



Universidad Internacional de La Rioja

Facultad de Ciencias de la Salud

Máster Universitario en Bioinformática

## Desarrollo de flujos de trabajo para la optimización de fármacos candidatos

Trabajo fin de Estudio presentado por:	Óscar Saiz Gutiérrez
Tipo de trabajo:	Redacción de trabajo original relacionado con bases de datos.
Director/a:	Ricardo Blázquez Encinas
Codirector/a:	Carlos Óscar Sánchez Sorzano
Fecha:	05/02/2025

## Resumen

El presente trabajo tiene como objetivo desarrollar un flujo de trabajo computacional eficiente para la optimización de ligandos dirigidos a la proteína CXCR4, una diana terapéutica clave en enfermedades como el cáncer y el VIH. Se emplearon herramientas bioinformáticas avanzadas como DeepFrag y AutoGrow4 para mejorar la afinidad y estabilidad de los ligandos seleccionados. A través de un proceso sistemático que incluyó la preparación de la proteína, identificación de pockets, docking molecular y lead optimization, se logró optimizar los compuestos candidatos, destacando la molécula 137 por su mejora significativa en la energía de unión. Los resultados obtenidos demuestran la eficacia de las metodologías in silico en la optimización de fármacos, reduciendo tiempos y costos en la fase inicial del desarrollo farmacéutico. Sin embargo, se identificaron limitaciones en las propiedades farmacológicas de los compuestos optimizados, lo que resalta la necesidad de evaluaciones experimentales adicionales.

**Palabras clave:** optimización de ligandos, docking molecular, CXCR4

## Abstract

The present study aims to develop an efficient computational workflow for the optimization of ligands targeting the CXCR4 protein, a key therapeutic target in diseases such as cancer and HIV. Advanced bioinformatics tools such as DeepFrag and AutoGrow4 were employed to improve the affinity and stability of the selected ligands. Through a systematic process that included protein preparation, pocket identification, molecular docking, and lead optimization, candidate compounds were successfully optimized, with molecule 137 standing out due to its significant improvement in binding energy. The obtained results demonstrate the effectiveness of in silico methodologies in drug optimization, reducing time and costs in the early stages of drug development. However, limitations in the pharmacological properties of the optimized compounds were identified, highlighting the need for additional experimental evaluations.

**Keywords:** ligand optimization, molecular docking, CXCR4.

## Índice de contenidos

<b>Introducción.....</b>	<b>5</b>
<b>Justificación.....</b>	<b>5</b>
Presentación del tema.....	5
Interés personal e importancia científica.....	6
<b>Planteamiento del problema .....</b>	<b>7</b>
Problema principal.....	7
<b>Objetivos del trabajo.....</b>	<b>8</b>
Objetivo general.....	8
Objetivos específicos.....	8
Resultados esperados.....	8
<b>Marco teórico.....</b>	<b>9</b>
Diseño de fármacos <i>in silico</i> .....	9
Proteína CXCR4 como diana terapéuticas.....	12
Descripción de la proteína.....	12
Relevancia patológica.....	13
Métodos para la optimización de ligandos .....	14
Scipion.....	14
DeepFrag.....	15
AutoGrow4.....	16
<b>Materiales y Métodos .....</b>	<b>17</b>
<b>1. Preparación de la Proteína/target:.....</b>	<b>17</b>
1.1. Importación de la estructura:.....	17
1.2. Preparación del receptor:.....	17
<b>2. Identificación de Pockets.....</b>	<b>19</b>
2.1. fpocket.....	19
2.2. p2rank.....	19
2.3. autodock.....	20
2.4. Consensus de ROIs Estructurales.....	21
2.5. Selección del <i>pocket</i> de interés:.....	22
<b>3. Preparación de ligandos .....</b>	<b>24</b>
3.1. Importación de ligandos.....	24
3.2. Preparación de ligandos.....	24

<b>4. Docking Molecular .....</b>	<b>25</b>
4.1. Ejecución del Docking.....	25
4.2. Análisis de Resultados del Docking.....	27
4.3. Selección de moléculas.....	28
<b>5. Lead Optimization.....</b>	<b>29</b>
5.1. DeepFrag .....	29
5.2. AutoGrow4.....	32
<b><i>Resultados.....</i></b>	<b><i>34</i></b>
Preparación de la proteína, ligandos e identificación de pockets. ....	34
Docking molecular .....	36
Lead optimization .....	37
Optimización con DeepFrag.....	37
AutoGrow4.....	38
Resultados generales.....	39
<b><i>Discusión y Conclusiones.....</i></b>	<b><i>40</i></b>
Preparación de Proteína/Ligandos e Identificación de pockets. ....	40
Docking .....	40
Lead Optimization.....	41
Conclusiones:.....	46
<b><i>Bibliografía .....</i></b>	<b><i>48</i></b>
<b><i>Anexo I .....</i></b>	<b><i>53</i></b>
<b><i>Índice de figuras.....</i></b>	<b><i>54</i></b>
<b><i>Índice de tablas .....</i></b>	<b><i>56</i></b>

## Introducción

### Justificación

#### Presentación del tema

La proteína CXCR4, también conocida como “fusina” es un receptor acoplado a proteínas G, que pertenece a la familia de receptores de quimiocinas y se expresa en varios tipos celulares durante el desarrollo y la vida, jugando un papel fundamental en múltiples procesos biológicos [1,2].

Sin embargo, su importancia no solo radica en funciones fisiológicas normales, sino también en su implicación en diversas patologías como el cáncer, el VIH y otras enfermedades inflamatorias[2]. Es uno de los receptores de quimiocinas mejor estudiados debido a su papel como correceptor para la entrada del VIH [3]. La expresión excesiva de CXCR4 se implica en la mediación de metástasis en diversos tipos de cáncer como el cáncer de próstata, leucemias, neuroblastomas, entre otros [4]. La interacción de CXCR4 con su ligando natural, CXCL12, facilita la comunicación entre las células tumorales y su microambiente, protegiendo a las células cancerosas de estreses genotóxicos como la quimioterapia[5].

El diseño de fármacos que puedan inhibir la interacción de CXCR4/CXCL12, tiene gran interés en la comunidad científica, ya que tiene un papel clave en la diseminación del cáncer y en la entrada del virus del VIH [6]. Actualmente, solo existe un medicamento aprobado, AMD3100 (plerixafor)[7], para bloquear esta interacción, lo que convierte a CXCR4 en un objetivo farmacológico prometedor para el desarrollo de nuevas terapias [8].

El uso de herramientas computacionales avanzadas, como el *docking molecular* y la optimización de ligandos, facilita la predicción de cómo los compuestos con potencial terapéutico interactúan con la proteína objetivo [9]. Este enfoque *in silico* permite ahorrar tiempo y recursos en las primeras fases del descubrimiento de fármacos [10].

### **Interés personal e importancia científica**

El diseño de fármacos *in silico* representa un campo en constante evolución combinando varias áreas como biología, bioquímica, química e informática. Mi interés personal en este trabajo surge de la posibilidad de contribuir y participar en el desarrollo de tratamientos contra enfermedades complejas. Además, este proyecto me permite integrar mis conocimientos adquiridos en bioinformática, un área que está adquiriendo un impacto notable en la investigación biomédica.

La importancia científica de este proyecto está en la potencial creación de herramientas y flujos de trabajo que permitan avanzar en el descubrimiento de fármacos, haciéndolos más estables y mejorando la afinidad por la proteína receptora.

## Planteamiento del problema

### Problema principal

El desarrollo y descubrimiento de fármacos es un proceso que requiere una gran inversión de tiempo y recursos. Los costes pueden superar los 1,8 mil millones de dólares para llevar un solo compuesto al mercado [11]. A pesar de ello, las tasas de éxito son muy bajas, existiendo una tasa de deserción de aproximadamente un 96% en etapas avanzadas. Las principales razones de este elevado índice de fracaso se deben a la baja eficacia de los compuestos, así como problemas con sus propiedades fisicoquímicas [10]

En los últimos años, los métodos computacionales para el desarrollo de fármacos han adquirido un importante papel en la investigación biomédica. Estos métodos permiten identificar posibles dianas terapéuticas y realizar un cribado virtual de enormes librerías de moléculas, optimizando los candidatos reduciéndose así el número de compuestos que necesitan una evaluación experimental, haciendo todo el proceso más eficiente [10]

Por ello, el problema principal de este trabajo es desarrollar una estrategia computacional para diseñar e implementar un flujo de trabajo reproducible que facilite la optimización de moléculas dirigidas a un objetivo terapéutico.

Con este trabajo se busca establecer un flujo de trabajo óptimo y eficiente para la optimización computacional de fármacos candidatos. Para desarrollarlo se enfocó en la proteína CXCR4 y 3 ligandos potencialmente válidos.



## Objetivos del trabajo

### Objetivo general

El objetivo general de este TFM es crear un flujo de trabajo efectivo y reproducible para la optimización *in silico* de fármacos dirigidos a la proteína CXCR4. A través de diversas herramientas bioinformáticas, se busca identificar y mejorar compuestos que puedan tener un potencial terapéutico, con especial interés en el tratamiento de enfermedades como el cáncer y el VIH.

### Objetivos específicos

Como objetivos específicos de este TFM cuento con:

- Realizar el *docking* molecular para identificar de qué forma interactúan los ligandos propuestos con la proteína CXCR4 y poder determinar sitios de unión y posiciones de unión efectivas para después optimizar los ligandos.
- Desarrollar un flujo de trabajo eficiente para la optimización de los candidatos integrando diversas herramientas bioinformáticas.
- Emplear las herramientas de DeepFrag y AutoGrow4 en el proceso de optimización de ligandos para mejorar la afinidad y propiedades farmacocinéticas de los compuestos candidatos.
- Evaluación y comparación las propiedades de los compuestos mediante simulaciones computacionales.

### Resultados esperados

Se esperan alcanzar los siguientes resultados:

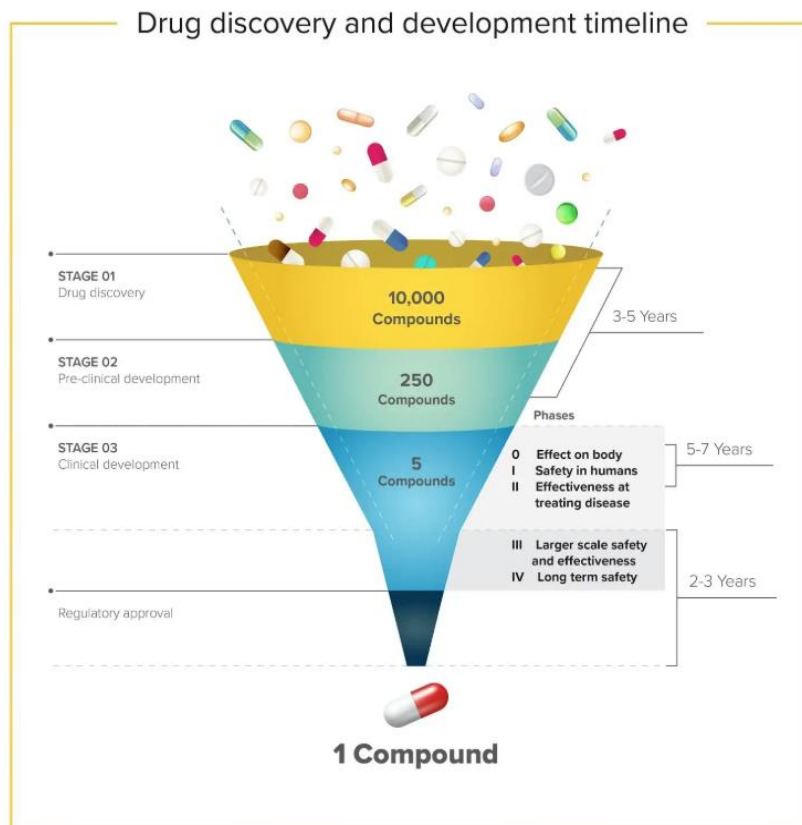
1. Diseño de un flujo de trabajo robusto, integrando las herramientas DeepFrag y Autogrow4, facilitando la mejora de las moléculas. Se pretende que este flujo se adapte a otras proteínas objetivo en investigaciones futuras.
2. Evaluar las herramientas disponibles, identificando posibles limitaciones de cada una.
3. Identificar y generar ligandos de mejor afinidad y propiedades fisicoquímicas.

