en esta librería. Este hecho puede tener un efecto positivo al limitar las moléculas que pueden entrar en el sitio de unión, ya que las más grandes serán restringidas por impedimentos estéricos debido al reducido tamaño de este bolsillo.



**Figura 6. Compuestos de la librería APEXBT-natural.** A modo de ejemplo, se muestran las estructuras en 2D de 16 moléculas de la base de datos comercial APEXBT-natural.

#### 4.6. Acoplamiento virtual con AutoDock

AutoDock (<https://autodock.scripps.edu/>) es un programa que utiliza una técnica de acoplamiento flexible, permitiendo la flexibilidad tanto del ligando como de las cadenas laterales del sitio de unión de la proteína. Para predecir la unión de una molécula pequeña a la proteína objetivo, AutoDock emplea estrategias de búsqueda heurística global, como el algoritmo de Monte Carlo, el templado simulado o el algoritmo genético Lamarckiano.

Estas estrategias exploran el espacio de conformaciones del ligando y el receptor para encontrar de manera eficiente la conformación de mínima energía (menor  $\Delta G$ ).

Debido al alto coste computacional del cribado (en términos de tiempo de procesamiento y memoria), es preferible realizar el proceso en un servidor remoto con un sistema de computación de alto rendimiento y múltiples nodos de procesamiento que permita la ejecución en paralelo. En este caso, AutoDock se ejecutó en el servidor Horus (Linux) del laboratorio.

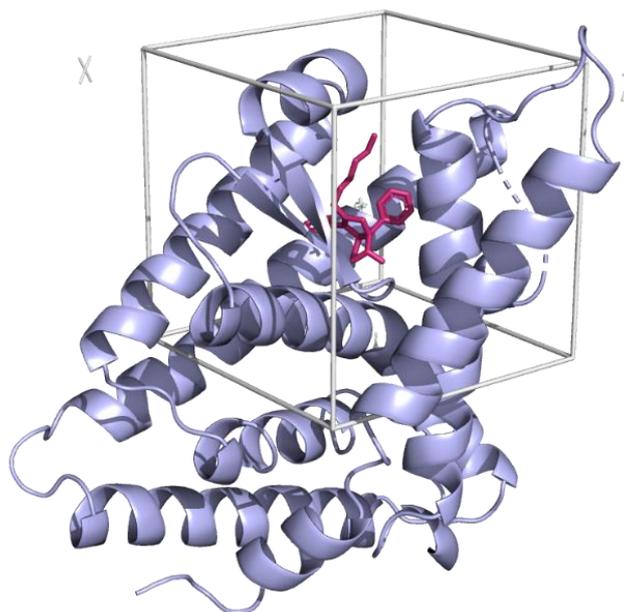
El protocolo de cribado sigue estos pasos:

1º. Preparación de la estructura del receptor (5L11). En una carpeta de trabajo se colocan los archivos receptor.pdb, receptor.pdbqt y receptor.gpf:

- El archivo receptor.pdb es simplemente el archivo PDB de la estructura, preparado recomponiendo residuos incompletos, eliminando moléculas de agua y añadiendo hidrógenos y cargas parciales a todos los átomos implicados.

- El archivo receptor.pdbqt se genera con el programa AutoDockTools (ADT) a partir del archivo receptor.pdb. Este formato contiene información adicional sobre la carga parcial atómica y los tipos de átomos.

- El archivo receptor.gpf también se genera con ADT. Este archivo contiene información sobre la caja de simulación, una rejilla que delimita la región espacial donde se llevará a cabo la búsqueda conformacional. En la figura 7 se muestra cómo se sitúa esta caja de simulación en el receptor y se verifica que está bien definida.



**Figura 7. Caja de simulación de AutoDock.** Se muestra el receptor en color azul y el ligando en color rosa. La caja se define de forma que queda centrada con respecto al ligando. Dimensiones de la caja: (x = 25, y = 25, z = 25 Å).

2º. Preparación de los ligandos. Todos los ligandos se colocan individualmente en formato mol2 en la carpeta de trabajo, junto con el ligando de referencia (RJW100) de forma aislada.

3º. Preparación y ejecución de los scripts de comandos. Se ejecutan varios scripts de Python, cada uno encargado de una parte específica del proceso de acoplamiento. La función de estos scripts se resume brevemente en la tabla 2.

**Tabla 2. Guiones de comandos (*scripts*) usados en el protocolo de cribado virtual con AutoDock.** Los *guiones de comandos* se ejecutan ordenadamente en una terminal de Linux, usando un entorno de Python2.7 (Conda).

Guiones de comandos	Función
<b>prepara_VS_cluster_Horus.py</b>	Genera archivos que contienen las energías de interacción precalculadas entre la proteína y cada uno de los átomos de los ligandos. Ahorra tiempo de cálculo durante el acoplamiento.
<b>launch_jobs_sh.py</b>	Manda los trabajos a la cola de ejecución del servidor para que sean calculados. Cuando un trabajo se completa, se genera un archivo que contiene información detallada sobre el acoplamiento (energía de unión, posición y orientación del ligando).
<b>Comprueba_Progreso_VS.py</b>	Comprueba el progreso para saber cuándo han acabado todos los cálculos.
<b>Recoleccion_Datos_VS_Horus.py</b>	Recopila los datos al finalizar el cribado. Crea un archivo con todos los ligandos ordenados por sus respectivas energías de interacción con el receptor.

#### 4.7. Acoplamiento virtual con Yasara

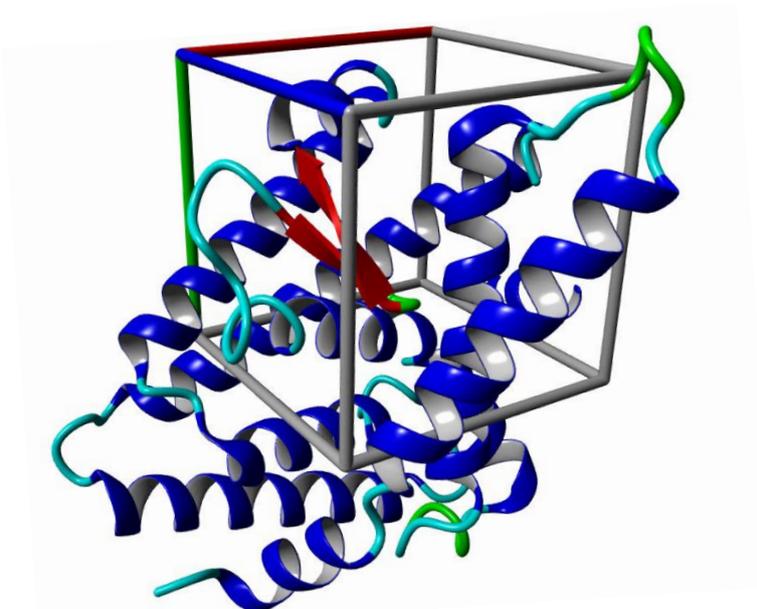
Yasara (<http://www.yasara.org/>) es un programa que permite realizar el acoplamiento proteína-ligando, similar a AutoDock. Utiliza algoritmos de acoplamiento molecular para predecir la orientación y la afinidad de unión de las moléculas a la proteína objetivo. En este trabajo se usa específicamente VINA (Arthur J. Olson et al., 2010), un algoritmo que emplea una búsqueda global seguida de un refinamiento local para encontrar la mejor conformación de cada ligando.

Debido a los altos recursos computacionales requeridos, el proceso se lleva a cabo en un servidor remoto, en este caso, el servidor Osiris del laboratorio.

Para realizar el cribado virtual, se sigue este procedimiento ordenado:

1º. Preparación del receptor y los ligandos. Se crea un directorio de trabajo donde se colocan:

- 5L11\_ligands.sdf. Un archivo sdf múltiple que contiene todos los ligandos.
- 5L11\_receptor.sce. El archivo del receptor en formato de escena de Yasara (.sce), que contiene dos objetos: el receptor.pdb preparado y la caja de simulación (idéntica a la utilizada en AutoDock). La escena se muestra en la figura 8.



**Figura 8. Caja de simulación de Yasara.** Los parámetros que definen la caja son idénticos a los usados en el protocolo de AutoDock (Tamaño de la caja X: 25, Y: 25, Z: 25 Å, centrada en el sitio definido por la unión del ligando).

2º. Acoplamiento. Para realizar el acoplamiento molecular, se selecciona el archivo 5L11\_receptor.sce preparado anteriormente. Luego, se ejecuta el archivo dock\_runscreening.mcr, un script de Yasara que contiene el código necesario para llevar a cabo todo el proceso de búsqueda y cálculo.

Al finalizar el acoplamiento, se obtienen los siguientes archivos importantes para este trabajo:

- 5L11.log: Un archivo de texto que contiene toda la información sobre las energías de unión. En Yasara, el sistema de puntuación funciona a la inversa: energías más positivas indican uniones más fuertes, mientras que energías más negativas indican peores uniones.
- Archivos yob: Son escenas de Yasara que contienen individualmente la mejor conformación de cada ligando acoplado al receptor.

#### 4.8. Otros programas de acoplamiento.

Además de AutoDock y Yasara, este trabajo también emplea los programas Dock6, X-Score, DSX, PLANTS y RxDock, que se describen brevemente a continuación:

**Dock6** ([https://dock.compbio.ucsf.edu/DOCK\\_6/index.htm](https://dock.compbio.ucsf.edu/DOCK_6/index.htm)): Utiliza un algoritmo de construcción incremental (anchor-and-grow) y búsqueda aleatorizada para el acoplamiento, empleando la coincidencia geométrica para superponer el ligando en una imagen negativa del bolsillo de unión. Este algoritmo ha mejorado con el tiempo, incorporando características importantes como la puntuación basada en campos de fuerza y la flexibilidad del ligando y el receptor.

**PLANTS** (<http://www.tcd.uni-konstanz.de/research/plants.php>): Basado en algoritmos de optimización estocástica inspirados en el comportamiento de las colonias de hormigas, que encuentran el camino más corto

entre su nido y una fuente de alimento. En el acoplamiento proteína-ligando, una colonia de hormigas artificiales encuentra la conformación de energía mínima del ligando en el sitio de unión.

**RxDock** (<https://rxdock.gitlab.io/>): Un programa rápido y versátil para acoplar ligandos a proteínas y ácidos nucleicos, diseñado específicamente para cribado virtual. Similar a AutoDock, calcula previamente las energías de interacción usando una rejilla para reducir el tiempo de cálculo posterior.

**X-Score** ([https://www.ics.uci.edu/~dock/manuals/xscore1.1\\_manual/usage.html](https://www.ics.uci.edu/~dock/manuals/xscore1.1_manual/usage.html)): Utiliza una función de puntuación empírica para calcular la unión ligando-receptor a partir de los acoplamientos obtenidos en AutoDock.

**DSX** (<https://agklebe.pharmazie.uni-marburg.de/?id=11>): Emplea una función de puntuación basada en el conocimiento, considerando interacciones por parejas dependientes de distancia, ángulo de torsión y superficie accesible al solvente. Al igual que X-Score, DSX utiliza los datos de acoplamiento obtenidos con AutoDock.

#### 4.9. Métodos de acoplamiento consenso

Todos los programas de acoplamiento mencionados anteriormente utilizan enfoques diferentes para puntuar la unión entre el ligando y el receptor, lo que resulta en listas de moléculas clasificadas según sistemas de puntuación distintos. A pesar de estas diferencias, no se puede considerar que un programa sea mejor que otro, lo que dificulta la selección de las mejores moléculas para pruebas experimentales. Para resolver este problema, se emplean métodos de consenso.