## Marco teórico

### Diseño de fármacos *in silico*

El desarrollo y diseño de fármacos es un proceso complejo que comienza con la identificación de compuestos capaces de interactuar con un *target* o blanco terapéutico, demostrando una actividad biológica en ensayos de cribado farmacológico [12]. Esto es un proceso largo y complicado (Figura 1), donde el fracaso es bastante común, además de requiere mucha inversión, tanto de dinero como de personal. En total, desarrollar un medicamento puede llegar a costar más de 2.500 millones de dólares y requerir entre 12 y 15 años de trabajo [13], siendo aproximadamente un 75% de estos costos atribuibles a fallos ocurridos a lo largo del proceso [14].

**Figura 1.** Esquema del desarrollo de fármacos



Fuente: modificado de farmaskills.com

Además de ser un proceso largo, costoso y complicado, el desarrollo de fármacos implica el problema de explorar un enorme espacio químico de compuestos similares a los fármacos, se estima que consta de  $10^{63}$  moléculas [15] .

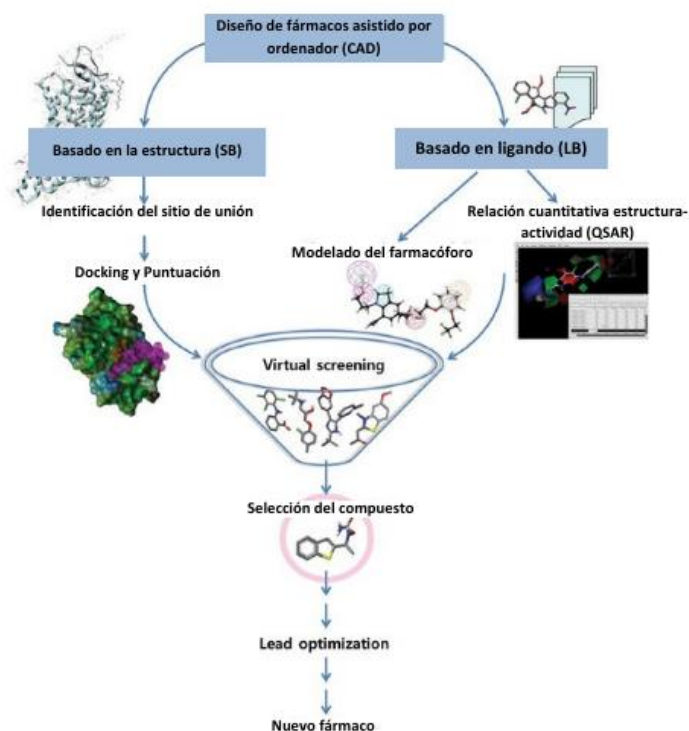
En este contexto, el diseño de fármacos asistido por ordenador (CADD), también conocido como diseño de fármacos *in silico*, representa una estrategia revolucionaria en la investigación biomédica moderna. Este enfoque combina conocimientos avanzados de biología molecular, química y ciencias computacionales [16]. El objetivo principal es optimizar el descubrimiento y desarrollo de los fármacos, disminuyendo de forma considerable la cantidad de ligandos que se deben de examinar en los ensayos experimentales [14]. De esta forma no solo se acelera el proceso de identificación de moléculas candidatas, sino que también se reduce significativamente los costes asociados a las primeras etapas del desarrollo farmacéutico [17]. Estas herramientas identifican moléculas líderes para pruebas posteriores, evaluando su efectividad, anticipando posibles efectos secundarios y optimizando su biodisponibilidad como candidatos a fármacos [14].

Este proceso se conoce comúnmente como Cribado Virtual de Fármacos, también denominado por sus siglas en inglés como *Virtual Drug Screening* (VDS) [15]. Normalmente estos métodos comprenden diferentes enfoques, que se pueden clasificar en dos (Figura 2) [18]:

- Diseño de fármacos basado en la estructura (SB): utiliza modelos estructurales de proteínas obtenidos por técnicas como difracción de rayos X, RMN o simulación molecular para identificar y analizar sitios de unión.
- Diseño de fármacos basado en ligando (LB): se centra en moléculas conocidas que interactúan con una diana, generando modelos de farmacóforo que definen características estructurales esenciales.

Dentro de estos dos campos, en el diseño basado en estructura (SB) se utilizan enfoques como el *docking* o acoplamiento molecular y el diseño *de novo*, mientras que en el diseño basado en ligandos (LB) destacan la relación cuantitativa estructura-actividad (QSAR), que establece una correlación matemática entre las propiedades estructurales de la molécula y su efecto biológico y el modelado de farmacóforos, entre otros [15].

**Figura 2.** Diseño de fármacos asistido por ordenador.



Fuente: Modificado de Prajapat, *et al.* (2017) [19]

Además del cribado virtual, el diseño de fármacos asistido por ordenador (CADD) incluye una amplia gama de aplicaciones que juegan un papel crucial en el descubrimiento y optimización de fármacos. Entre ellas, el *docking* molecular permite predecir cómo una molécula interactúa con un sitio activo de una proteína diana, evaluando su afinidad de unión y las posibles configuraciones espaciales del complejo ligando-proteína [9]. La predicción de propiedades ADMET utiliza modelos computacionales para analizar características clave como la absorción, distribución, metabolismo, excreción y toxicidad de los compuestos candidatos, asegurando que cumplan con los estándares de seguridad y eficacia antes de ser evaluados experimentalmente[20]. Por último, la optimización de ligandos se enfoca en modificar las estructuras moleculares para mejorar su afinidad y especificidad hacia la diana seleccionada [21].

El diseño computacional de fármacos ha contribuido significativamente al desarrollo de numerosos medicamentos exitosos [22] como el Zanamivir contra la gripe [23], el Captopril contra la hipertensión [24] y el Saquinavir contra el VIH [25] entre otros (Anexo I).

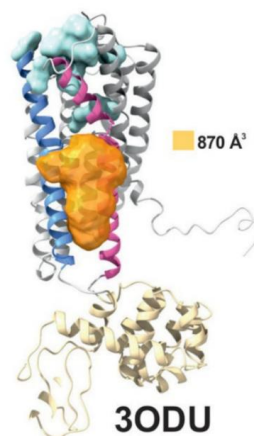
## Proteína CXCR4 como diana terapéuticas

### Descripción de la proteína

El receptor de quimiocinas CXCR4 es una proteína que se encuentra en las membranas celulares y que está conectada a proteínas G, actuando como un mediador clave en varios procesos biológicos [26]. Se encuentra tanto en tejidos embrionarios como en tejidos adultos altamente expresado en diversos tipos de células, como linfocitos, células madre endoteliales, epiteliales y hematopoyéticas, fibroblastos estromales, e incluso células cancerosas [26]. Este receptor desempeña un papel crucial en las respuestas inmunitarias, y es fundamental para la organización y el funcionamiento adecuado de la médula ósea [2]. Además, junto con su ligando CXCL12, CXCR4 controla procesos importantes como el movimiento, el retorno y la supervivencia de las células madre hematopoyéticas en la médula ósea, permitiendo así que se desempeñen las funciones vitales de renovación y mantenimiento del sistema sanguíneo [2].

Estructuralmente, CXCR4 está formado por siete hélices transmembrana conectadas por bucles intracelulares y extracelulares (Figura 3). Esto le permite interactuar específicamente con moléculas pequeñas y proteínas, ya que las hélices crean un bolsillo en el que puede unirse el ligando, la quimiocina CXCL12 [27].

**Figura 3.** Superficie de CXCR4 destacando hendidura TMV-TMVI, residuos de unión a CXCL12 y cavidad identificada.



Fuente: García-Cuesta, *et al.* (2024)

### **Relevancia patológica**

La expresión de CXCR4 está frecuentemente elevada en diversos tipos de cáncer, incluyendo el de mama, ovario, próstata, melanoma y neuroblastoma [2]. Además, el eje CXCR4/CXCL12 está implicado en el desarrollo de enfermedades autoinmunes. Y junto con el receptor CD4, actúa como un correceptor clave utilizado por el VIH-1 para invadir células del sistema inmunitario [26], es por ello por lo que se convierte en una diana terapéutica atractiva.

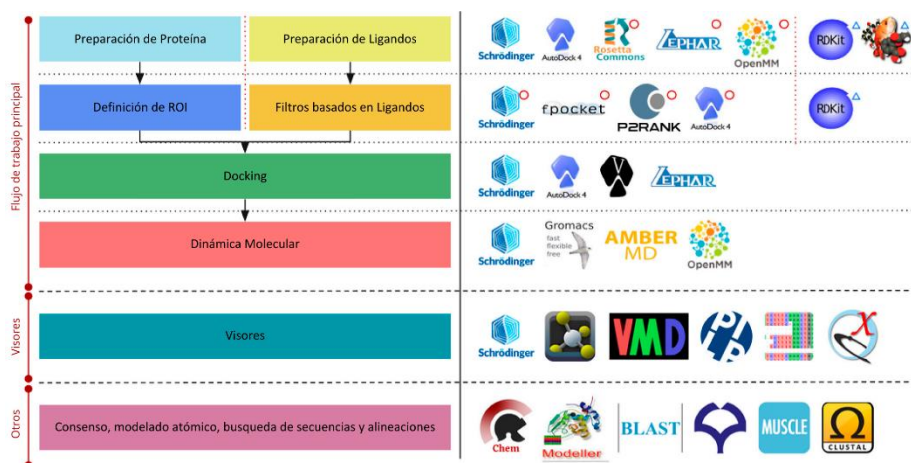
## Métodos para la optimización de ligandos

### Scipion

Scipion es un programa de código abierto desarrollado en el Centro Nacional de Biotecnología (CNB, CSIC), disponible de forma gratuita en <https://scipion.i2pc.es>. Este software facilita la integración de múltiples paquetes de software 3DEM (Microscopía Electrónica Tridimensional) al proporcionar una interfaz gráfica unificada que simplifica su uso tanto para programadores como para desarrolladores [28].

Los flujos de trabajo de los paquetes de procesamiento de imágenes EM son similares, pero las diferencias en formatos y convenciones dificultan su interoperabilidad. Con Scipion se pretende buscar una solución este problema integrando algoritmos de los principales paquetes 3DEM y facilitando su uso conjunto [28].

**Figura 4.** Scipion-chem integra programas organizados por etapas del flujo de trabajo VDS. Los programas específicos para objetivos están marcados con círculos rojos, y los específicos para ligandos con triángulos azules, además de incluir herramientas de visualización y utilidades adicionales.



Fuente: modificado de Del Hoyo, et al. (2023)

Scipion-chem es una extensión de Scipion orientada al cribado virtual de fármacos (VDS). Actualmente, integra 11 complementos de los programas más comunes en este campo, los cuales pueden utilizarse a través de su interfaz gráfica para ejecutar y analizar tareas típicas

de VDS (Figura 4). Además, incluye protocolos de consenso que combinan los resultados de diferentes programas para generar predicciones más sólidas [15].

Scipion-chem facilita la interoperabilidad entre estos paquetes de software, gestionando automáticamente archivos intermedios, parámetros y decisiones del usuario [15]. Todo ello está disponible de forma abierta en <https://github.com/scipion-chem>.

### **DeepFrag**

DeepFrag es una herramienta basada en la inteligencia artificial diseñada para mejorar ligandos en el desarrollo de fármacos disponible en línea a través de la página web del laboratorio de Durrant (<http://durrantlab.com/deepfrag>) [29]. Su principal función es que tiene capacidad para sugerir fragmentos y enlaces químicos que pueden añadirse al ligando para así mejorar la afinidad y estabilidad hacia la proteína diana.

El proceso de optimización con esta herramienta se basa, como se ha comentado, en un modelo de inteligencia artificial que estudia como un ligando se une a la proteína diana, analizando tanto la estructura de la proteína como las interacciones específicas entre la diana y los ligandos. A partir de este análisis, Deepfrag sugiere fragmentos moleculares y enlaces químicos que podrían añadirse de forma eficiente a ese fármaco para mejorar esa interacción. En pruebas, fue capaz de identificar el fragmento correcto en un 58% de los casos entre más de 6500 opciones [30].

Para facilitar su uso, DeepFrag cuenta con una aplicación web de código abierto que permite a los investigadores acceder a sus funcionalidades sin necesidad de conocimientos avanzados en programación [29]. Además, DeepFrag está integrado en Scipion, lo que amplía su accesibilidad y versatilidad. Sin embargo, una limitación aquí, es que, para utilizarlo, es necesario seleccionar manualmente el átomo del ligando al que se harán las pruebas de unión del fragmento, lo que, a priori parece una ventaja al tener todo el control, esta tarea puede volverse tediosa y propensa a errores, sobre todo cuando se trabaja con ligandos muy grandes.



### **AutoGrow4**

AutoGrow4 es una herramienta de código abierto que emplea algoritmos genéticos para el diseño de fármacos *de novo* y la optimización de fármacos candidatos. El objetivo principal de esta herramienta es generar nuevas estructuras moleculares con una alta afinidad y especificidad que se unan eficazmente a una diana específica, convirtiéndose en una herramienta muy valiosa en el campo de la medicina personalizada y el diseño de fármacos [31].

El funcionamiento de AutoGrow4 se basa en la evolución iterativa de una población inicial de compuestos químicos, donde cada compuesto es considerado como “candidato” para ser optimizado. En cada ciclo, esta herramienta utiliza operaciones de mutación y cruce para modificar y combinar estructuras existentes, generando nuevas moléculas con propiedades mejoradas. Las mutaciones se basan en una biblioteca de 94 reacciones químicas predefinidas, permitiendo modificar la estructura del compuesto original mediante cambios en sus enlaces, grupos funcionales o geometría. Por otro lado, el cruce combina dos compuestos diferentes identificando su subestructura común más grande y fusionando los fragmentos más favorables [31]. A través de la combinación de mutación y cruce, AutoGrow4 es capaz de explorar de forma muy eficaz una gran variedad de estructuras moleculares y seleccionar las que sean mas favorables en términos de afinidad por la diana.

Sin embargo, a pesar de todas las capacidades que presenta AutoGrow4, tiene una limitación importante que se tiene que tener en cuenta a la hora de usar esta herramienta en el diseño de fármacos. Una de las principales restricciones es que requiere librerías de compuestos grandes para funcionar correctamente y de forma óptima, lo que puede dificultar su uso en proyectos con pocos compuestos iniciales.

## Materiales y Métodos

### 1. Preparación de la Proteína/target:

La proteína CXCR4, conocida por su papel crucial en diversas funciones biológicas y en la migración celular, fue seleccionada como objetivo para este estudio. Se utilizó la estructura cristalina de CXCR4 depositada en el Protein Data Bank (PDB) bajo el código 3Oe0. La preparación de la proteína se realizó utilizando el entorno Scipion, siguiendo un procedimiento metódico:

#### 1.1. Importación de la estructura:

La estructura de la proteína fue importada desde la base de datos PDB utilizando el módulo “pwem - import atomicstructure” en Scipion [28].

#### 1.2. Preparación del receptor:

Para completar la estructura de la proteína, se utilizó el módulo “openmm” de Scipion [32], este módulo utiliza PDBFixer [33] para la preparación de receptores.

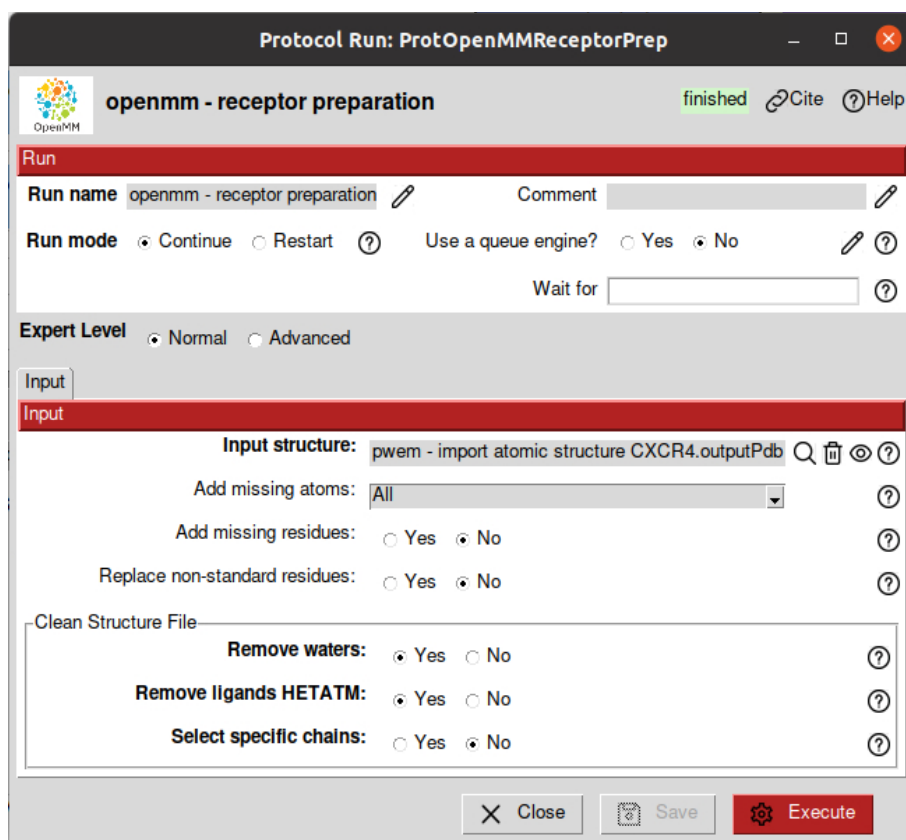
Con esta herramienta se llevó a cabo el proceso de adición de átomos faltantes. Este paso es fundamental, ya que la ausencia de átomos puede comprometer la calidad del modelo y la precisión del docking [34]. Scipion permite automatizar este proceso, garantizando que se añadan los átomos correctos y se mantenga la química adecuada.

Se procedió a eliminar todas las moléculas de agua y los ligandos no estándar (etiquetados como HETATM) de la estructura. Este proceso se realizó para minimizar interferencias durante el docking, asegurando que únicamente se consideraran los átomos de la proteína. La eliminación de estas moléculas fue verificada visualmente mediante PyMOL [35] (REF), confirmando que la estructura del receptor era adecuada para el análisis posterior.

En este paso, se decidió no añadir residuos faltantes, no se realizaron selecciones de cadenas específicas ni se reemplazaron residuos no estándar, esto hizo posible simplificar el modelo y enfocarlo en la región de interés, optimizando los recursos computacionales sin afectar la

calidad del análisis. La estructura de CXCR4 fue tratada como un todo, para así simplificar el proceso de docking.

**Figura 5.** Protocolo OpenMM para la preparación del receptor.



Scipion

## 2. Identificación de Pockets

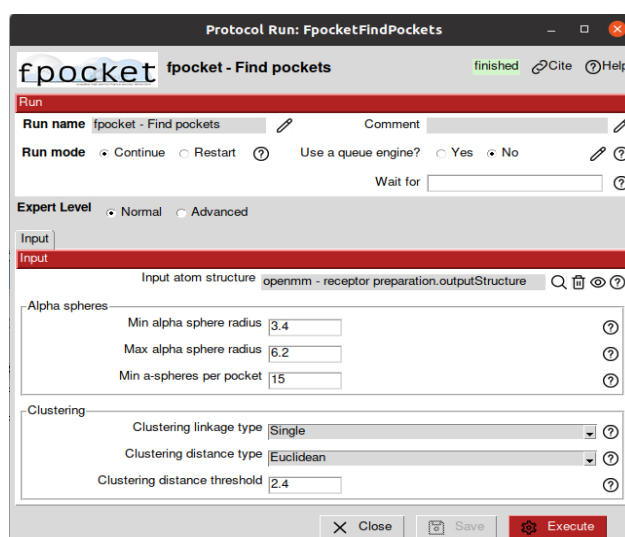
La identificación de posibles sitios de unión (*pockets*) en la proteína CXCR4 se realizó utilizando varias herramientas integradas en Scipion, con el objetivo de seleccionar las regiones donde los ligandos podrían interactuar:

### 2.1. fpocket

La herramienta fpocket es un algoritmo de predicción de *pockets* a partir de una estructura de proteína PDB, se utilizó para identificar los sitios de unión en la superficie de la proteína CXCR4 [36]. Se utilizaron los parámetros por defecto.

Se ejecutó el programa, y se generó una lista de 39 posibles pockets que podrían ser candidatos para el docking.

**Figura 6.** Protocolo fpocket para la identificación de pockets.



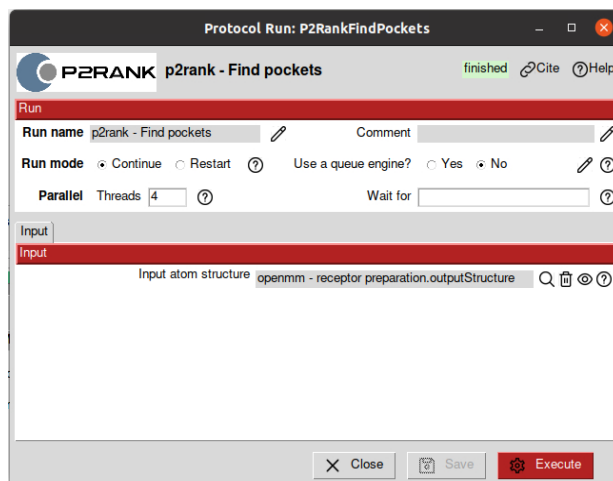
Scipion

### 2.2. p2rank

También se utilizó p2rank buscar *pockets* fuertes en la proteína [37]. Este software también busca los puntos de unión de ligandos a partir de una estructura dada.

Se ejecutó el programa, y se generó una lista de 29 posibles pockets que podrían ser candidatos para el docking.

**Figura 7.** Protocolo *p2rank* para la identificación de *pockets*.

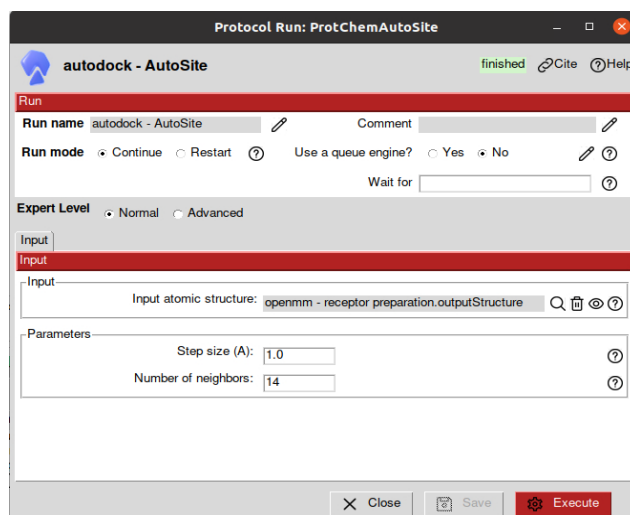


Scipion

### 2.3. autodock

Por otro lado, se empleó el módulo AutoSite de AutoDock para predecir *pockets* potenciales de interacción ligando-proteína [38]

**Figura 8.** Protocolo *AutoSite* para la identificación de *pockets*.



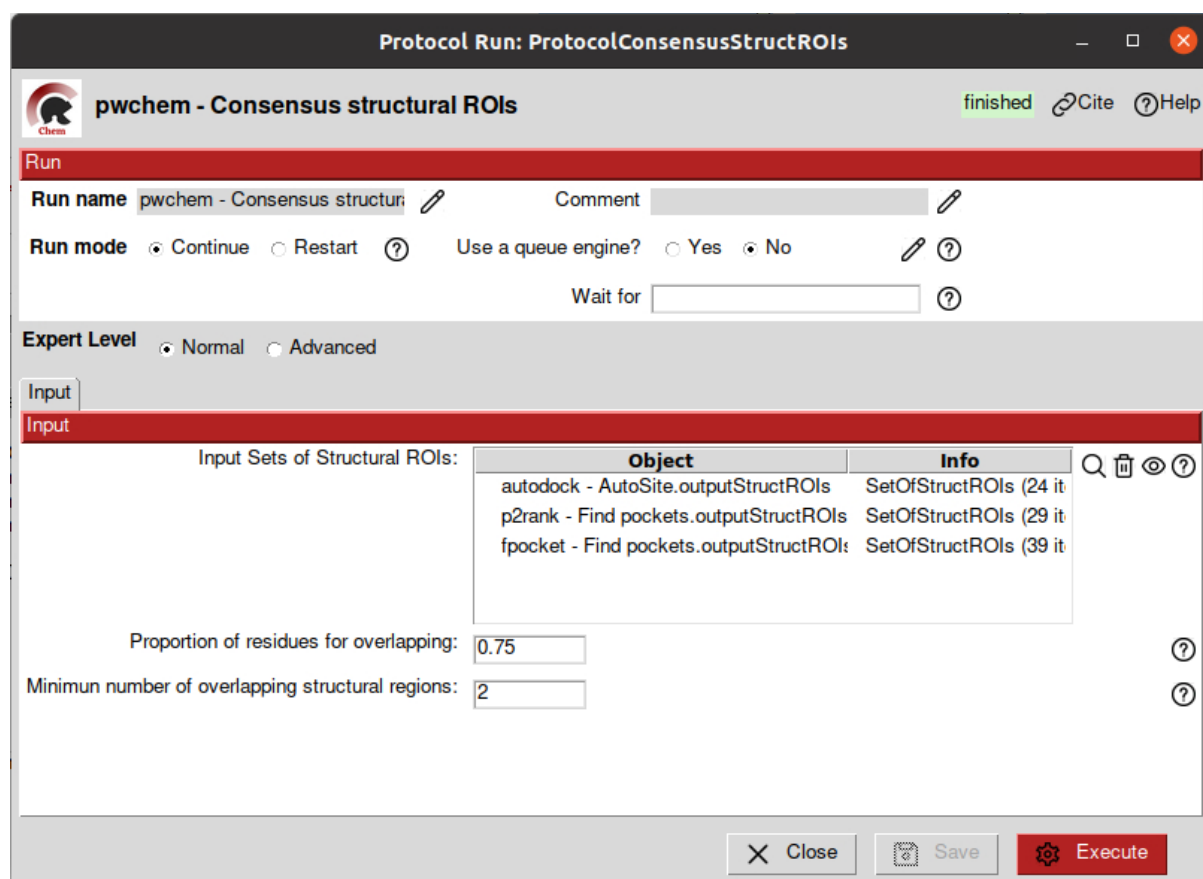
Scipion

Se ejecutó el programa, y se generó una lista de 24 posibles pockets que podrían ser candidatos para el docking.