Los métodos de acoplamiento consenso buscan mejorar los resultados del cribado virtual (Blanes-Mira et al., 2022; Palacio-Rodríguez et al., 2019) promediando las puntuaciones de las moléculas individuales obtenidas en los diferentes programas de acoplamiento. En este trabajo se han utilizado los siguientes métodos de acoplamiento consenso: relación de puntuación normalizada (NSR), clasificación por número (RBN), promedio de puntuación autoajustada (AASS), z-score, clasificación por rango (RBR) y clasificación por voto (RBV). A continuación, se describe el procedimiento seguido para calcular cada uno de ellos.

Relación de puntuación normalizada (NSR): Se basa en normalizar los valores de energía en función del valor del ligando de referencia. Para cada uno de los programas de acoplamiento utilizados, se divide el valor obtenido para cada molécula entre el valor del ligando de referencia para ese programa y luego se hace la media entre todos ellos. En otros términos:

$$NSR_i = \frac{1}{n} \sum_j \frac{s_i^j}{r^j}$$

$s_i^j$  representa la puntuación de la molécula  $i$  en el programa de acoplamiento  $j$ ;  $r^j$  es la puntuación del ligando de referencia en el programa  $j$  y  $n$  el número de programas de acoplamiento utilizados.

Clasificación de consenso exponencial (*Exponential consensus ranking*; ECR): Se asigna una puntuación exponencial a cada molécula ( $i$ ) para cada programa ( $j$ ) utilizando el rango de la molécula ( $r$ ) proporcionado por cada programa de acoplamiento individual,

$$p(r_i^j) = \frac{1}{\sigma} e^{-\frac{r_i^j}{\sigma}}$$

Donde  $\sigma$  es el valor esperado de la distribución exponencial. Este parámetro establece el número de moléculas para cada función de puntuación que se considerará, es decir, el umbral del conjunto de datos que se tomará en cuenta para el consenso. La puntuación final de cada molécula  $i$  se define como la suma de la puntuación exponencial de todas las funciones de puntuación  $j$ :

$$ECR_i = \sum_j p(r_i^j) = \frac{1}{\sigma} \sum_j \exp\left(-\frac{r_i^j}{\sigma}\right)$$

Clasificación por número (*rank-by-number*; RBN): La puntuación de cada molécula se corresponde con la media de los valores obtenidos en cada uno de los programas de acoplamiento:

$$RBN_i = \frac{1}{n} \sum_j s_i^j$$

Promedio de puntuación autoajustada (*auto-scaled-scores*; AASS): Consiste en normalizar los valores de energía entre 0 y 1. Para ello, la puntuación de cada molécula se ajusta según la puntuación mínima ( $s_{min}^j$ ) y máxima ( $s_{max}^j$ ) en cada programa. La puntuación final viene dada por la media de todas las puntuaciones normalizadas, de forma que:

$$AASS_i = \frac{1}{n} \sum_j \frac{s_i^j - s_{min}^j}{s_{max}^j - s_{min}^j}$$

Z-score: La puntuación de las moléculas ( $s_i^j$ ) se normaliza usando la media de los resultados de cada programa ( $\mu^j$ ) y su desviación estándar ( $\sigma^j$ ). De nuevo, la puntuación final es el promedio de las puntuaciones normalizadas entre todos los programas usados:

$$Z - score_i = \frac{1}{n} \sum_j \frac{s_i^j - \mu^j}{\sigma^j}$$

Clasificación por rango (*rank-by-rank*; RBR): En este método, se ordenan las moléculas en cada programa y se les asigna el número según la posición o *ranking* ( $r_i^j$ ) que ocupan en la lista total.

$$RBR_i = \frac{1}{n} \sum_j r_i^j$$

Clasificación por voto (*rank-by-vote*; RBV): En esta estrategia, las moléculas reciben un voto (+1) si están situadas en el top x% (10% en el caso de este trabajo) de los resultados de cierto programa de acoplamiento, de forma que al ser sumados todos los votos se obtendrá una puntuación entre 0 y el número de programas usados.

#### 4.11. Evaluación de propiedades farmacocinéticas

Los fármacos que son administrados por vía oral sufren una serie de procesos comúnmente conocidos en farmacología con el acrónimo ADME (absorción, distribución, metabolismo y excreción) a los que el fármaco se verá sometido cuando entre en el organismo (Doogue et al., 2013). Para que una molécula pueda finalmente convertirse en un fármaco, debe presentar una farmacocinética favorable (biodisponibilidad suficiente para alcanzar la diana). Por ello, las moléculas con mejor puntuación, seleccionadas tras el cribado virtual y el análisis de consenso, deben someterse a una evaluación adicional basada en criterios farmacocinéticos para mejorar la probabilidad de su eficacia terapéutica futura.

La regla del 5 de Lipinski es una regla empírica, basada en la observación de las características fisicoquímicas de muchos fármacos, que permite predecir cualitativamente si un principio activo podría ser suministrado oralmente de forma satisfactoria (Benet et al., 2016). Según esta regla (tabla 3), una molécula debe cumplir con los siguientes requisitos: (1) Tener un peso molecular menor de 500 g/mol, lo que indica que no debe ser excesivamente grande. (2) Poseer un coeficiente de partición octanol-agua (logP) inferior a 5, lo cual garantiza una hidrofobicidad suficiente para atravesar membranas biológicas sin comprometer su solubilidad en agua. (3) No tener más de cinco donadores de enlaces de hidrógeno (átomos de nitrógeno u oxígeno con al menos un hidrógeno). (4) No tener más de diez aceptores de enlaces de hidrógeno (átomos de nitrógeno, oxígeno o flúor).

**Tabla 3. Regla del 5 de Lipinski.** Propiedades fisicoquímicas que poseen frecuentemente las moléculas con actividad farmacológica.

Parámetro	Valor
PM (g/mol)	< 500
logP	< 5
HBD	< 5
HBA	< 10

PM: peso molecular, logP: coeficiente de partición octanol-agua, HBD: donadores de hidrógeno, HBA: aceptores de hidrógeno.

Además, también se examinarán tres propiedades farmacocinéticas clave: la absorción gastrointestinal (GIA, *gastrointestinal absorption*), la permeabilidad de la barrera hematoencefálica (BBB, *blood-brain barrier*), y si la molécula es sustrato de la glicoproteína-P (PGP, *P-glycoprotein*). Cada uno de estos parámetros puede tener diferentes valores, atendiendo a los criterios mostrados en la tabla 4 (Zoete et al., 2016).

Para GIA, un valor de 0 indica ausencia de absorción intestinal; 1, absorción baja; y 2, absorción alta. En el caso de BBB, un valor de 0 significa que el compuesto no atraviesa la barrera hematoencefálica, mientras que un valor

de 1 indica que sí lo hace. De manera similar, para PGP, un valor de 0 indica que el compuesto no es sustrato de la glicoproteína-P y un valor de 1 señala que lo es.

**Tabla 4. Parámetros farmacocinéticos adicionales.**

Parámetro	Valor		
GIA	0	1	2
BBB	0	1	
PGP	0	1	

GIA: absorción gastrointestinal, BBB: permeabilidad de la barrera hematoencefálica, PGP: sustrato de glicoproteína-P.

Por último, se puede caracterizar la solubilidad de los ligandos, que también dará una idea de la biodisponibilidad oral del fármaco y cómo de sencillo será manejarlo durante su síntesis y durante el resto de fases del desarrollo farmacológico. De entre los diferentes métodos que existen para estimar la solubilidad, se usará el calculado por ESOL (*Estimated Solubility*), (Delaney, 2004), cuya escala se indica en la tabla 5:

**Tabla 5. ESOL, solubilidad estimada.** Se calcula a partir de un método simple que estima la solubilidad acuosa de un compuesto directamente a partir de su estructura.

< -10	Insoluble
-10 a -6	Poco soluble
-6 a -4	Moderadamente soluble
-4 a -2	Soluble
-2 a 0	Muy soluble
> 0	Altamente soluble

#### 4.12. Programas accesorios de edición molecular

Se han utilizado varias herramientas que han sido de utilidad para desempeñar diversas funciones específicas. Estos son:

- PyMol (<https://pymol.org/2/>) como programa de visualización, edición y modelado molecular. Versión 2.5.4.
- OpenBabel ([http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)) para la conversión entre diferentes formatos de archivos de moléculas (pdb, sdf, mol2) y también para agrupar varias moléculas en un único archivo. Versión 3.1.1.
- MarvinView (<https://chemaxon.com/marvin>) como herramienta de visualización. Versión 6.0

- AutoDockTools (<https://autodocksuite.scripps.edu/adt/>) para la preparación y generación de archivos específicos (pdbqt, gpf), como se ha explicado en el apartado 4.6. Versión 1.5.6.
- Bitvise SSH (<https://www.bitvise.com/>) para realizar la conexión remota al servidor Horus. Versión 9.24.
- Sublime Text (<https://www.sublimetext.com/>) como editor de texto y código. Versión 4.1.4.3.

## 5. Resultados

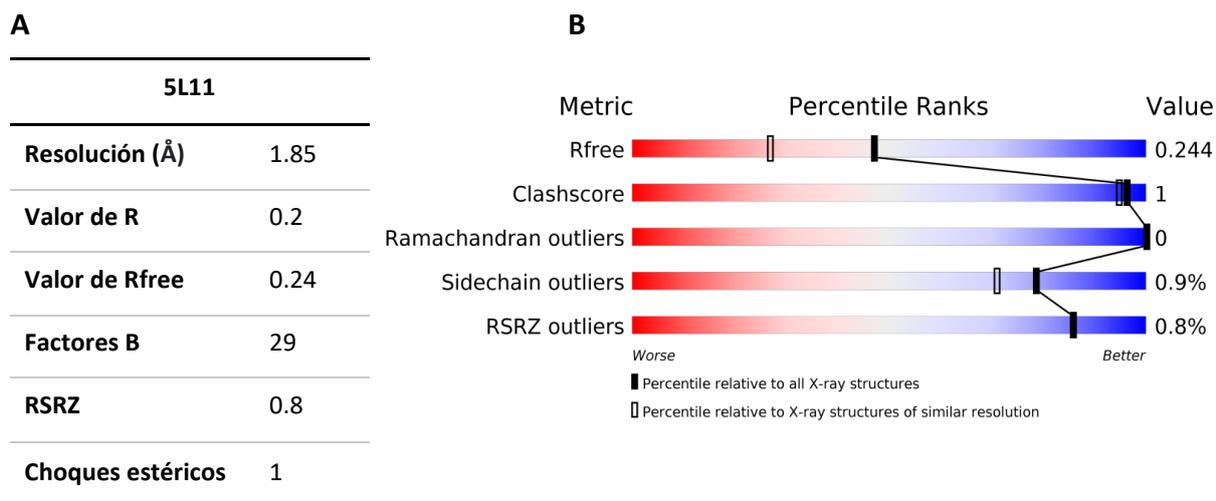
### 5.1. Selección de la estructura tridimensional del receptor LRH-1

En la base de datos PDB (Protein Data Bank) se dispone de varias estructuras del LRH-1 de *Homo sapiens* (ver tabla 6). Basándose en los criterios establecidos en el apartado 4.2, se ha decidido seleccionar la estructura con el código PDB 5L11. Este modelo, obtenido mediante cristalografía de rayos X, presenta las métricas especificadas en la figura 9 (derivadas del informe completo de validación de la estructura).