#### 2.4. Consensus de ROIs Estructurales

Finalmente se generaron un total de 20 Regiones de Interés (ROIs) estructurales a partir de los resultados obtenidos mediante las tres herramientas mencionadas. Esta selección se llevó a cabo mediante un proceso de consenso, que combinó la información de fpocket, p2rank y AutoDock.

**Figura 9.** Protocolo para realizar un consenso de la información obtenida al identificar pockets en la proteína.



Scipion

Se realizó un análisis visual y cuantitativo para determinar cuáles pockets ofrecían la mejor combinación de características.

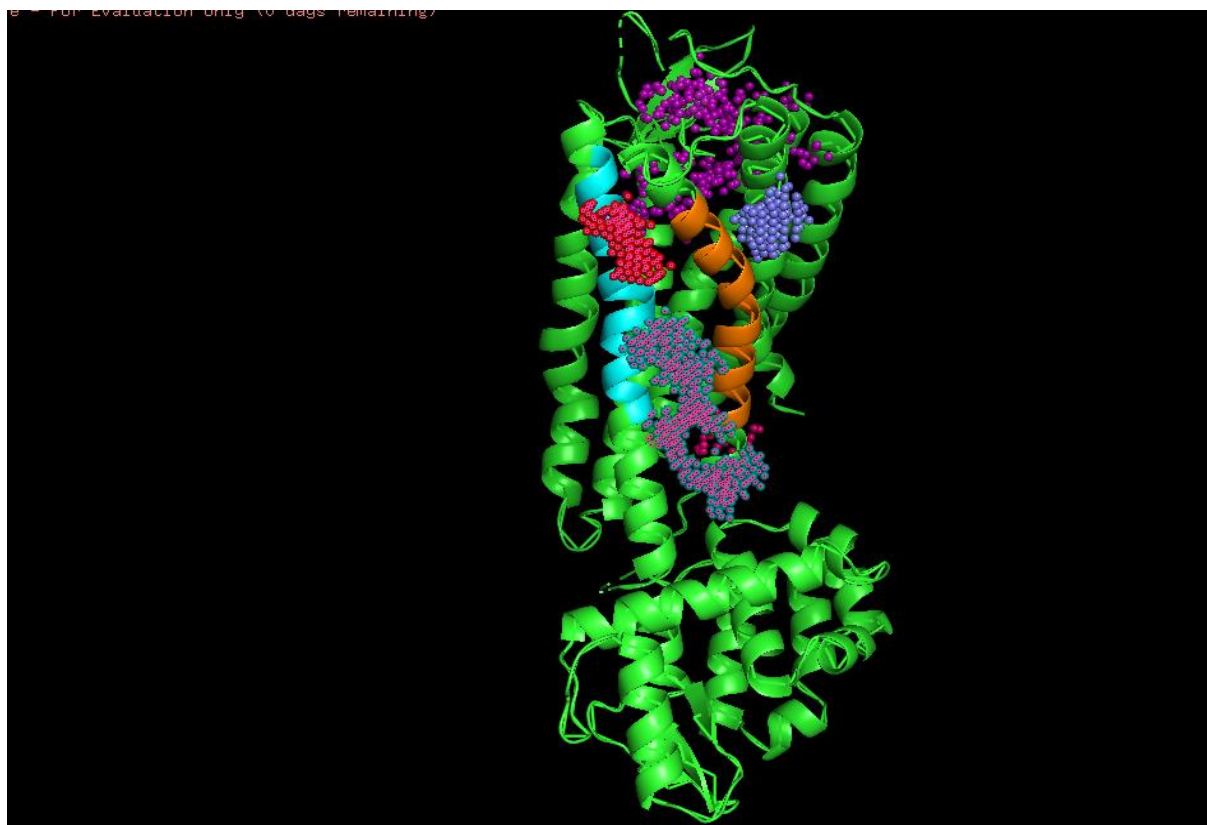
## 2.5. Selección del *pocket* de interés:

Finalmente, se seleccionó un pocket específico entre las hélices 5 y 6 para el proceso de docking. Esta decisión se basó en la accesibilidad del pocket, su tamaño y su potencial para interactuar con los ligandos seleccionados, lo que se consideró crítico para la posterior evaluación de la unión [2]. Para ello, se buscaron los residuos de las hélices 5 y 6 en UniProt [39]:

- Hélice 5: residuos 196-216
- Hélice 6: residuos 242-261

Realizando un análisis visual en PyMol y seleccionando estas dos hélices se seleccionaron los pockets que estaban entre estas hélices (Figura 10).

**Figura 10.** Estructura de la proteína y pockets identificados. Las helices 5 y 6 se colorean de azul y naranja respectivamente para identificar los pockets que se encuentran entre ellos (esferas rojas), que se corresponden con el pocket 3 y el pocket 19.



PyMol

De estos dos pockets, se seleccionó el pocket 19, ya que tiene un volumen y un score mayor, y una energía de unión más baja, lo que le confiere más estabilidad [40](Tabla 1).

**Tabla 1.** Atributos de cada pocket.

ID pocket	Volumen	Score	Energía
3	1496,3307	22,5200	-44,6500
19	4720,6890	40,3500	-183,4700

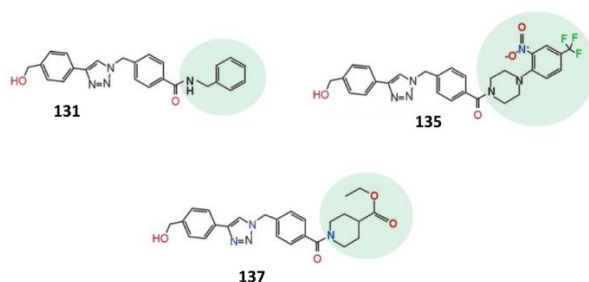


### 3. Preparación de ligandos

#### 3.1. Importación de ligandos

Los ligandos para el estudio se importaron desde los archivos .pdb utilizando el módulo “pwem - import small mols” en Scipion (Figura 11) [28].

**Figura 11.** Moléculas iniciales. En verde aparecen sombreadas las diferencias entre las cadenas laterales.

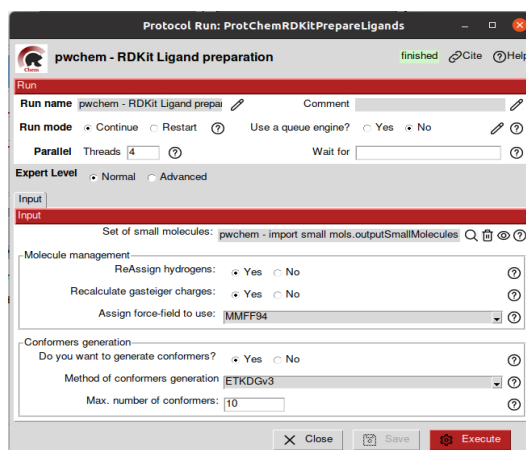


Fuente: García-Cuesta EM, et al. (2024)

#### 3.2. Preparación de ligandos

La preparación de los ligandos se llevó a cabo utilizando el módulo “RDKit Ligand Preparation” dentro de Scipion. Este proceso permitió una preparación sistemática de los ligandos para su posterior uso en los experimentos de docking y optimización.

**Figura 12.** Protocolo RDKit para la preparación de los ligandos.



Scipion

Se añadieron hidrógenos a las estructuras moleculares para completar la representación de los ligandos. Se asignaron cargas parciales adecuadas a los átomos, lo cual es esencial para los cálculos de docking. Para cada ligando, se generaron 10 conformeros energéticamente favorables.

## 4. Docking Molecular

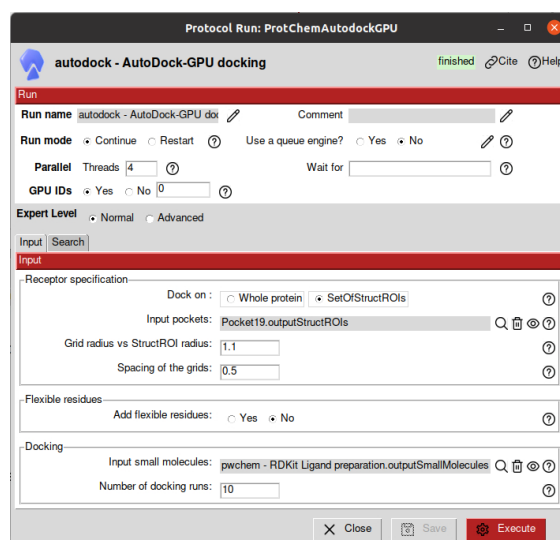
El proceso de docking molecular se llevó a cabo para evaluar la interacción entre la proteína CXCR4 y los ligandos seleccionados, empleando múltiples algoritmos para asegurar la robustez de los resultados.

### 4.1. Ejecución del Docking

Se realizaron experimentos de docking utilizando varios algoritmos disponibles en Scipion, cada uno con características específicas:

- **AutoDock-GPU:** Se utilizó este algoritmo para realizar docking utilizando la potencia de procesamiento gráfico (GPU), lo que permite una exploración más rápida y eficiente de las conformaciones de los ligandos.

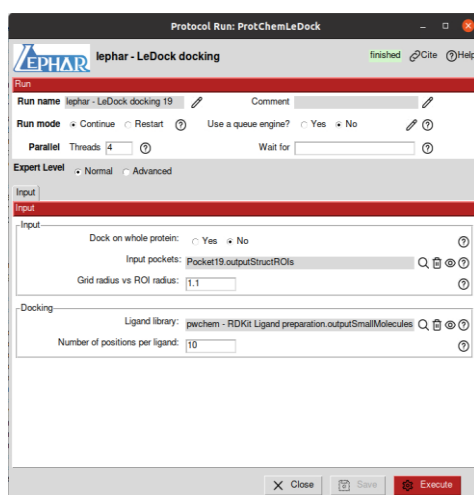
**Figura 13.** Protocolo Autodock para realizar el docking con AutoDock-GPU.



Scipion

- **LeDock**: Esta herramienta se empleó para obtener una visión complementaria sobre las interacciones potenciales entre los ligandos y el pocket. LeDock es conocido por su capacidad para realizar un scoring eficaz de las interacciones.

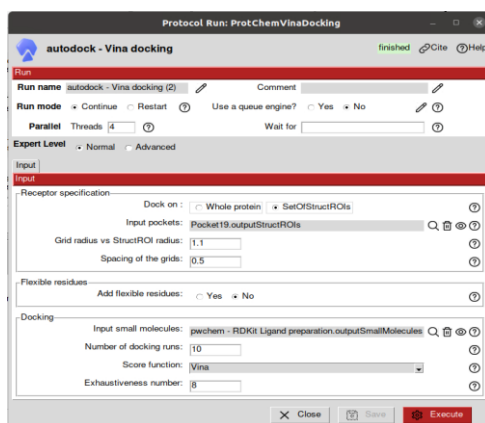
**Figura 14.** Protocolo lephar para realizar el docking con LeDock.



Scipion

- **Vina**: Se utilizó el algoritmo Vina debido a su eficacia en la predicción de las interacciones de ligandos con sitios de unión en proteínas. Vina ofrece un balance entre velocidad y precisión, lo que lo hace ideal para estudios de docking.

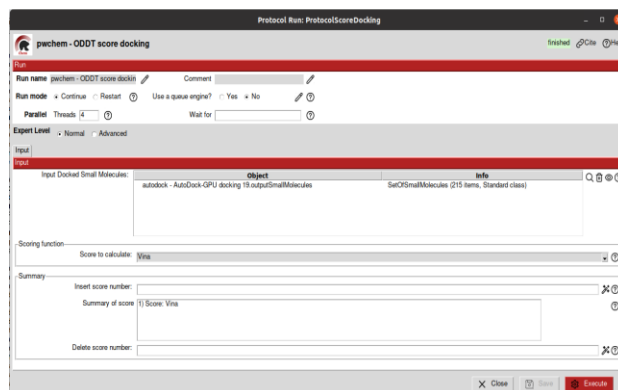
**Figura 15.** Protocolo autodock para realizar el docking con Vina.



Scipion

Para combinar y comparar las poses de acoplamiento generadas por cada uno de los programas, se evalúan utilizando la misma función de puntuación. En este caso, se utilizó el protocolo de puntuación ODDT para volver a puntuar todas las poses de acoplamiento con su función de puntuación Vina.

**Figura 16.** Protocolo de ODDT para reevaluar las puntuaciones de los acomplamientos.

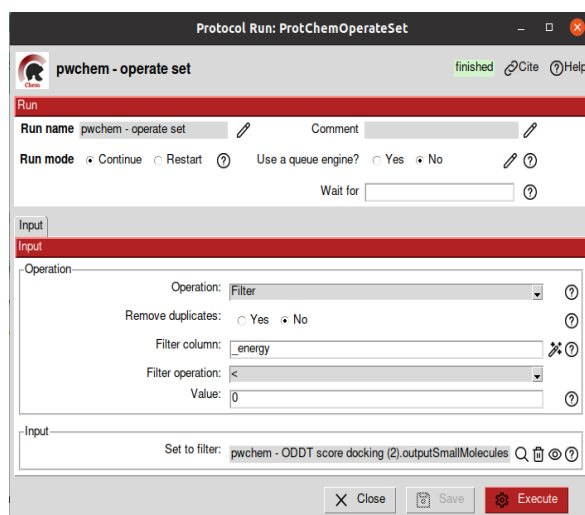


Scipion

#### 4.2. Análisis de Resultados del Docking

Los acoplamientos se filtraron seleccionando únicamente los que tienen una energía de acoplamiento negativa (Figura 17).

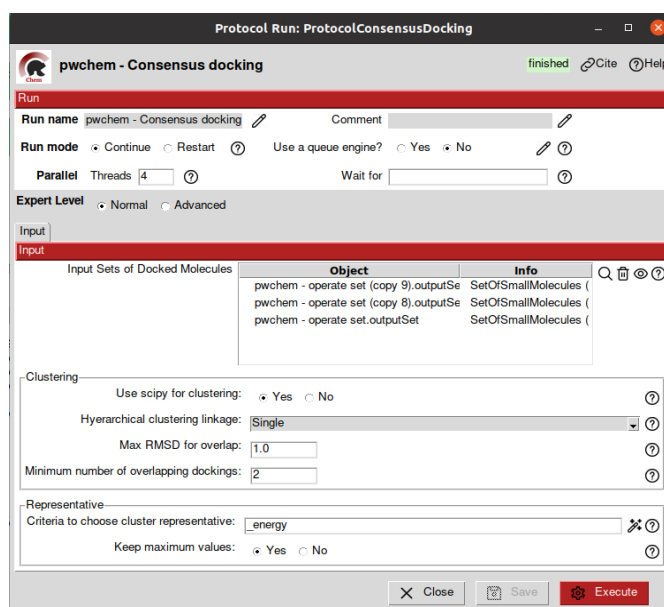
**Figura 17.** Protocolo para filtrar los acomplamientos.



Scipion

Tras esto, para extraer las poses de acoplamiento más prometedoras, se combinaron y clasificaron con el protocolo de consenso (Figura 18), seleccionando las moléculas que mostraban las mejores características de unión, con un resultado final de 8 moléculas.

**Figura 18.** Protocolo consenso.



Scipion

#### 4.3. Selección de moléculas

Tras la ejecución del docking, se seleccionaron las moléculas que presentaron las mejores puntuaciones de unión y energías. De cada una, se seleccionó la que tenía un ODDT score y una energía de unión mayor (Tabla 2), para la optimización posterior.

**Tabla 2.** Moléculas seleccionadas antes de optimizar.

Molécula	Energía	ODDT Score	Energía de Unión
131	-9.4100	7,3976	0,8734
135	-5,4800	6,8372	-11,2380
137	-8,7500	6,3736	31,8413

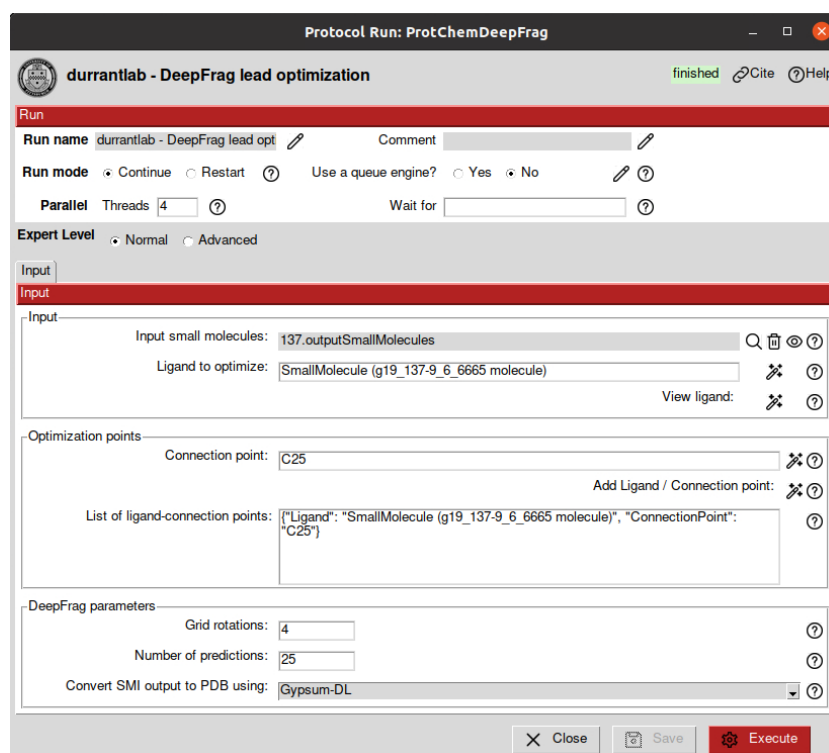
## 5. Lead Optimization

### 5.1. DeepFrag

La optimización de los ligandos seleccionados tras el proceso de *docking* se realizó utilizando la herramienta *DeepFrag* [30] esta herramienta se basa en extender el ligando con fragmentos moleculares de forma que esta modificación mejore la afinidad por el receptor.

Para cada molécula se calculó la energía libre de enlace con el protocolo MM-GBSA de Schrodinger [41] (Tabla 3) y se buscó visualmente con PyMol una región libre en cada molécula, identificándose esta posición como punto de conexión para las nuevos fragmentos propuestos por *DeepFrag*. En el caso de la molécula 131 se usó el átomo O2, para la molécula 135 el átomo O3 y para la molécula 137 el átomo C25.

**Figura 19.** Protocolo *DeepFrag* para lead optimization.



Scipion

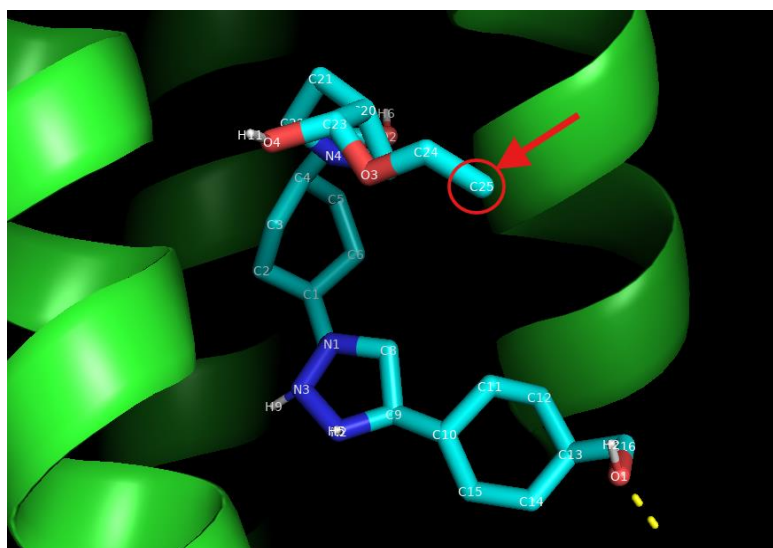
Tras ejecutar el protocolo de *DeepFrag* se obtuvieron 25 moléculas candidatas optimizadas. Con ellas se siguió el procedimiento de docking anterior (4. Docking Molecular), y al final de

este proceso se seleccionó la molécula más estable, es decir, aquella que presentaba una energía de unión menor (Tabla 3) [42].

**Tabla 3.** Moléculas seleccionadas después de optimizar.

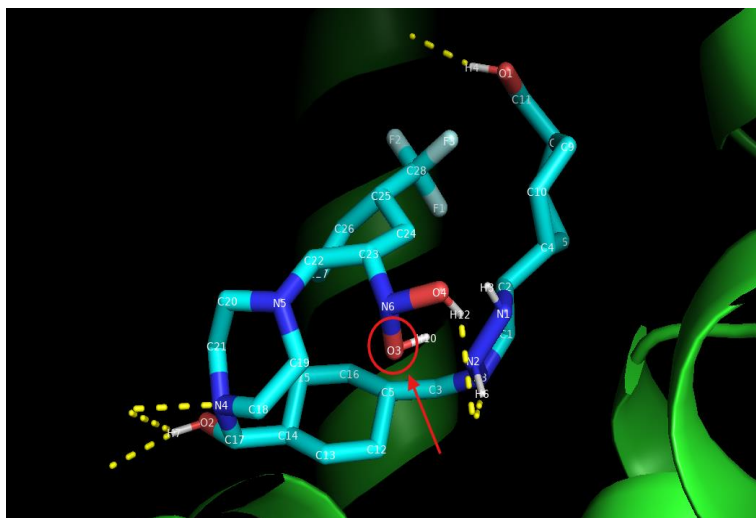
Molécula	Energía	ODDT Score	Energía de Unión
131	-5,6100	3,4315	-13,7349
135	-4,9200	4,3510	-13,5632
137	-5,2800	3,7114	-17,5982

**Figura 20.** Molécula 137 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína.



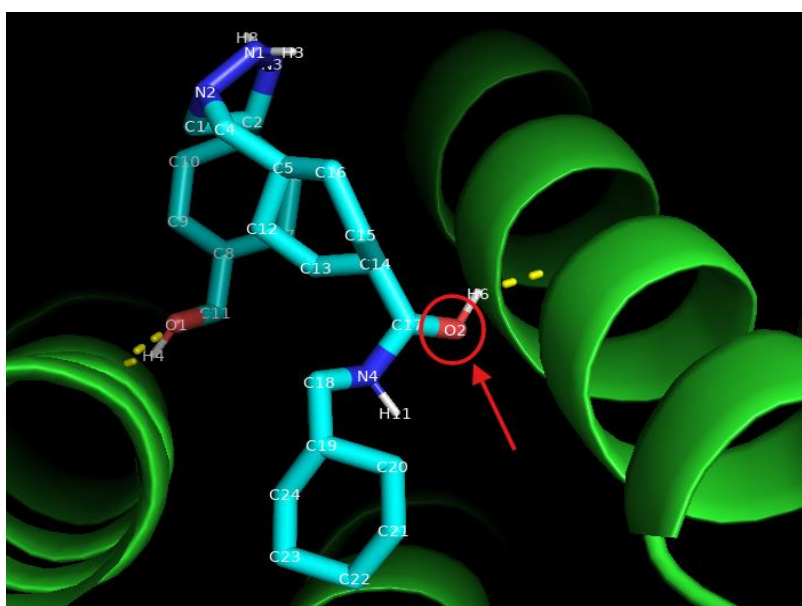
PyMol

**Figura 21.** Molécula 135 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína.



PyMol

**Figura 22.** Molécula 131 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína.



PyMol



## 5.2. AutoGrow4

Tras el docking, también se hizo la optimización de ligandos con la herramienta AutoGrow4 [31], este programa utiliza un código genético para desarrollar ligandos, es decir, partiendo de una población inicial de ligandos y aplica en un número determinado de generaciones cruce, mutación y filtros para generar iterativamente nuevas moléculas.