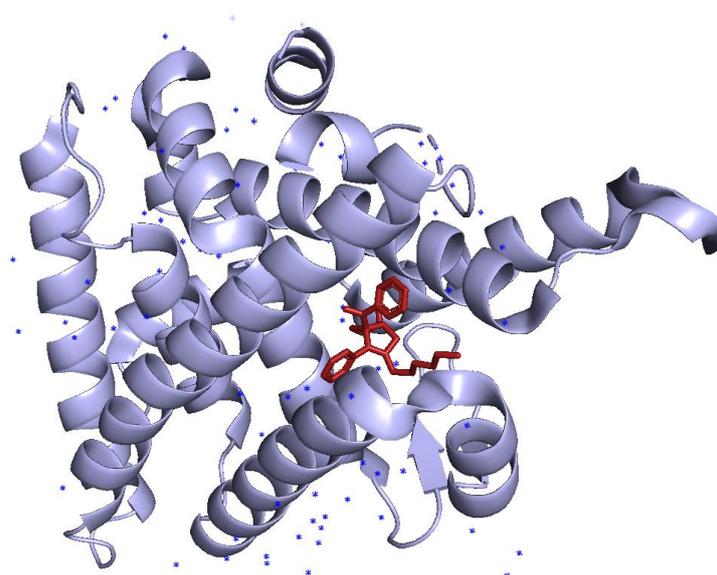
**Tabla 6. Estructuras del receptor LRH-1 en la base de datos PDB.** Se muestran 5 modelos resueltos por difracción de rayos X, donde se puede apreciar que 5L11 tiene la mejor resolución de todos ellos.

PDB	Resolución (Å)	Valor de R
5L11	1.85	0.24
5SYZ	1.93	0.23
6VIF	2.26	0.24
1ZDU	2.50	0.24
1YOW	3.00	0.25

Como se muestra en el panel A de la figura 9, los valores cumplen con todos los requisitos establecidos para la selección de estructuras de alta calidad (ver sección 4.2). La figura 9B ilustra visualmente que la estructura seleccionada es de alta calidad. Respecto al mapa de Ramachandran, el informe indica que ninguno de los 242 residuos analizados (que representan el 93% del total) se encuentra en regiones prohibidas o atípicas. En la figura 10 se muestra la estructura tridimensional del receptor LRH-1 (5L11.pdb) en complejo con un ligando orgánico.



**Figura 9. Parámetros y métricas estructurales del modelo 5L11. A)** Parámetros de calidad de estructura (Resolución, Valor de R y Rfree, factores B, RSRZ y choques estéricos). **B)** Métricas en rango de percentiles (*Percentile Ranks*). La gráfica proviene de la base de datos PDB y muestra algunas de las métricas situadas en un rango de percentiles, de forma que los valores situados en la zona azul se consideran mejores que los situados en la zona roja. *Ramachandran outliers* hace referencia a los residuos en zonas prohibidas del mapa de Ramachandran y *sidechain outliers* hace referencia a conformaciones atípicas de las cadenas laterales.



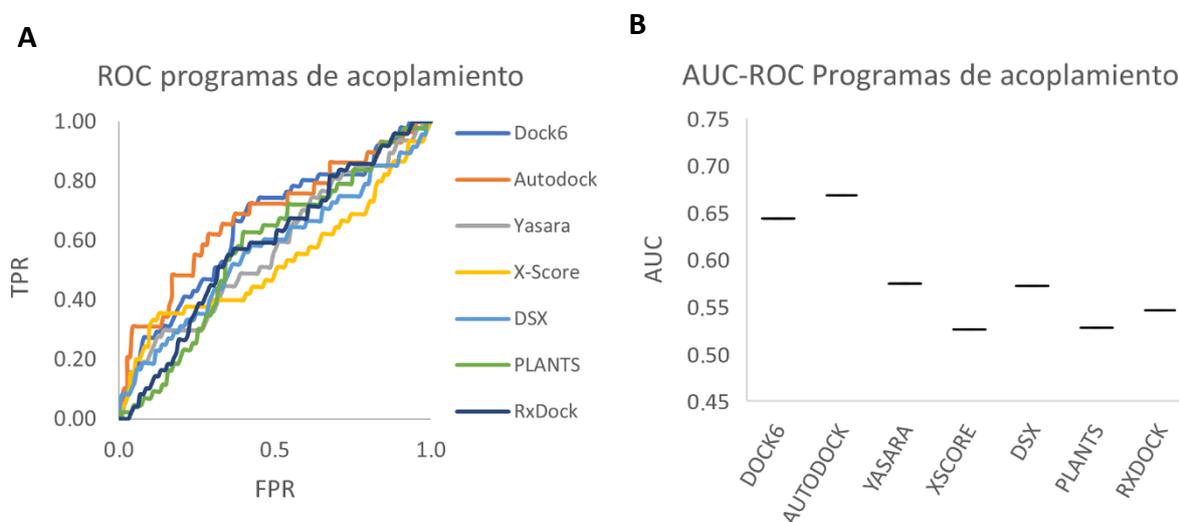
**Figura 10. Estructura tridimensional del modelo 5L11.** La estructura incluye el ligando RJW100 (rojo) que marca el sitio de unión del receptor y múltiples moléculas de agua (azul). Tanto el ligando como las moléculas de agua se eliminan para realizar el cribado virtual.

## 5.2. Análisis de los resultados de acoplamiento molecular. Curvas ROC y EP

Para evaluar la precisión de los programas de acoplamiento, se realizó un cribado virtual utilizando la librería etiquetada (ver sección 4.4). El proceso incluyó el acoplamiento de esta librería con todos los programas (Dock6, AutoDock, Yasara, X-Score, DSX, PLANTS y RxDock), y los resultados se consolidaron en una hoja de Excel. En esta hoja, se listan en una columna las identificaciones de cada ligando, y en columnas adyacentes, las puntuaciones otorgadas por cada programa (una columna por método), considerando las siguientes especificaciones: Algunos ligandos no pudieron ser procesados por todos los programas; estos se excluyeron del análisis. En total, la librería contiene 57 activos (56 de ChEMBL y el ligando de referencia) y 589 señuelos (totalizando 646 compuestos).

Además de la columna de ligandos, se incluye una columna adicional que indica la actividad de cada compuesto (1 para activos y 0 para señuelos). Esta columna se utiliza para contar los activos al calcular la matriz de confusión y los factores de enriquecimiento.

Posteriormente, las moléculas se ordenan según la puntuación obtenida en cada programa utilizando la hoja de cálculo, y se calcula la matriz de confusión (TP, FP, FN, TN) para distintos conjuntos de datos de la librería, en fracciones que van del 0.1% al 100% de la librería. Con las tasas de verdaderos positivos (TPR) y falsos positivos (FPR) para cada porcentaje, se generan las curvas ROC y se calcula el área bajo la curva (AUC) para evaluar la fiabilidad de cada programa, como se muestra en la figura 11.

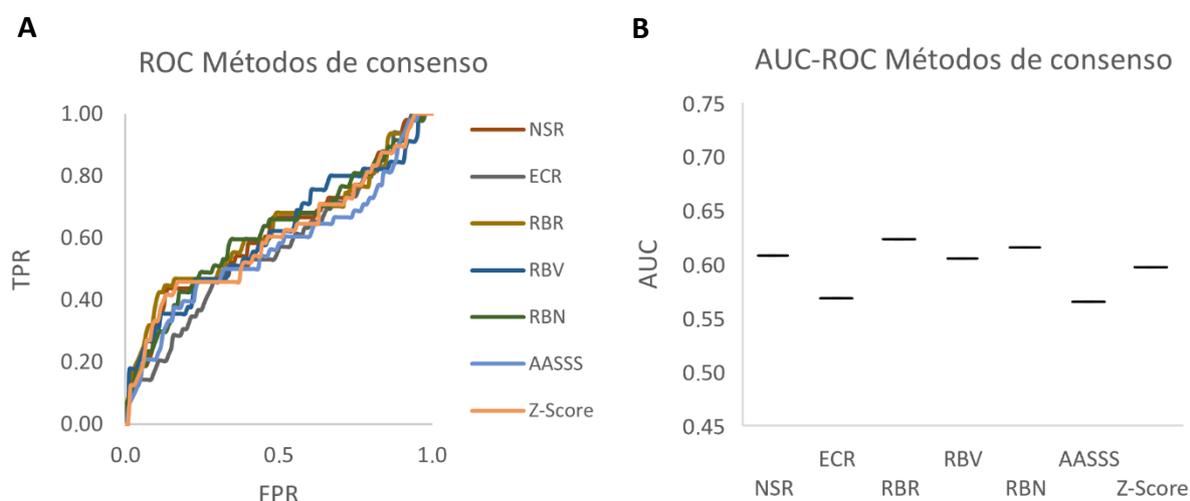


**Figura 11. Curvas ROC y AUC-ROC de los programas de acoplamiento usados. A)** Curvas ROC para todos los programas de acoplamiento. Cada una de las líneas coloreadas representa un programa. **B)** AUC de todos los programas de acoplamiento. Cada pequeña línea de color negro representa los valores obtenidos por cada programa.

Atendiendo al área bajo la curva ROC (AUC-ROC), se puede argumentar que ninguno de los programas parece comportarse especialmente bien en este sistema, siendo AutoDock (AUC = 0.67), el que mejor distingue activos de no activos, mientras que programas como PLANTS (AUC = 0.53) y X-Score (AUC = 0.53) son los que peor rinden según esta métrica. Dado que no hay un programa que destaque significativamente sobre los demás con un

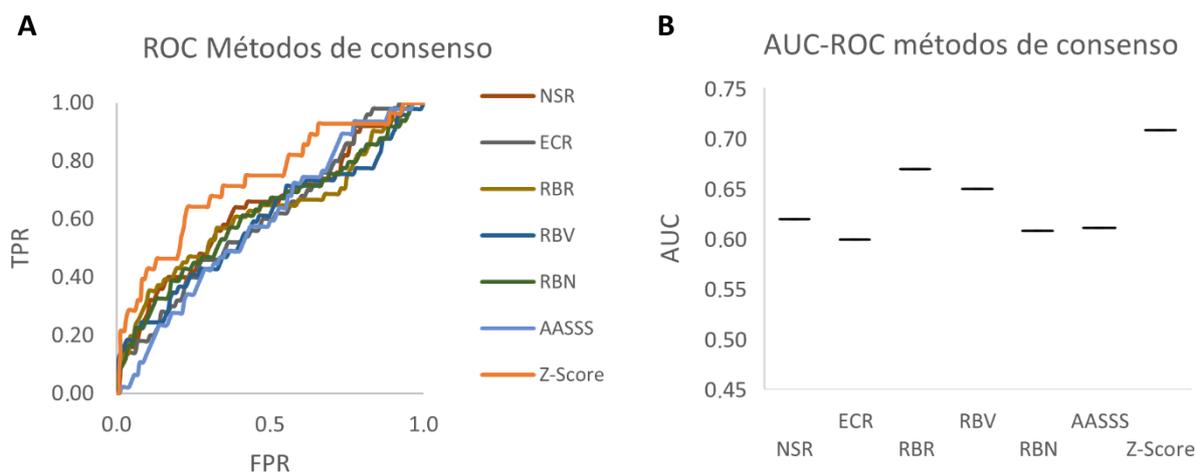
desempeño y fiabilidad alta, el cribado virtual de la librería de compuestos naturales se realizará con todos los programas de acoplamiento.

Seguidamente se obtuvieron las curvas ROC y sus respectivas AUC para los métodos de acoplamiento consenso (NSR: *normalized score ratio*, ECR: *exponential consensus ranking*; RBR: rank-by-rank, RBV: rank-by-vote, RBN: rank-by-number, AASS: *average of auto-scaled score* y Z-Score). En teoría, estos métodos deberían mejorar las métricas, sin embargo, tal y como se muestra en la figura 12, en este caso los métodos de consenso no parecen mejorar la fiabilidad del cribado, ya que todos tienen un área bajo la curva (AUC) muy similar a los programas de acoplamiento individuales. Si bien el método RBR se comporta ligeramente mejor que el resto de métodos, ninguno destaca significativamente sobre los demás.



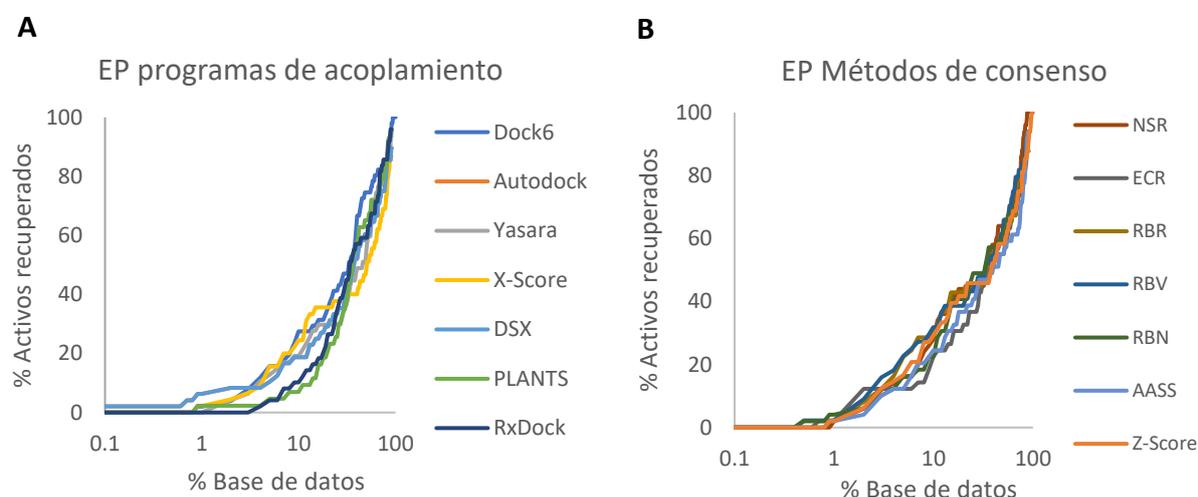
**Figura 12. Curvas ROC y AUC-ROC de los métodos de acoplamiento consenso. A)** Curvas ROC para todos los métodos consenso. Cada una de las líneas coloreadas representa un método. **B)** AUC de todos los métodos consenso. Cada pequeña línea de color negro representa un método.

Con el objetivo de optimizar estos resultados, se decide eliminar algunos programas de la ponderación sobre el consenso, sobre todo los menos fiables (AUC-ROC cercano a 0.5), como los programas PLANTS y X-Score. Una vez eliminados, los cálculos obtenidos del acoplamiento consenso son más fiables, y servirán para hacer la selección final de compuestos. En la figura 13, se puede apreciar que al excluir del consenso los programas PLANTS y X-Score, ahora todos los métodos rinden un AUC superior a 0.6, además, se observa como el método Z-Score posee un rendimiento superior al resto de métodos (AUC = 0.71), siendo este el más fiable, seguido del método RBR (AUC = 0.68).



**Figura 13. Curvas ROC y AUC-ROC de los métodos de acoplamiento consenso excluyendo los programas PLANTS y X-Score.** **A)** Curvas ROC para todos los métodos consenso. Cada una de las líneas coloreadas representa un método. **B)** AUC de todos los métodos consenso. Cada pequeña línea de color negro representa un método.

A continuación, con los mismos porcentajes de la librería usados en las curvas ROC y el número de activos recuperados para cada uno de ellos, se pueden calcular las gráficas de enriquecimiento (*Enrichment Plots*) para los 7 programas y para los métodos consenso, tal y como se representa en la figura 14.



**Figura 14. EP, representación semilogarítmica.** **A)** EP de los programas de acoplamiento. Cada una de las líneas coloreadas representa un programa. **B)** EP de los métodos de acoplamiento consenso. Cada una de las líneas coloreadas representa un método.

Al igual que se observa con las curvas ROC, los programas individuales no parecen rendir de forma óptima y ninguno destaca especialmente en su capacidad para recuperar los activos de la librería etiquetada. Por otra

parte, los EP para los métodos de acoplamiento consenso muestran una ligera mejora. Los métodos de consenso excluyendo los programas PLANTS y X-Score (no mostrado), muestran una mejora muy sutil.

### 5.3. Acoplamiento consenso

Tal y como se ha determinado en la validación metodológica del apartado anterior, el cribado virtual se realizará con todos los programas y la librería APEXBT-natural.

Los resultados de los programas de acoplamiento se recopilan y se emparejan en una hoja Excel. A continuación, se calculan los métodos consenso, obteniendo los resultados que se muestran en la figura 15. Los resultados se ordenaron inicialmente por el método de consenso Z-Score, ya que es el que mejor rendimiento obtenía con la librería etiquetada, sin embargo, esta clasificación no mostró ningún ligando que puntuara mejor que el ligando cristalográfico, por lo tanto, se optó por realizar la clasificación de acuerdo con el segundo mejor método, es decir, RBR (*Rank-by-rank*). Los programas PLANTS y X-Score no se tuvieron en cuenta para el consenso.

El análisis de los resultados indica que hay múltiples ligandos que puntúan por encima o de forma muy similar al ligando de referencia en este método de consenso (algunos de ellos son: Apexbt\_0\_252, Apexbt\_0\_106, Apexbt\_0\_194, Apexbt\_0\_329, Apexbt\_0\_353, etc). Este hecho pone de manifiesto que hay compuestos con una gran afinidad teórica por el receptor LRH-1, por lo que podrían tener potencial como fármacos candidatos.

ID	Métodos de acoplamiento consenso							
	NSR	ECR	RBR	RBV	RBN	AASS	Z-Score	
apexbt_0_252	1.12	0.137	13.6	5	-50.0	0.219	-0.58	
apexbt_0_106	1.04	0.068	27.8	5	-43.5	0.222	-0.51	
apexbt_0_194	1.07	0.108	29.0	4	-42.9	0.238	-0.52	
apexbt_0_329	1.14	0.100	34.8	3	-44.5	0.217	-0.50	
apexbt_0_353	1.15	0.101	37.4	3	-45.6	0.211	-0.56	
apexbt_0_509	1.00	0.065	37.8	4	-42.3	0.226	-0.51	
apexbt_0_336	1.06	0.088	39.0	4	-42.4	0.217	-0.51	
apexbt_0_240	0.99	0.041	45.0	3	-41.4	0.234	-0.46	
apexbt_0_179	1.00	0.070	45.2	3	-42.4	0.230	-0.54	
apexbt_0_244	1.03	0.064	50.2	2	-40.7	0.229	-0.45	
apexbt_0_346	0.98	0.038	55.0	3	-42.5	0.226	-0.48	
apexbt_0_402	1.01	0.096	59.6	3	-45.4	0.218	-0.51	
apexbt_0_325	1.03	0.042	60.0	2	-41.7	0.228	-0.48	
apexbt_0_263	1.07	0.069	63.0	3	-42.2	0.241	-0.46	
apexbt_0_366	1.06	0.085	64.8	3	-40.2	0.220	-0.46	
apexbt_0_403	0.97	0.051	65.6	3	-43.3	0.214	-0.49	
apexbt_0_130	0.94	0.014	66.6	2	-38.3	0.240	-0.43	
apexbt_0_345	0.96	0.050	68.2	3	-37.6	0.249	-0.46	
<b>ligando cristalográfico</b>	1.00	0.128	69.4	4	-46.4	0.214	-0.55	

**Figura 15. Resultados del acoplamiento consenso.** Detalle de la hoja Excel con los resultados del acoplamiento consenso. Se ha ordenado de menor a mayor por las puntuaciones del método de consenso RBR (*rank-by-rank*), excluyendo de la ponderación los programas PLANTS y X-Score. La columna ID indica la identificación del ligando. El ligando cristalográfico se destaca en rojo, puntuado con un valor de 1 en el método NSR.

#### 5.4. Ligandos seleccionados. Propiedades farmacocinéticas.

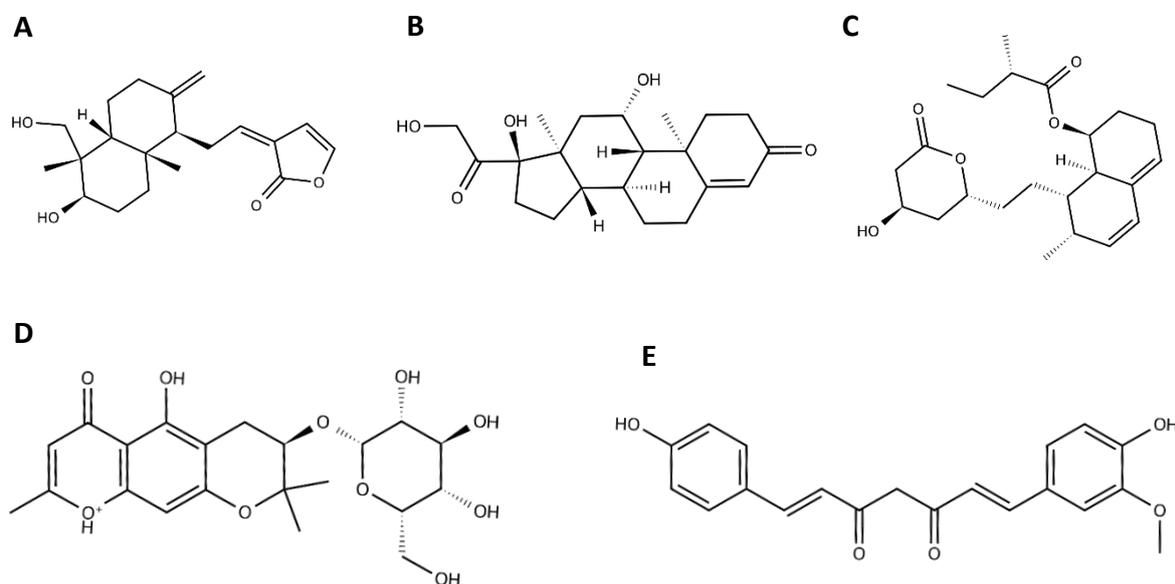
La selección de los ligandos en el apartado anterior se ha hecho en base a la puntuación obtenida mediante los métodos de acoplamiento consenso, sin embargo, la puntuación por sí sola no es suficiente para predecir la eficacia final de un compuesto, por lo que también se deben considerar factores farmacocinéticos, que se analizan a continuación.

**Tabla 7. Parámetros farmacocinéticos de los ligandos candidatos.** Propiedades extraídas de la base de datos APEXBT\_natural. Se destacan en fondo azul los ligandos seleccionados con las mejores propiedades.