Se seleccionaron tres ligandos como punto de partida, los cuales fueron previamente identificados mediante el proceso de *docking*. Estos ligandos fueron preparados utilizando *RDKit* para garantizar su correcta estructura y representación 3D, compatible con *AutoGrow4*.

Al intentar ejecutar el proceso con *AutoGrow4*, dio error ya que este programa está diseñado para funcionar con una librería grande de compuestos iniciales, por lo que al tener solo tres, el proceso de crossover falló.

El proceso de Lead Optimization se llevó a cabo con los siguientes parámetros:

- **Número de generaciones:** 10.
- **Número de crossovers por generación:** 10.
- **Número de mutantes por generación:** 10.
- **Filtros de ligandos:** *LipinskiLenientFilter*.
- **Centro de la caja de *docking*:** [43.999, 3.133, 15.009].
- **Tamaño de la caja de *docking*:** [14.108, 14.108, 14.108].
- **Receptor:** CXCR4 (PDB ID: 3OE0).
- **Software de *docking*:** *QuickVina2*.

Se buscaron alternativas para solucionar este problema:

- Modificar parámetros de AutoGrow4: se cambió el número de crossovers por generación a 0 para que no combine fragmentos de estos 3 ligandos y solo haga mutaciones en ellos, pero el error siguió.

- Usar una biblioteca de ligandos iniciales más grande: El principal problema es que en estas bases de datos ZINC por ejemplo, no encuentro las moléculas problema.
- Usar una biblioteca de ligandos iniciales más grande:

Se importaron ligandos desde la base de datos de la FDA, importando unas 1600 moléculas nuevas. Estas se prepararon con RDKit y al igual que para las tres moléculas principales se añadieron hidrógenos a las estructuras moleculares, se asignaron cargas parciales y se generaron 5 conformeros energéticamente favorables.

Se hizo el docking que estas moléculas en el pocket donde se hizo anteriormente, con el fin de seleccionar un número manejable de moléculas para posteriormente, ejecutar el proceso de AutoGrow4 con las moléculas más favorables de la FDA junto con los tres ligandos del estudio. Se seleccionaron las 10 moléculas con una energía de enlace menor de la base de datos de la FDA para que junto a los tres ligandos iniciales, AutoGrow4 pueda cruzarlos, mutarlos y filtrar estos últimos con el fin de optimizarlos y hacerlos más estables.

**Tabla 4.** Moléculas seleccionadas después de optimizar con AutoGrow4.

Molécula	Energía	ODDT Score	Energía de Unión
131	-5,6100	3,4315	-13,7349
135	-4,9200	4,3510	-13,5632
137	-5,2800	3,7114	-17,5982

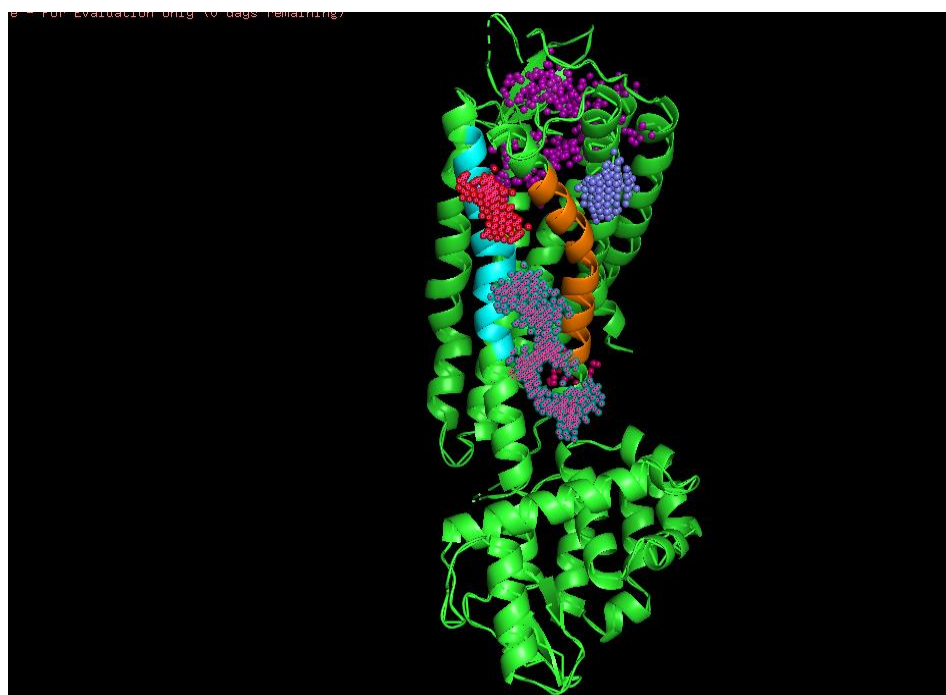
## Resultados

### Preparación de la proteína, ligandos e identificación de pockets.

Como resultado del análisis y preparación de la estructura de la proteína CXCR4, se obtuvo un modelo procesado en el que se eliminaron las moléculas de agua y ligandos no estándar. Esto permitió un análisis más preciso a la hora de buscar regiones de potencial interés para los experimentos de docking.

Se obtuvieron un total de 92 pockets en el análisis de sitios de unión de las moléculas. Sin embargo, como la región de interés en el estudio es la zona situada entre las hélices 5 y 6, se seleccionó el pocket situado entre estas hélices (Figura 23). El pocket seleccionado fue el pocket 19, debido a sus propiedades favorables: un volumen significativo ( $4720.69 \text{ \AA}^3$ ), un score elevado (40.35) y la energía de unión más baja obtenida ( $-183.47 \text{ kcal/mol}$ ). Esto sugiere que el pocket 19 es el más adecuado para la interacción con los ligandos seleccionados.

**Figura 23.** Estructura de la proteína y pockets identificados. Las hélices 5 y 6 se colorean de azul y naranja respectivamente para identificar los pockets que se encuentran entre ellos (esferas rojas), que se corresponden con el pocket 3 y el pocket 19.



PyMol

**Tabla 5.** *Propiedades de los pockets situados entre las hélices 5 y 6.*

ID pocket	Volumen	Score	Energía
3	1496,3307	22,5200	-44,6500
19	4720,6890	40,3500	-183,4700

Los ligandos se prepararon y se evaluaron realizar el docking. Se seleccionaron las moléculas **131**, **135** y **137** generándose 10 conformaciones energéticamente estables por de cada una, listas para su análisis con la proteína objetivo.

Los ligandos seleccionados cumplieron con las características necesarias para los estudios de docking molecular, incluyendo la asignación correcta de cargas parciales y la adición de hidrógenos a sus estructuras. En el caso del ligando 131, se obtuvo un volumen promedio de 245.12 Å<sup>3</sup>. El ligando 135, por su parte, presentó un volumen ligeramente superior, alcanzando los 265.48 Å<sup>3</sup>. Por último, el ligando 137 resultó ser el más compacto, con un volumen promedio de 230.67 Å<sup>3</sup>.

## Docking molecular

El docking molecular permitió analizar cómo interactúan los ligandos seleccionados con el pocket 19 de CXCR4. Los resultados que se muestran en la Tabla 6, obtenidos a partir de tres algoritmos diferentes (AutoDock-GPU, Vina y LeDock), fueron analizados para identificar las mejores poses y energías de unión. Solo se consideraron aquellas poses con energías negativas, ya que estas indican una interacción favorable entre el ligando y el receptor.

Estos resultados sugieren que el ligando 131 tiene el mayor potencial, ya que muestra una buena estabilidad y afinidad. Sin embargo, el ligando 137 también tiene propiedades interesantes que vale la pena considerar para futuros análisis.

**Tabla 6.** Energías de unión y puntuaciones ODDT de los ligandos seleccionados tras el docking molecular.

Molécula	Energía (kcal/mol)	ODDT Score	Energía de Unión GBSA (kcal/mol)
131	-9.4100	7,3976	0,8734
135	-5,4800	6,8372	-11,2380
137	-8,7500	6,3736	31,8413

## Lead optimization

### Optimización con DeepFrag

DeepFrag permite añadir fragmentos moleculares a las estructuras originales, con el objetivo de optimizar su interacción con el receptor CXCR4. Antes de comenzar, se identificaron las zonas de cada ligando donde podrían añadirse los fragmentos. Estas áreas tenían que estar libres y ser accesibles dentro del pocket.

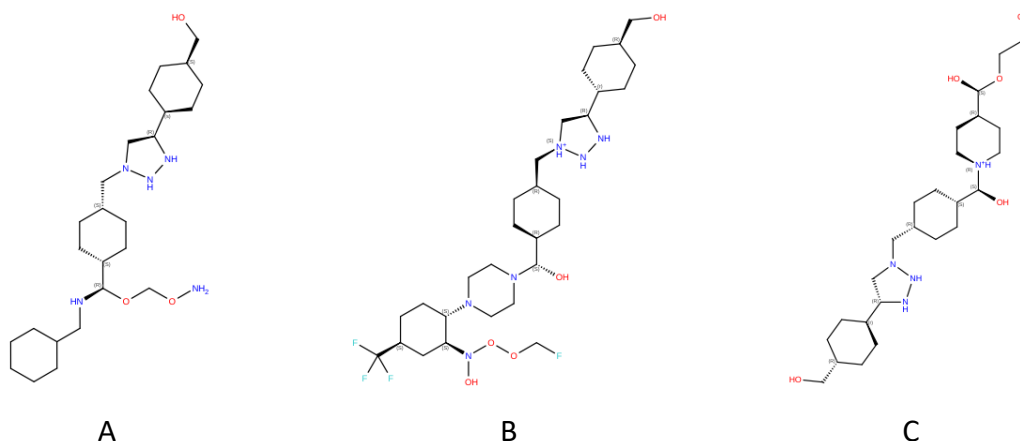
En el caso del ligando 131, se escogió el átomo O2 como el punto de conexión, ya que se encontraba en una posición estratégica para ampliar la interacción con el pocket. Para el ligando 135, se seleccionó el átomo O3, mientras que, para el ligando 137, se optó por el átomo C25, situado en un extremo opuesto al pocket, lo que lo hacía ideal para extender la molécula sin interferir con las interacciones existentes.

Una vez definido el punto de unión para cada ligando, se ejecutó el protocolo de DeepFrag, que generó 25 moléculas optimizadas por cada estructura inicial. Estas nuevas moléculas fueron evaluadas mediante un análisis de docking igual al utilizado anteriormente. A partir de este análisis, se seleccionaron aquellas moléculas que presentaban las energías de unión más bajas y los puntajes ODDT más altos, indicadores claros de una mejora en la afinidad y estabilidad de las interacciones.

**Tabla 7.** Comparativa de energías y puntuaciones de los ligandos originales y optimizados tras la optimización con DeepFrag

Molécula	Energía de unión inicial (kcal/mol)	ODDT Score inicial	Energía de unión optimizada (kcal/mol)	ODDT Score Optimizado
131	0,87	7,40	-13,73	3,43
135	-11,23	5,65	-13,56	4,35
137	31,85	6,37	-17,59	3,71

**Figura 24.** Moléculas tras optimizar con Deepfrag. A: ligando 131, B: ligando 135, C: ligando 137

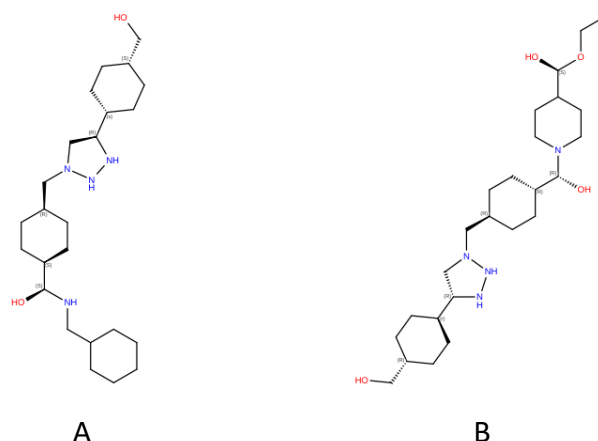


En general, los resultados fueron positivos. Las moléculas optimizadas mostraron una mejora notable en comparación con los ligandos originales, tanto en términos de afinidad como de estabilidad. Esto confirmó que la optimización con DeepFrag había sido efectiva. La Tabla 1 resume los resultados obtenidos tras esta etapa.