Ligando	PM (g/mol) [< 500]	logP [< 5]	HBD [< 5]	HBA [< 10]	GIA [= 2]	BBB [= 0]	PGP [= 0]	ESOL [> -6]
Apexbt_0_252	494.5	-1.2	6	11	1	0	0	-2.19
Apexbt_0_106	390.51	3.95	1	5	2	1	0	-4.22
Apebt_0_194	346.46	2.67	2	4	2	1	1	-3.08
Apexbt_0_329	428.43	-2.02	6	10	1	0	0	-0.93
Apexbt_0_353	493.44	3.34	7	9	1	0	0	-5.71
Apexbt_0_509	408.4	-0.18	4	8	1	0	1	-1.5
Apexbt_0_336	515.44	1.03	7	11	0	0	0	-4.2
Apexbt_0_240	332.43	3.12	2	4	2	1	0	-3.83
Apexbt_0_179	442.37	2.53	7	10	1	0	1	-3.7
Apexbt_0_244	480.46	-1.51	5	11	1	0	1	-1.85
Apexbt_0_346	390.38	0.45	6	8	1	0	0	-2.9
Apexbt_0_402	492.47	-1.33	5	11	1	0	0	-1.79
Apexbt_0_325	338.35	3.36	2	5	2	0	0	-3.92
Apexbt_0_263	438.43	0.07	5	9	1	0	0	-2.44
Apexbt_0_366	457.43	-3.11	7	12	0	0	0	-0.65
Apexbt_0_403	512.46	-2.01	6	13	1	0	0	-1.75
Apexbt_0_130	362.46	1.78	3	5	2	0	1	-2.97
Apexbt_0_345	445.35	0.28	6	9	1	0	0	-3.46

**PM (g/mol):** peso molecular [< 500], **logP** [< 5], **HBD:** donadores de hidrógeno [< 5], **HBA:** aceptores de hidrógeno [< 10], **GIA:** absorción gastrointestinal [= 2], **BBB:** permeabilidad de la barrera hematoencefálica [= 0], **PGP:** sustrato de glicoproteína-P [= 0], **ESOL:** solubilidad estimada [> -6]. Se han destacado en verde los valores que satisfacen los criterios y en rojo los valores que los incumplen. Se han destacado sobre fondo azul los ligandos seleccionados.

Como se puede observar en la tabla 8, la mayoría de ligandos seleccionados cuentan con unas propiedades farmacocinéticas bastante favorables, en particular, destacan 5 compuestos. En la figura 16, se pueden observar las estructuras de estos ligandos. Los compuestos son químicamente diferentes entre sí, lo que permite disponer de cierta variabilidad química en el experimento.



**Figura 16. Estructura química de los ligandos sugeridos. A)** Estructura química del ligando Apexbt\_0\_240. **B)** Estructura química del ligando Apexbt\_0\_130 (Cortisol). **C)** Estructura química del ligando Apexbt\_0\_106 (Mevastatin). **D)** Estructura química del ligando Apexbt\_0\_263 (Sec-O-glucosylhamaudol). **E)** Estructura química del ligando Apexbt\_0\_325 (Demetoxicurcumina).

#### Apexbt\_0\_240

El ligando Apexbt\_0\_240 cumple con todos los parámetros de la regla de Lipinski, cuenta con una buena absorción gastrointestinal y no es sustrato de la glicoproteína P, sin embargo, es permeable a la barrera hematoencefálica. Su puntuación según ESOL de -3.83 ( $1.47E-4$  mol/L) predice que se trata de un compuesto soluble.

#### Apexbt\_0\_130

El ligando Apexbt\_0\_130 (cortisol) no presenta ninguna violación de Lipinski y predice una buena absorción gastrointestinal e impermeabilidad por parte de la barrera hematoencefálica. Sin embargo, el compuesto es sustrato de pgg, lo que supone que su acumulación en las células se vería disminuida. Especialmente en el caso de LRH-1, teniendo el cáncer como objetivo terapéutico, es muy probable que apareciese un fenotipo de resistencia al tratamiento. En cuanto a la solubilidad, el valor de ESOL es de -2.97 ( $1.07E-3$  mol/L), lo que se corresponde con un compuesto soluble.

#### Apexbt\_0\_106

El ligando Apexbt\_0\_106 no presenta violaciones de Lipinski, cuenta con una buena absorción gastrointestinal, no es sustrato de la glicoproteína P, pero sí es capaz de atravesar la barrera hematoencefálica, lo que supondría una desventaja en caso de que el compuesto tuviese efectos adversos en el sistema nervioso. Por último, el valor de solubilidad estimado de  $-4.22$  ( $5.96E-5$  mol/L) se corresponde con una solubilidad moderada.

#### Apexbt\_0\_263

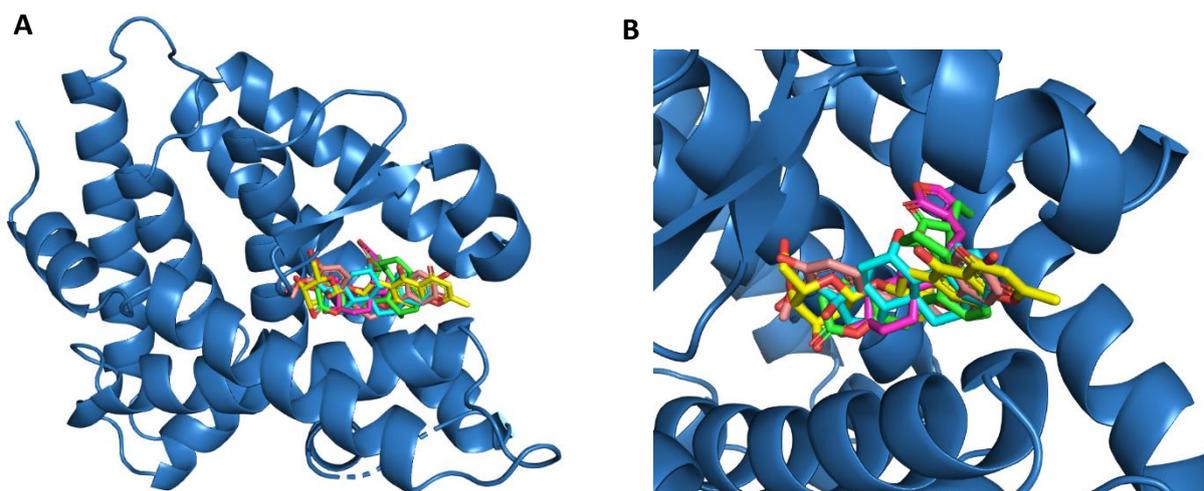
Este ligando no posee violaciones de la regla de Lipinski. Se observa que Apexbt\_0\_263 (Sec-O-glucosylhamaudol) tiene una absorción intestinal moderada, no atraviesa la barrera hematoencefálica y tampoco es sustrato de PGP. Además, el valor de solubilidad predicho es de  $-2.44$  ( $1.43E-3$  mol/L), correspondiéndose con un compuesto soluble.

#### Apexbt\_0\_325

El ligando Apexbt\_0\_325 (Demetoxicurcumina) no posee ninguna violación de Lipinski, posee una absorción intestinal buena y no atraviesa la barrera hematoencefálica, por lo que, en teoría, no habría interacciones en el sistema nervioso central. Además, el compuesto no es sustrato de la glicoproteína P. En cuanto a la solubilidad, la puntuación según ESOL es de  $-3.92$  (equivalente a  $1.2E-4$  mol/L), que se corresponde con una molécula soluble.

### 5.5. Ligandos seleccionados. Interacciones proteína-ligando

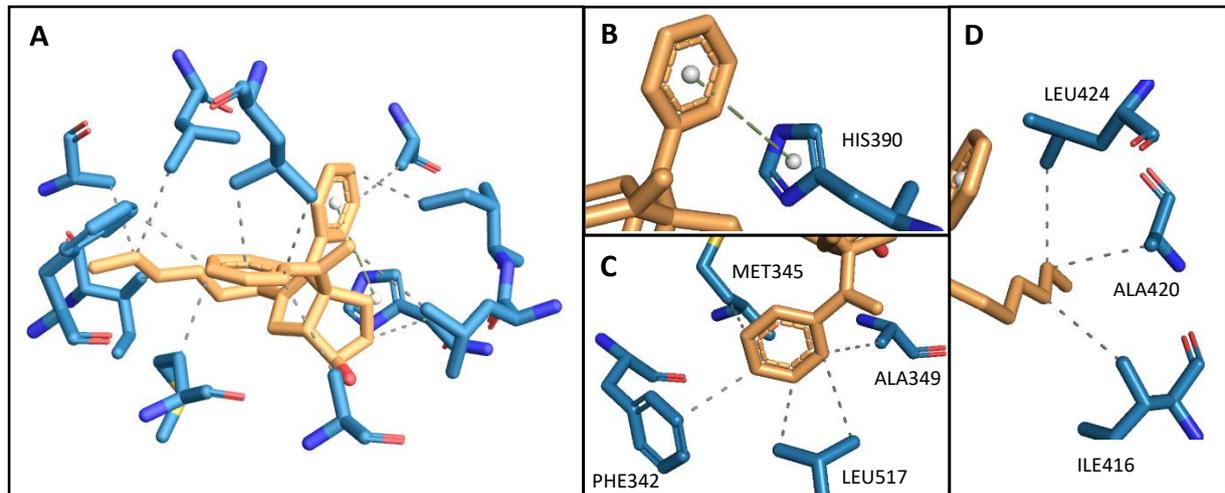
En este apartado, se realiza una breve descripción de los ligandos con mejores propiedades farmacocinéticas descritos en el apartado anterior (figura 17). Además, se describen sus interacciones con el receptor y se comparan con las del ligando cristalográfico.



**Figura 17. Ligandos sugeridos acoplados en el sitio de unión del receptor. A)** Vista global del receptor y los ligandos situados adecuadamente en el sitio de unión. **B)** Ampliación del sitio de unión donde se muestra la configuración espacial que adoptan los ligandos superpuestos. Apexbt\_0\_240 (color rosa), Apexbt\_0\_130 (color cian), Apexbt\_0\_106 (color verde), Apexbt\_0\_263 (color amarillo), Apexbt\_0\_325 (color rosa palo).

### Ligando cristalográfico

Atendiendo al modelo estructural usado (PDB 5L11), se pueden describir en detalle las interacciones entre el receptor y el ligando cocrystalizado (figura 18) con ayuda de la herramienta online Protein-Ligand Interaction Profiler o PLIP (<https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index>).



**Figura 18. Interacciones del ligando cristalográfico (RJW100) con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. **A)** Vista global. Se muestran las interacciones hidrofóbicas mediante líneas segmentadas de color gris. **B)** Interacción de apilamiento perpendicular de anillos aromáticos ( $\pi$ - $\pi$ ). La interacción se representa con una línea segmentada de color verde que une los centros aromáticos representados por esferas de color blanco. **C)** Detalle de interacciones establecidas entre uno de los grupos fenilo del ligando encajado en un bolsillo hidrofóbico. **D)** Interacciones establecidas con la cadena alifática del ligando.

Como se puede observar en el panel A de la figura 18, todas las interacciones son de naturaleza hidrofóbica. En el panel B de la misma figura, se destaca la interacción de apilamiento  $\pi$ - $\pi$  perpendicular que se establece entre el anillo aromático del residuo de histidina 390 y uno de los anillos aromáticos del ligando cristalográfico, a una distancia de 4.98 Å.

En el panel C, se observa como otro anillo aromático del ligando establece múltiples interacciones con 4 residuos del bolsillo de unión del receptor, concretamente con: metionina 345 a una distancia de 3.65 Å, fenilalanina 342 a 3.5 Å, leucina 517 a 3.52 y 3.77 Å y alanina 349 a 3.82 Å.

Finalmente, en el panel D, se muestran las interacciones que se establecen con la cadena alifática del ligando, estas son: leucina 424 a una distancia de 3.82 Å, alanina 420 a 3.84 Å e isoleucina 416 a 3.74 Å.

Una vez conocidas las interacciones entre el receptor y el ligando cristalográfico se pueden describir los ligandos sugeridos y comparar sus interacciones.

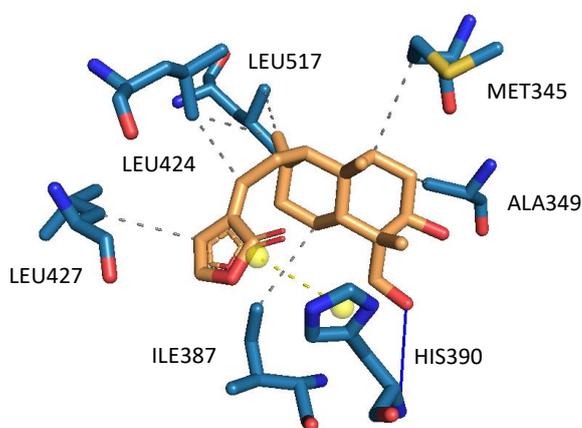
Apexbt\_0\_240

El compuesto Apexbt\_0\_240, PubChem ID: 122173144, (figura 16, panel A) posee 5 anillos, todos ellos alifáticos. No tiene cargas netas y presenta 5 centros quirales.

Como se observa en la figura 19, este ligando posee mayoritariamente interacciones de naturaleza hidrofóbica, como las que forman la leucina 427, la leucina 424, la leucina 517 y la isoleucina 387, pero también presenta dos interacciones polares en las que interviene el residuo de histidina 390, con el que forma un puente de hidrógeno con un grupo hidroxilo del ligando y un puente salino entre el grupo NH de su cadena lateral y el grupo carbonilo del ligando, lo que proporciona una mayor estabilidad al complejo.

El trazo amarillo discontinuo del enlace de tipo puente salino se ha representado uniendo los centros de carga del ligando (carga negativa) y del aminoácido histidina 390 (carga positiva), cuyos centros se han representado como esferas amarillas.

A diferencia del ligando cristalográfico, en esta unión no se forma ninguna apilación  $\pi$ - $\pi$  de anillos aromáticos. Además, el número de interacciones hidrofóbicas totales es menor (12 frente a 7).

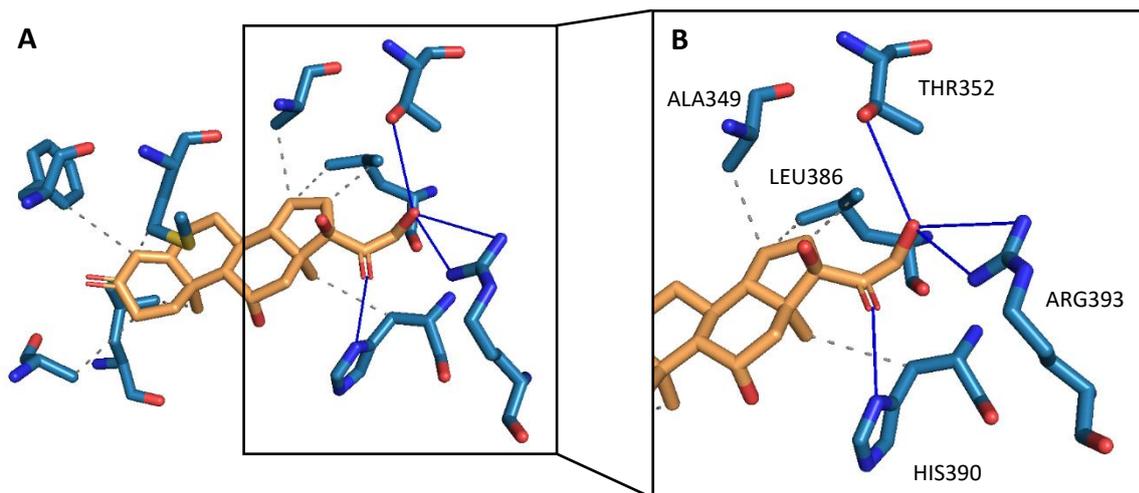


**Figura 19. Interacciones del ligando Apexbt\_0\_240 con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. Vista global con interacciones detalladas y los residuos involucrados etiquetados. Se muestran las interacciones hidrofóbicas representadas mediante líneas discontinuas de color gris, los puentes de hidrógeno mediante líneas continuas de color azul y los puentes salinos mediante líneas discontinuas de color amarillo.

#### Apexbt\_0\_130

El compuesto Apexbt\_0\_130 (figura 16, panel B) se trata del cortisol. La molécula posee 4 anillos, todos ellos alifáticos. No tiene cargas netas y presenta 7 centros quirales.

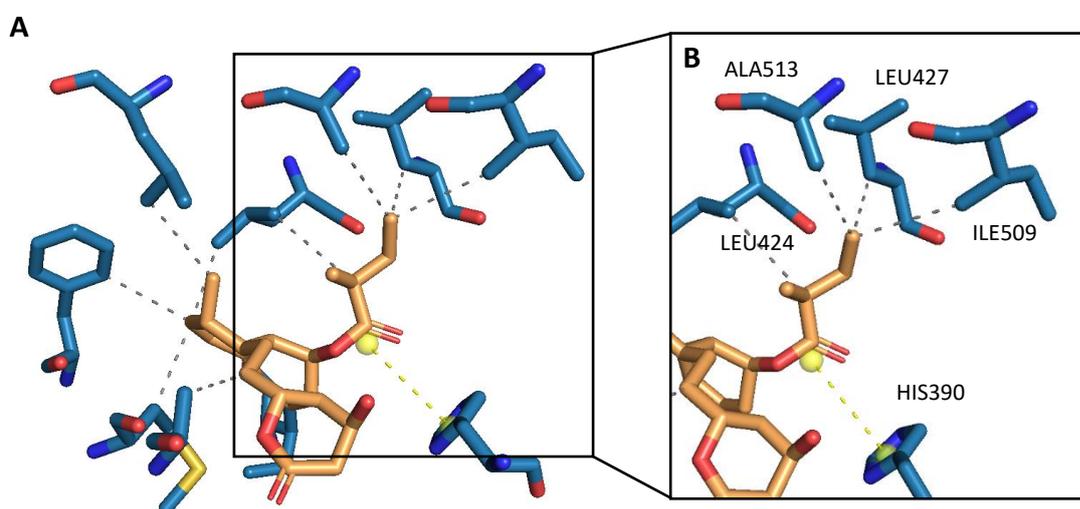
En la figura 20 se puede observar que en este ligando también destacan algunas interacciones polares, aunque siguen predominando las hidrofóbicas. Se destacan en el panel B los puentes de hidrógeno que generan los residuos de histidina 390, arginina 393 y treonina 352. Es importante destacar como estos residuos que forman interacciones polares se encuentran más enterrados en el bolsillo de unión del receptor, mientras que hacia el exterior se encuentran los residuos que establecen interacciones hidrofóbicas.



**Figura 20. Interacciones del ligando Apexbt\_0\_130 con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. **A)** Vista global. Interacciones hidrofóbicas representadas mediante líneas segmentadas de color gris y puentes de hidrógeno representados mediante líneas continuas de color azul. **B)** Detalle de las interacciones polares e hidrofóbicas de la región resaltada en el panel A.

#### Apexbt\_0\_106

El compuesto Apexbt\_0\_106 (figura 16, panel C) es el fármaco Mevastina. La molécula posee 3 anillos, todos ellos alifáticos. No tiene cargas netas y presenta 7 centros quirales.

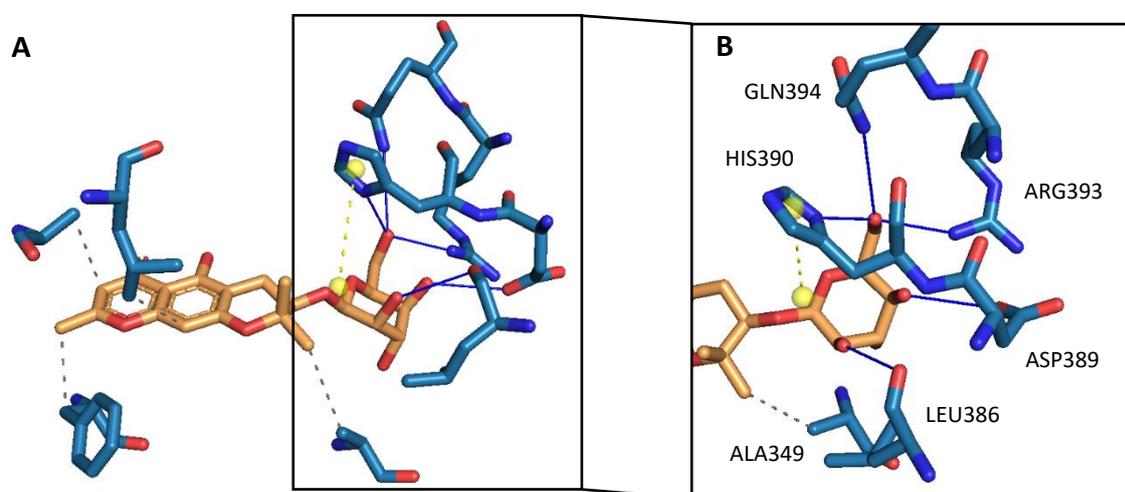


**Figura 21. Interacciones del ligando Apexbt\_0\_106 con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. **A)** Vista global. Interacciones hidrofóbicas representadas mediante líneas segmentadas de color gris y puente salino representado mediante una línea segmentada de color amarillo. **B)** Detalle de las interacciones polares e hidrofóbicas de la región resaltada en el panel A.

Estudiando las interacciones para el ligando Apexbt\_0\_106 o Mevastina (figura 21) se observa que, de igual forma que en el ligando cristalográfico, todas las interacciones son hidrofóbicas, a excepción de una interacción de tipo puente salino. Concretamente, en el panel B se puede apreciar como el residuo de histidina 390 forma un enlace con el grupo carbonilo del ligando. También se muestra la participación de la leucina 424, la alanina 513, la leucina 427 y la isoleucina 509. La mayoría de residuos que participan en la interacción con este ligando también participan en la interacción con el ligando cristalográfico.

#### Apexbt\_0\_263

El compuesto Apexbt\_0\_263 o Sec-O-glucosylhamaudol, ChemSpider ID: 8653686, (figura 16, panel D) posee 4 anillos, dos de ellos aromáticos y dos alifáticos. No tiene cargas netas y presenta 6 centros quirales.

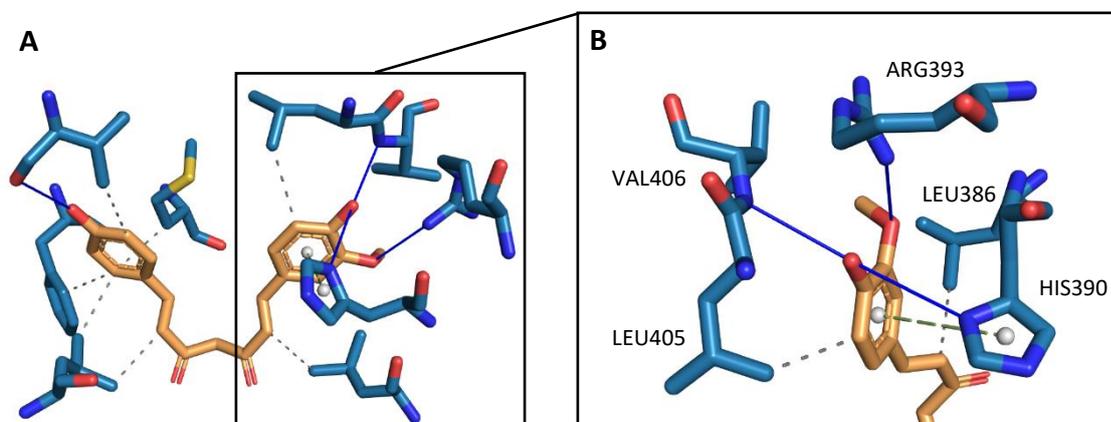


**Figura 22. Interacciones del ligando Apexbt\_0\_263 con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. **A)** Vista global. Interacciones hidrofóbicas representadas mediante líneas segmentadas de color gris, puentes de hidrógeno representados mediante líneas continuas de color azul y puente salino representado mediante una línea segmentada de color amarillo. **B)** Detalle de las interacciones polares de la región resaltada en el panel A.