#### AutoGrow4

AutoGrow4 utiliza un enfoque genético aplicando cruces, mutaciones y filtros para generar de forma iterativa nuevos ligandos con mejores propiedades. Sin embargo, este proceso requiere un número inicial de moléculas suficientemente amplio, y esta fue la principal limitación al utilizar esta herramienta. Para afrontarla, se incorporaron 1600 compuestos de la base de datos de la FDA, los cuales fueron preparados y evaluados mediante docking. De este análisis se seleccionaron las 10 moléculas con mejores energías de unión y se combinaron con los 3 ligandos iniciales. Esto permitió generar una población inicial más robusta y se pudo ejecutar AutoGrow4 sin problemas.

**Figura 25.** Moléculas tras optimizar con AutoGrow4. A: 131, B: 137



Tras esto, AutoGrow4 generó moléculas optimizadas que mostraron mejoras significativas en términos de afinidad y estabilidad (Tabla 8).

**Tabla 8.** Moléculas seleccionadas después de optimizar con AutoGrow4.

Molécula	Energía inicial (kcal/mol)	ODDT Score inicial	Energía optimizada (kcal/mol)	ODDT Score Optimizado
131	-9,41	7,40	-10,42	6,6751
135	-5,48	6,84	-	-
137	-8,75	6,37	-10,31	6,98

## Resultados generales

La optimización con DeepFrag y AutoGrow4 permitió generar moléculas con mejores propiedades en comparación con las originales. El ligando 137, optimizado, destacó como el más estable, con una energía de unión de -17.59 kcal/mol tras DeepFrag y -10,31 kcal/mol tras AutoGrow4, consolidándose como el mejor candidato para estudios posteriores.



## Discusión y Conclusiones

### Preparación de Proteína/Ligandos e Identificación de pockets.

Se prepararon tanto las proteínas como los ligandos, eliminando moléculas de agua y ligandos no estándar para generar un modelo que permitiera un análisis más preciso de los pockets. Optimizar y preparar tanto las proteínas como los ligandos es un paso crucial en los estudios de cribado virtual. La optimización de la red de enlaces de hidrógeno y la minimización de la geometría de la proteína son fundamentales para mejorar significativamente los resultados de estos experimentos [43].

La identificación de un total de 92 pockets, cosa que evidenció la complejidad de la proteína, pero el análisis se centró en la región comprendida entre las hélices V y VI. Estas hélices son conocidas por experimentar cambios conformacionales significativos durante su activación [2,44].

De los dos pockets encontrado en esta zona, se eligió el pocket 19 por sus propiedades más favorables: un volumen considerable de  $4720,69 \text{ \AA}^3$ , un score alto de 40,35 y una energía de unión baja de  $-183,47 \text{ kcal/mol}$ , que le confiere más estabilidad. Estos parámetros indican que el pocket 19 tiene un alto potencial para interacciones estables con los ligandos seleccionados.

### Docking

El docking molecular inicial permitió evaluar cómo interactúan los ligandos seleccionados con el pocket 19 de CXCR4, identificando características clave para la afinidad y estabilidad de estas interacciones. Las interacciones entre el ligando y el receptor, aunque presentan algunas limitaciones importantes, suelen explicarse en términos de energía electrostática y de *van der Waals*, fuerzas que afectan directamente la estabilidad del complejo [45].

El análisis de los resultados obtenidos en este trabajo, coincide con los resultados reportados en otros estudios, como el de García-Cuesta et al. (2024). En particular, se observa que la molécula 131 tiene una puntuación energética más favorable ( $-9.41 \text{ kcal/mol}$ ) y un ODDT score de 7,40, esto puede deberse por la presencia de un grupo amida y un CH<sub>2</sub> adicional que conecta el anillo de fenilo [2]. Por otro lado, la molécula 137 aunque presenta una energía de

unión menos favorable (31,84 kcal/mol), obtuvo una energía de docking de -8,75 kcal/mol y un ODDT Score de 6,37. Esto es debido a que el grupo éster etílico de 137 es menos polar y mas hidrofóbico en comparación con el grupo nitro de la molécula 135, que obtuvo una energía de -5,48 kcal/mol y un ODDT Score de 6,84 [2].

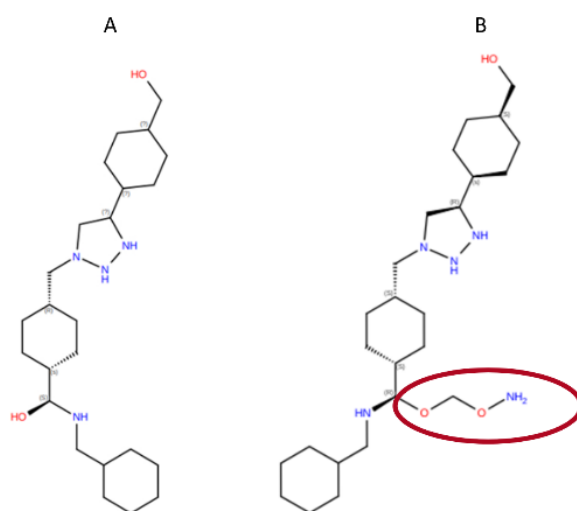
Estos resultados iniciales señalan a la molécula 131 como la más prometedora para futuras optimizaciones, gracias a su energía más negativa, que la convierte en la opción más favorable energéticamente. Por otro lado, las características estructurales de la molécula 137 también la posicionan como una buena candidata para el proceso de optimización.

### Lead Optimization

En este estudio se utilizaron DeepFrag y AutoGrow4 para optimizar las propiedades de interacción de ligandos dirigidos al pocket 19 de la proteína CXCR4, un objetivo terapéutico relevante en el tratamiento de enfermedades como el cáncer y el VIH [26]. Los resultados destacan el potencial de estas metodologías y abren nuevas perspectivas para el desarrollo de fármacos in silico.

La optimización con DeepFrag permitió extender los ligandos seleccionados añadiendo fragmentos moleculares en posiciones libres que se identificaron mediante un análisis visual en PyMOL. La adición de nuevos fragmentos moleculares a los ligandos mejora de forma significativa la afinidad de estos al receptor [30].

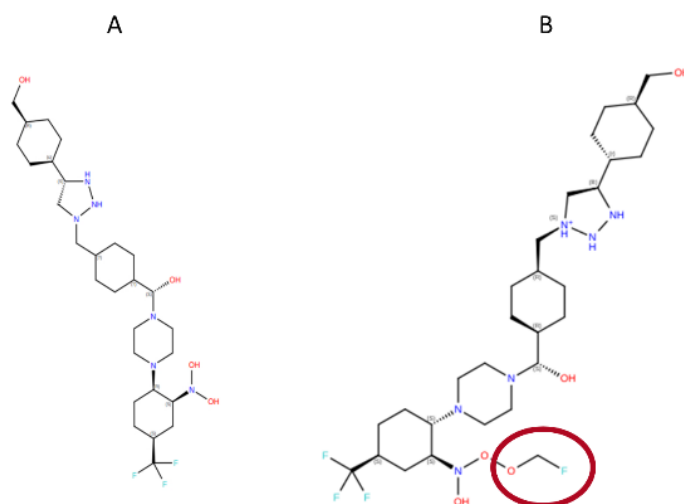
**Figura 26.** Comparación de la molécula 131 tras la optimización con Deepfrag.



La molécula 131 presentó una energía inicial de unión de 0,87 kcal/mol, lo que reflejaba una interacción poco favorable con el pocket 19 de la proteína CXCR4. Tras ser optimizada con DeepFrag, su energía de unión mejoró notablemente, alcanzando los -13,73 kcal/mol.

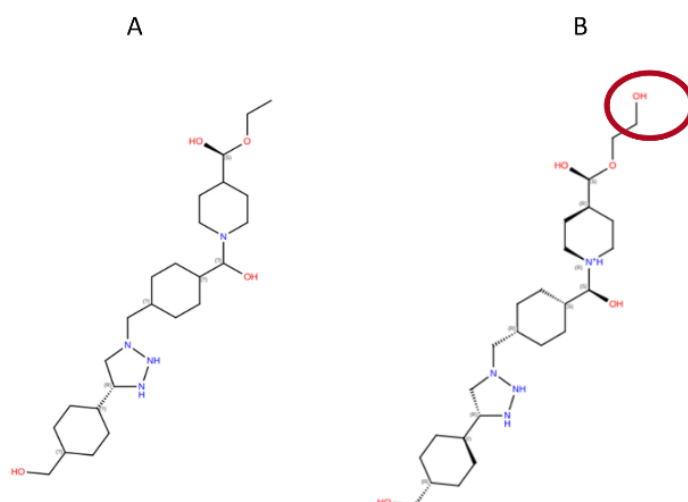
Lo mismo ocurre con la molécula 135 que partía de una energía inicial de unión ya favorable (-11.23 kcal/mol), y tras la optimización, este valor se mejoró a -13.56 kcal/mol.

**Figura 27.** Comparación de la molécula 135 tras la optimización con Deepfrag.



Por último, la molécula 137 destacó por mostrar el cambio más significativo tras la optimización, su energía inicial de unión era desfavorable (31.85 kcal/mol), pero tras el proceso de optimización, este valor mejoró drásticamente a -17.59 kcal/mol, convirtiéndose en el ligando con la mejor afinidad entre los optimizados

**Figura 28.** Comparación de la molécula 137 tras la optimización con Deepfrag.



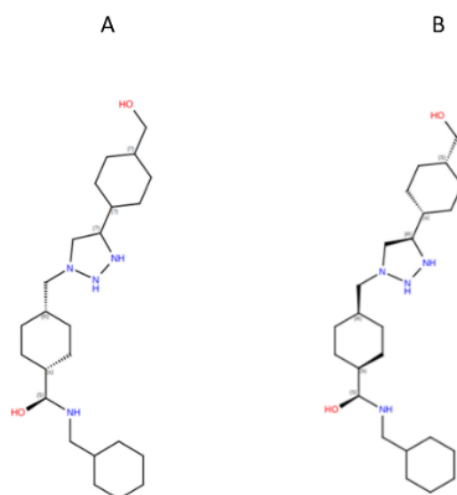
La selección de los átomos O2, O3 y C25 como puntos de conexión permitió extender las moléculas hacia regiones previamente desocupadas del pocket, maximizando el uso de áreas poco utilizadas y aumentando el potencial para formar enlaces de hidrógeno y contactos hidrofóbicos adicionales, siguiendo las recomendaciones del diseño molecular basado en fragmentos [30].

Las optimizaciones realizadas con DeepFrag lograron mejoras significativas en la energía de unión al maximizar las interacciones de los ligandos con el pocket 19 de CXCR4 mediante la adición de fragmentos moleculares, demostrando su eficacia para aumentar la afinidad en el diseño de fármacos.

AutoGrow4, una herramienta basada en algoritmos genéticos, se utilizó para explorar la diversidad molecular y generar variantes optimizadas de los ligandos iniciales mediante cruces, mutaciones y filtros iterativos que mejoran la afinidad y estabilidad de los compuestos [31]. Esta herramienta necesita una población inicial amplia para funcionar, lo que fue un problema al contar solo con tres ligandos. Para solucionarlo, se añadieron 1600 compuestos de la base de datos FDA [2], lo que reforzó la población inicial y permitió llevar a cabo las optimizaciones con éxito.

La molécula 131 mostró una reducción en su energía de unión tras la optimización con AutoGrow4, pasando de -9.41 kcal/mol a -10.42 kcal/mol.