En el caso del ligando Apexbt\_0\_263 (figura 22), se observa que la mayoría de interacciones son de naturaleza polar, contrariamente a los ligando anteriores y al ligando cristalográfico, donde predominaban las interacciones de naturaleza hidrofóbica. Concretamente, en el panel B se puede apreciar como los residuos de histidina 390, glutamina 394, arginina 393, ácido aspártico 389 y leucina 386 consiguen formar un total de 5 puentes de hidrógeno. Además, la histidina 390 también forma un puente salino con un grupo carboxilo del ligando. Nuevamente, se observa como los residuos que forman enlaces polares están más enterrados en el sitio de unión y los residuos hidrofóbicos se encuentran hacia la entrada del sitio de unión.

### Apexbt\_0\_325

El compuesto Apexbt\_0\_325 (figura 16, panel E), se corresponde con la demetoxicurcumina. La molécula posee 2 anillos, ambos aromáticos. No tiene cargas netas y no posee centros quirales.



**Figura 23. Interacciones del ligando Apexbt\_0\_325 con el receptor LRH-1.** Se representa el ligando cristalográfico en color sepia y los residuos del receptor involucrados en color azul. **A)** Vista global. Se muestran las interacciones hidrofóbicas mediante líneas discontinuas de color gris, los puentes de hidrógeno mediante líneas continuas de color azul y el apilamiento de anillos aromáticos mediante una línea discontinua de color verde. **B)** Detalle de las interacciones polares e hidrofóbicas de la región resaltada en el panel A.

Del análisis de las interacciones para el ligando Apexbt\_0\_325 (figura 23) se observa que siguen predominando las interacciones hidrofóbicas, aunque destacan algunas interacciones por puente de hidrógeno, como lo son las que generan los residuos de arginina 393, histidina 390 y valina 406. Concretamente, en el panel B se puede apreciar como el NH del residuo histidina 390 genera un puente de hidrógeno con el oxígeno del grupo hidroxilo del anillo del ligando, además, este mismo anillo forma un apilamiento perpendicular  $\pi$ - $\pi$  con el anillo de histidina 390, como ocurría en el caso del ligando cristalográfico, lo que genera una mayor estabilidad. También se muestra la interacción entre el NH de la arginina 393 y el oxígeno del grupo carbonilo del ligando. Por otra parte, se muestra la participación de la leucina 405 y la leucina 386 mediante interacciones hidrofóbicas.

### 5.6. Ligandos seleccionados. Disponibilidad comercial.

Se comprobó la disponibilidad comercial de cada molécula seleccionada en algunas bases de datos como PubChem (<https://pubchem.ncbi.nlm.nih.gov/>), ChemSpace (<https://chem-space.com/>), Mcule (<https://mcule.com/search/>), ZINC (<https://zinc12.docking.org/search/>) o Molport (<https://www.molport.com/shop/index>).

-Apexbt\_0\_240 está disponible en ChemSpace (ID: CSSS14863683590) y en Molport (ID: MolPort-051-762-680). Su síntesis puede ser encargada a Apexbio Technology LLC con una pureza del 99%.

-Apexbt\_0\_130 se encuentra disponible en todas las bases de datos, ya que se trata del compuesto cortisol. Algunas casas comerciales donde se puede encargar son: Sigma-Aldrich (ID: 286095\_ALDRICH), con un 98% de pureza; Mcule (ID: MCULE-8463642691), con un de pureza del 98%; o Molport (ID: MolPort-001-794-637), también con una pureza del 98%.

- Apexbt\_0\_106 o Mevastina, también se encuentra disponible en todas las bases de datos, algunos ejemplos donde se puede obtener son: Molport (ID: MolPort-002-885-840) que asegura una pureza del 97%; Thermo Fisher Scientific (ID: GID\_900000007732314), con una pureza superior al 97%; Sigma-Aldrich (ID: M2537\_SIGMA) que ofrece una pureza del 98%; y Mcule (ID: MCULE-5661885893), con un 95% de pureza.

- Apexbt\_0\_263 está disponible en ChemSpace (ID: CSSB00020666987). Se puede encargar a varios distribuidores: Key Organics Limited, Ambeed Inc, BLD PHARMATECH LTD CN, etc. Todos ellos garantizan un 95% de pureza.

- Apexbt\_0\_325 o Demetoxicurcumina, se encuentra en PubChem (ID: 5469424) y se puede obtener a través de Molport (ID: MolPort-016-638-373), con una pureza del 98% y a través de Mcule (ID: MCULE-8565366410) también con una pureza del 98%.

## 6. Discusión

El avance de la biología estructural ha permitido la identificación de objetivos biológicos clave para tratar enfermedades como el cáncer y la diabetes. La mejora de las herramientas computacionales junto con la expansión de las bases de datos biológicas públicas han facilitado el análisis virtual de miles de moléculas orgánicas en los sitios de unión de las dianas antes de la verificación experimental.

La estrategia computacional ofrece ventajas como el bajo costo, la rapidez, la escalabilidad y la automatización, lo que ha transformado el descubrimiento de fármacos. Entre estas estrategias, el acoplamiento molecular y el cribado virtual destacan por su capacidad para identificar moléculas con alto potencial terapéutico. Estos métodos se caracterizan por su velocidad y eficacia para determinar la posición y orientación más favorables del ligando en su sitio de unión, lo que los hace computacionalmente eficientes (Blanes Mira et al., 2022). Sin embargo, presentan algunos desafíos: (1) requieren estructuras de alta resolución, (2) a menudo no se conoce el sitio de unión del ligando, (3) no pueden predecir el tipo de actividad del ligando (activador o inhibidor), y (4) la flexibilidad del receptor está limitada a unas pocas cadenas laterales en el bolsillo de unión.

Además, los resultados del cribado virtual dependen críticamente del muestreo conformacional del ligando en el sitio de unión y de la calidad de las funciones de puntuación, lo que puede generar numerosos falsos positivos y negativos. Debido a que distintos programas de acoplamiento emplean diversas estrategias de muestreo y funciones de puntuación, se observan variaciones significativas en el rendimiento entre ellos. Por ello, el uso de métodos de acoplamiento consenso representa un avance significativo, mejorando la fiabilidad de los programas individuales al compensar las limitaciones y sesgos de cada algoritmo. El principal inconveniente de este enfoque es el tiempo de cálculo, ya que es necesario evaluar las quimiotecas con todos los programas de acoplamiento. En el cribado virtual con el receptor LRH-1 y la librería etiquetada, el programa de acoplamiento que demostró mayor rendimiento para discriminar activos de no activos fue AutoDock, como lo demuestran las métricas de las curvas ROC, el área bajo la curva y el gráfico de enriquecimiento (EP). En el caso de los métodos de consenso

utilizados en este trabajo, Z-Score y RBR obtuvieron mejores resultados en comparación con el resto de métodos. Una desventaja específica del proceso de validación usado en este trabajo es la fiabilidad de las moléculas señuelo, ya que no podemos tener la seguridad de que las moléculas señuelo generadas se vayan a comportar como moléculas no activas y por lo tanto cabe la posibilidad de que algunas de ellas realmente tengan afinidad por el receptor y se acaben acoplado al sitio de unión como lo haría una molécula activa. Esto puede distorsionar las métricas AUC-ROC e infravalorar la clasificación de algunos programas y métodos.

No obstante, se intentó optimizar al máximo las métricas de validación, eliminando del cálculo los programas que peor puntuaban (PLANTS y X-Score), permitiendo así determinar los métodos más adecuados para ser aplicados en el sistema LRH-1 y la librería APEXBT\_natural. El resultado de aplicar esta estrategia con el método de consenso RBR (segundo método más fiable después de Z-Score) fue la obtención de una lista priorizada de compuestos con potencial unión al receptor LRH-1, de los cuales 18 puntuaban por encima del ligando cristalográfico. Las mejores moléculas obtenidas se filtraron según sus propiedades farmacocinéticas, obteniendo así 5 compuestos con propiedades favorables, es decir, sin violaciones de la regla de Lipinski (Apexbt\_0\_240, Apexbt\_0\_130, Apexbt\_0\_106, Apexbt\_0\_263 y Apexbt\_0\_325). Aunque algunos de ellos presentan algunas desventajas: Apexbt\_0\_240 y Apexbt\_0\_106 son capaces de atravesar la barrera hematoencefálica y Apexbt\_0\_130 es sustrato de la glicoproteína P, lo que los convierte en candidatos no idóneos. Al analizar las interacciones proteína-ligando, se verificó que todos ellos se acoplaban adecuadamente en el sitio de unión del receptor, además se pudo observar que las interacciones polares se daban en residuos que se encontraban más enterrados en el bolsillo de unión, lo que sugiere que estos enlaces son cruciales para la estabilidad y especificidad de la interacción, en contraste, los residuos que establecían interacciones hidrofóbicas se encontraban hacia la entrada del sitio de unión, lo que podría facilitar la entrada del ligando al sitio de unión y ayudar a excluir el agua, favoreciendo la interacción inicial y contribuyendo a la orientación del ligando hacia el sitio activo más enterrado. Cabe destacar el papel de la histidina 390 que parece ser un residuo muy relevante a la hora de establecer interacciones polares, tanto puentes de hidrógeno como puentes salinos, con diversos ligandos.

De los 5 compuestos propuestos, tres son agentes bioactivos conocidos, Apexbt\_0\_130 (Cortisol), Apexbt\_0\_106 (Mevastina) y Apexbt\_0\_325 (Demetoxicurcumina). El hallazgo de que el cortisol sea un posible ligando de LRH-1 sugiere que esta hormona podría influir directamente en las funciones reguladas por LRH-1, lo que podría tener implicaciones significativas en enfermedades metabólicas e inflamatorias (Yu, H. et al., 2024). En cuanto a la Mevastina, se trata de un agente reductor de colesterol aislado del hongo *Penicillium citinium*, es tradicionalmente conocida como un inhibidor de la enzima HMG-CoA reductasa y fue el primer agente descubierto perteneciente a la clase de medicamentos hipolipemiantes, también conocidos como estatinas (Moskowitz M. A. et al., 2001). El hecho de que la Mevastina se una al receptor LRH-1 sugiere una nueva vía de interacción entre este medicamento y otros receptores, que no ha sido documentada previamente en la literatura científica. Por otra parte, la demetoxicurcumina es un compuesto natural encontrado en la cúrcuma y que se ha reportado que tiene un papel como agente antineoplásico y antiinflamatorio (Sahebkar, A. et al., 2018), aunque en la bibliografía no se ha descrito hasta la fecha ninguna relación con el receptor LRH-1.

El hecho de que estas moléculas puedan interactuar con el receptor LRH-1 y tener algún efecto terapéutico

podría resultar interesante como estrategia de reposicionamiento de fármacos.

Además, cabe destacar que el compuesto Apexbt\_0\_263 (Sec-O-glucosylhamaudol) es un flavonoide que se encuentra en la planta *Amomum subulatum* (cardamomo negro) y, aunque se desconoce si el compuesto posee actividad biológica *per se*, el extracto de dicha planta se ha relacionado con un efecto antioxidante e hipoglucemiante en ratas diabéticas (Zhang, X. F. et al., 2022), lo que podría ser indicativo de un posible efecto terapéutico de este compuesto.

De los 5 compuestos, solo Apexbt\_0\_263 (Sec-O-glucosylhamaudol) y Apexbt\_0\_325 (Demetoxicurcumina) son realmente buenos candidatos potenciales para ser probados experimentalmente, puesto que cumplen con todas las condiciones farmacocinéticas impuestas a los ligandos, poseen unas interacciones moleculares sólidas en el sitio de unión (múltiples interacciones hidrofóbicas, interacciones polares específicas y apilamiento de anillos aromáticos) y además, estudios anteriores sugieren un posible potencial terapéutico para estos compuestos.

## 7. Conclusiones

En este trabajo se ha expuesto la relevancia del receptor LRH-1 como nuevo objetivo terapéutico en algunos tipos de cáncer y de la diabetes tipo I. Con la intención de aportar valor en el tratamiento de estas enfermedades, se ha llevado a cabo una estrategia de cribado virtual usando métodos de acoplamiento consenso y se han propuesto 5 moléculas candidatas con potencial farmacológico.

De los 5 compuestos propuestos, Apexbt\_0\_325 (demetoxicurcumina) y Apexbt\_0\_263 (Sec-O-glucosylhamaudol) son probablemente los mejores candidatos, por su puntuación en el método consenso RBR, por satisfacer de forma adecuada todos los criterios farmacocinéticos evaluados y por sus diversas interacciones moleculares en el sitio de unión.

Además, en este trabajo se ha demostrado la mejora de la calidad de las predicciones que ofrecen los métodos de acoplamiento consenso, así como la aplicación de las curvas ROC y las gráficas de enriquecimiento (EP) en la elección de un método fiable para realizar cribado virtual, a pesar de la desventaja que supone para el análisis estadístico de estas métricas, el hecho de que las moléculas señuelo no garanticen comportarse realmente como compuestos no activos dentro de una librería etiquetada.

Finalmente, se espera que, en el futuro, los candidatos derivados de estas estrategias sean probados y validados experimentalmente, y logren superar con éxito los ensayos clínicos para llegar al mercado. Aunque el proceso de descubrimiento de fármacos es largo, costoso y con un alto porcentaje de fracaso, invertir más esfuerzo en la fase preclínica puede reducir significativamente el riesgo y los costos en las etapas posteriores.

## Bibliografia

Ahmed, A., Reinhold, C., Breunig, E., Phan, T. S., Dietrich, L., Kostadinova, F., Urwyler, C., Merk, V. M., Noti, M., Toja da Silva, I., Bode, K., Nahle, F., Plazzo, A. P., Koerner, J., Stuber, R., Menche, C., Karamitopoulou, E., Farin, H. F., Gollob, K. J., & Brunner, T. (2023). Immune escape of colorectal tumours via local LRH-1/Cyp11b1-mediated synthesis of immunosuppressive glucocorticoids. *Molecular Oncology*, 17(7), 1545–1566. <https://doi.org/10.1002/1878-0261.13414>

Alisa, L., Isigkeit, L., Schubert-Zsilavec, M., & Merk, D. (2021). The medicinal chemistry and therapeutic potential of LRH-1 modulators. *Journal of Medicinal Chemistry*, 64(23), 16956–16973. <https://doi.org/10.1021/acs.jmedchem.1c01663>

Andole, S., Sd, H., Sudhula, S., Vislavath, L., Boyina, H. K., Gangarapu, K., Bakshi, V., & Devarakonda, K. P. (2023). 3D QSAR based virtual screening of flavonoids as acetylcholinesterase inhibitors. *Advances in Experimental Medicine and Biology*, 1424, 233–240. [https://doi.org/10.1007/978-3-031-31982-2\\_26](https://doi.org/10.1007/978-3-031-31982-2_26)

Bajusz, D., Ferenczy, G. G., & Keseru, G. M. (2017). Structure-based virtual screening approaches in kinase-directed drug discovery. *Current Topics in Medicinal Chemistry*, 17(20), 2235–2259. <https://doi.org/10.2174/1568026617666170224121313>

Benod, C., Carlsson, J., Uthayaruban, R., Hwang, P., Irwin, J. J., Doak, A. K., Shoichet, B. K., Sablin, E. P., & Fletterick, R. J. (2013). Structure-based discovery of antagonists of nuclear receptor LRH-1. *The Journal of Biological Chemistry*, 288(27), 19830–19844. <https://doi.org/10.1074/jbc.M112.411686>

Blanes-Mira, C., Fernández-Aguado, P., de Andrés-López, J., Fernández-Carvajal, A., Ferrer-Montiel, A., & Fernández-Ballester, G. (2022). Comprehensive survey of consensus docking for high-throughput virtual screening. *Molecules*, 28(1), 175. <https://doi.org/10.3390/molecules28010175>

Cato, M. L., Cornelison, J. L., Spurlin, R. M., Courouble, V. V., Patel, A. B., Flynn, A. R., Johnson, A. M., Okafor, C. D., Frank, F., D'Agostino, E. H., Griffin, P. R., Jui, N. T., & Ortlund, E. A. (2022). Differential modulation of nuclear receptor LRH-1 through targeting buried and surface regions of the binding pocket. *Journal of Medicinal Chemistry*, 65(9), 6888–6902. <https://doi.org/10.1021/acs.jmedchem.2c00235>

Daina, A., & Zoete, V. (2016). A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem*, 11(11), 1117–1121. <https://doi.org/10.1002/cmdc.201600182>

Doogue, M. P., & Polasek, T. M. (2013). The ABCD of clinical pharmacokinetics. *Therapeutic Advances in Drug Safety*, 4(1), 5–7. <https://doi.org/10.1177/2042098612469335>

Gauthier, B. R., Najia, L. P., & Lorenzo, P. I. (2018). Targeting LRH-1/NR5A2 to treat type 1 diabetes mellitus. *Cell Stress*, 2(6), 141–143. <https://doi.org/10.15698/cst2018.06.140>

Gauthier, B. R., Najia, L. P., & Lorenzo, P. I. (2022). NR5A2/LRH-1 regulates the PTGS2-PGE2-PTGER1 pathway contributing to pancreatic islet survival and function. *iScience*, 25(5), 104345. <https://doi.org/10.1016/j.isci.2022.104345>

- Hatamipour, M., Ramezani, M., Tabassi, S. A. S., Johnston, T. P., Ramezani, M., & Sahebkar, A. (2018). Demethoxycurcumin: A naturally occurring curcumin analogue with antitumor properties. *Journal of Cellular Physiology*, 233(12), 9247–9260. <https://doi.org/10.1002/jcp.27029>
- Heikamp, K., & Bajorath, J. (2014). Support vector machines for drug discovery. *Expert Opinion on Drug Discovery*, 9(1), 93–104. <https://doi.org/10.1517/17460441.2014.866943>
- Lang, A., Isigkeit, L., Schubert-Zsilavec, M., & Merk, D. (2021). The medicinal chemistry and therapeutic potential of LRH-1 modulators. *Journal of Medicinal Chemistry*, 64(23), 16956–16973. <https://doi.org/10.1021/acs.jmedchem.1c01663>
- Malabanan, M. M., Chapagain, P., Haratipour, Z., Choi, W. J., Orun, A. R., & Blind, R. D. (2023). New high-throughput screen discovers novel ligands of full-length nuclear receptor LRH-1. *ACS Chemical Biology*, 18(5), 1101–1114. <https://doi.org/10.1021/acscchembio.2c00805>
- Mays, S. G., Okafor, C. D., Whitby, R. J., Goswami, D., Stec, J., Flynn, A. R., Dugan, M. C., Jui, N. T., Griffin, P. R., & Ortlund, E. A. (2016). Crystal structures of the nuclear receptor, liver receptor homolog 1, bound to synthetic agonists. *The Journal of Biological Chemistry*, 291(49), 25281–25291. <https://doi.org/10.1074/jbc.M116.753541>
- Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14), 6582–6594. <https://doi.org/10.1021/jm300687e>
- Nadolny, C., & Dong, X. (2015). Liver receptor homolog-1 (LRH-1): A potential therapeutic target for cancer. *Cancer Biology & Therapy*, 16(7), 997–1004. <https://doi.org/10.1080/15384047.2015.1045693>
- Ortlund, E. A., Lee, Y., Solomon, I. H., Hager, J. M., Safi, R., Choi, Y., Guan, Z., Tripathy, A., Raetz, C. R., McDonnell, D. P., Moore, D. D., & Redinbo, M. R. (2005). Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nature Structural & Molecular Biology*, 12(4), 357–363. <https://doi.org/10.1038/nsmb910>
- Palacio-Rodríguez, K., Lans, I., Cavasotto, C. N., & Cossio, P. (2019). Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Scientific Reports*, 9, 5142. <https://doi.org/10.1038/s41598-019-41594-3>
- Scott, B. (2017). Computational studies of liver receptor homolog 1 in the presence of small molecule agonists: Allosteric communication and virtual screening for new potential drug candidates (Thesis). *Georgia State University*. <https://doi.org/10.57709/10496114>
- Seacrist, C. D., Kuenze, G., Hoffmann, R. M., Moeller, B. E., Burke, J. E., Meiler, J., & Blind, R. D. (2020). Integrated structural modeling of full-length LRH-1 reveals inter-domain interactions contribute to receptor structure and function. *Structure*, 28(7), 830–846. <https://doi.org/10.1016/j.str.2020.04.020>
- Sun, W., Shi, Q., Li, J., Li, J., & Yu, L. (2020). LRH1 promotes tumour cell proliferation and migration and is correlated with poor prognosis in ovarian cancer. *Frontiers in Oncology*, 10, 583566. <https://doi.org/10.3389/fonc.2020.583566>

Sun, Y., Demagny, H., & Schoonjans, K. (2021). Emerging functions of the nuclear receptor LRH-1 in liver physiology and pathology. *Molecular Basis of Disease*, 1867(8), 166145. <https://doi.org/10.1016/j.bbadis.2021.166145>

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461. <https://doi.org/10.1002/jcc.21334>

Weikum, E. R., Liu, X., & Ortlund, E. A. (2018). The nuclear receptor superfamily: A structural perspective. *Protein Science*, 27(11), 1876–1892. <https://doi.org/10.1002/pro.3496>

Whitby, R. J., Dixon, S., Maloney, P. R., Delerive, P., Goodwin, B. J., Parks, D. J., & Willson, T. M. (2006). Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. *Journal of Medicinal Chemistry*, 49(23), 6652–6655. <https://doi.org/10.1021/jm060990k>

Wu, T., Lu, Z., & Yu, H. (2024). Liver receptor homolog-1: Structures, related diseases, and drug discovery. *Acta Pharmacologica Sinica*. <https://doi.org/10.1038/s41401-024-01276-x>

Zhang, X. F., Tang, Y. J., Guan, X. X., Lu, X., Li, J., Chen, X. L., Deng, J. L., & Fan, J. M. (2022). Flavonoid constituents of *Amomum tsao-ko* Crevost et Lemarie and their antioxidant and antidiabetic effects in diabetic rats - in vitro and in vivo studies. *Food & Function*, 13(1), 437–450. <https://doi.org/10.1039/d1fo02974f>