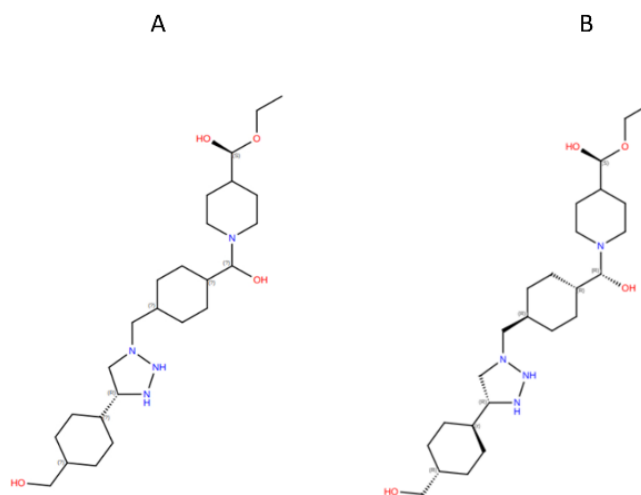
**Figura 29.** Comparación de la molécula 131 tras la optimización con AutoGrow4.



De manera similar, la molécula 137 presentó una mejora moderada, reduciendo su energía de unión de -8.75 kcal/mol a -10.31 kcal/mol. Estos resultados indican que las mutaciones, cruces y filtros introducidos aumentaron la afinidad de los ligandos por el pocket 19 de CXCR4.

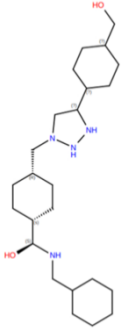
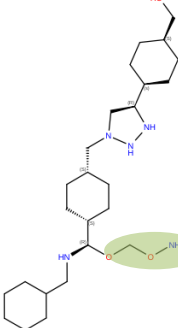
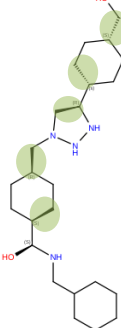
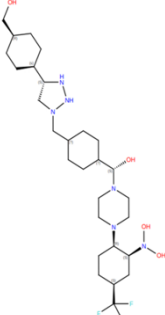
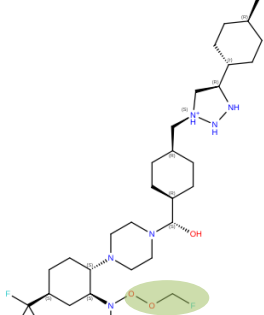

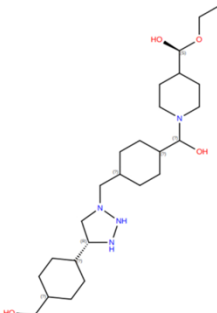
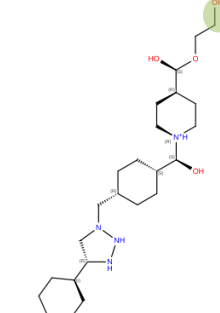
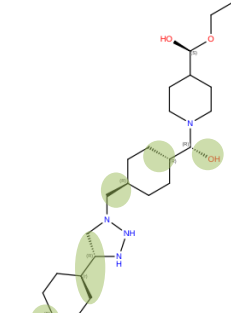
AutoGrow4 no encontró ninguna conformación ni mutación que mejore la molécula 135.

**Figura 30.** Comparación de la molécula 137 tras la optimización con AutoGrow4.



Con respecto a los ODDT Scores, los resultados obtenidos tras la optimización con ambos programas, revelaron una tendencia común entre los tres ligandos estudiados: una disminución significativa en comparación con sus valores iniciales. Este comportamiento sugiere que, aunque las modificaciones estructurales lograron mejorar considerablemente las energías de unión, pero podrían haber comprometido otras propiedades clave de los ligandos, como su estabilidad o accesibilidad dentro del pocket. Esto puede ser debido a que esta optimización se basa en la afinidad y unión ligando-proteína, sin tener en cuenta las propiedades farmacológicas.

**Tabla 9.** Resumen de la Comparación de las modificaciones por cada programa. Se colorean en verde las modificaciones que ha realizado cada programa.

Ligando Original	Modificaciones DeepFrag	Modificaciones AutoGrow4
		
		
		

La principal diferencia que se observa entre las modificaciones de ambos programas es que DeepFrag añade fragmentos a la molécula, y AutoGrow4 principalmente.

## Conclusiones:

Con este trabajo se desarrolló un flujo de trabajo bioinformático para la optimización de ligandos dirigidos a la proteína CXCR4, un receptor clave implicado en diversas patologías, como el cáncer y el VIH. Para ello, se emplearon herramientas como DeepFrag y AutoGrow4, que están integradas dentro de Scipion, las cuales permitieron realizar optimizaciones estructurales de los ligandos en función de sus interacciones con el pocket 19 de CXCR4. Los resultados obtenidos subrayan la eficacia de estas metodologías para mejorar la afinidad y la estabilidad de los ligandos, lo que representa un avance significativo en el diseño de fármacos terapéuticos más potentes.

- **Optimización de ligandos:** Durante el proceso de optimización, se lograron mejoras significativas en las energías de unión de los ligandos con el pocket 19 de CXCR4. Estas mejoras se alcanzaron mediante la adición de fragmentos moleculares para aumentar la superficie de interacción del ligando con la proteína. Además, se generaron variantes optimizadas utilizando tanto el enfoque molecular de DeepFrag como el enfoque genético de AutoGrow4. DeepFrag permitió una optimización precisa al seleccionar el átomo clave por el que se debía agrandar la molécula, lo que favoreció la creación de nuevas interacciones con el pocket. Por otra parte, AutoGrow4 exploró una amplia diversidad molecular, generando una población de ligandos candidatos con características estructurales y energéticas mejoradas.
- **Moléculas candidatas:** Entre los ligandos optimizados, el ligando 137 sobresale como el más prometedor, mostrando una energía de unión optimizada de -17,59 kcal/mol en DeepFrag y -16,45 kcal/mol en AutoGrow4. Estos valores no solo indican una afinidad elevada con CXCR4, sino también un potencial significativo para futuras investigaciones experimentales. La optimización mediante ambas herramientas permitió una reducción en la flexibilidad conformacional del ligando, lo que favorece la interacción con la proteína. Como resultado, el ligando 137 podría ser un excelente candidato para su evaluación en estudios *in vitro* e *in vivo*, con el fin de confirmar sus propiedades farmacológicas.
- **Limitaciones y futuras direcciones:** A pesar de los avances en la optimización, se observó una reducción en los ODDT Scores después de las optimizaciones, lo que sugiere que, aunque las energías de unión mejoraron, algunas propiedades

farmacológicas de los ligandos podrían haberse comprometido. Para superar esta limitación, se sugiere que futuros estudios se enfoquen en la simulación de las propiedades de solubilidad, permeabilidad y toxicidad de los ligandos optimizados, verificar la estabilidad y la biodisponibilidad de los mismos en sistemas biológicos.

En resumen, el flujo de trabajo computacional desarrollado en este estudio ha demostrado ser una herramienta poderosa para la optimización de ligandos dirigidos a CXCR4, destacándose la eficacia de DeepFrag y AutoGrow4 en la mejora de la interacción ligando-proteína. Sin embargo, las futuras investigaciones deberán abordar las posibles limitaciones relacionadas con las propiedades farmacológicas para asegurar que los compuestos optimizados puedan ser desarrollados como terapias efectivas.



## Bibliografía

- [1] Weiss ID, Jacobson O. Molecular Imaging of Chemokine Receptor CXCR4. *Theranostics* 2013;3:76. <https://doi.org/10.7150/THNO.4835>.
- [2] García-Cuesta EM, Martínez P, Selvaraju K, Ulltjörn G, Gómez Pozo AM, D'Agostino G, et al. Allosteric modulation of the CXCR4: CXCL12 axis by targeting receptor nanoclustering via the TMV-TMVI domain. *Elife* 2024;13. <https://doi.org/10.7554/ELIFE.93968>.
- [3] Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996;272:872–7. <https://doi.org/10.1126/SCIENCE.272.5263.872>.
- [4] Furusato B, Rhim JS. CXCR4 and Cancer. *Chemokine Receptors in Cancer* 2009:31–45. [https://doi.org/10.1007/978-1-60327-267-4\\_2](https://doi.org/10.1007/978-1-60327-267-4_2).
- [5] Domanska UM, Kruizinga RC, Nagengast WB, Timmer-Bosscha H, Huls G, De Vries EGE, et al. A review on CXCR4/CXCL12 axis in oncology: No place to hide. *Eur J Cancer* 2013;49:219–30. <https://doi.org/10.1016/J.EJCA.2012.05.005>.
- [6] Saotome K, McGoldrick LL, Ho J-H, Ramlall TF, Shah S, Moore MJ, et al. Structural insights into CXCR4 modulation and oligomerization. *Nature Structural & Molecular Biology* 2024 2024:1–11. <https://doi.org/10.1038/s41594-024-01397-1>.
- [7] De Clercq E. Mozobil® (Plerixafor, AMD3100), 10 years after its approval by the US Food and Drug Administration. *Antivir Chem Chemother* 2019;27. [https://doi.org/10.1177/2040206619829382/ASSET/IMAGES/LARGE/10.1177\\_2040206619829382-FIG3.JPEG](https://doi.org/10.1177/2040206619829382/ASSET/IMAGES/LARGE/10.1177_2040206619829382-FIG3.JPEG).
- [8] Caspar B, Cocchiara P, Melet A, Emelen K Van, Aa A Van der, Milligan G, et al. CXCR4 as a novel target in immunology: moving away from typical antagonists. *Future Drug Discov* 2022;4. <https://doi.org/10.4155/FDD-2022-0007>.
- [9] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* 2015, Vol 20, Pages 13384-13421 2015;20:13384–421. <https://doi.org/10.3390/MOLECULES200713384>.

- [10] Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. *Comput Biol Med* 2021;137:104851. <https://doi.org/10.1016/J.COMPBIOMED.2021.104851>.
- [11] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* 2010 9:3 2010;9:203–14. <https://doi.org/10.1038/nrd3078>.
- [12] Saldívar-González F, Prieto-Martínez FD, Medina-Franco JL, Saldívar-González F, Prieto-Martínez FD, Medina-Franco JL. Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química* 2017;28:51–8. <https://doi.org/10.1016/J.EQ.2016.06.002>.
- [13] Catacutan DB, Alexander J, Arnold A, Stokes JM. Machine learning in preclinical drug discovery. *Nature Chemical Biology* 2024 20:8 2024;20:960–73. <https://doi.org/10.1038/s41589-024-01679-1>.
- [14] Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry* 12:267 2016;12:2694–718. <https://doi.org/10.3762/BJOC.12.267>.
- [15] Del Hoyo D, Salinas M, Lomas A, Ulzurrun E, Campillo NE, Sorzano CO. Scipion-Chem: An Open Platform for Virtual Drug Screening. *J Chem Inf Model* 2023;63:7873–85. [https://doi.org/10.1021/ACS.JCIM.3C01085/ASSET/IMAGES/LARGE/CI3C01085\\_0014.JPEG](https://doi.org/10.1021/ACS.JCIM.3C01085/ASSET/IMAGES/LARGE/CI3C01085_0014.JPEG).
- [16] Medina-Franco JL, Fernán-Dezde Gortari E, Jesús Naveja J. Avances en el diseño de fármacos asistido por computadora. *Educación Química* 2015;26:180–6. <https://doi.org/10.1016/J.EQ.2015.05.002>.
- [17] Vattikuti MC. Improving Drug Discovery and Development Using AI: Opportunities and Challenges. *Research-Gate Journal* 2024;10.
- [18] Lin X, Li X, Lin X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* 2020, Vol 25, Page 1375 2020;25:1375. <https://doi.org/10.3390/MOLECULES25061375>.

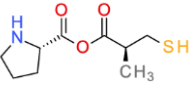
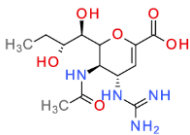
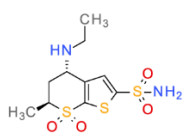
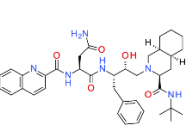
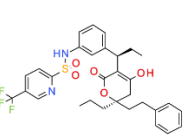
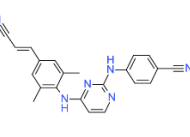
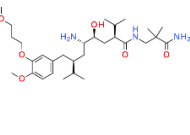
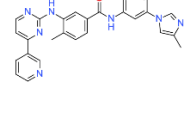
- [19] Prajapat P, Agarwal S, Telesara G. Significance of Computer Aided Drug Design and 3D QSAR in Modern Drug Discovery. 2017.
- [20] Yi J, Shi S, Fu L, Yang Z, Nie P, Lu A, et al. OptADMET: a web-based tool for substructure modifications to improve ADMET properties of lead compounds. *Nature Protocols* 2024 19:4 2024;19:1105–21. <https://doi.org/10.1038/s41596-023-00942-4>.
- [21] Liu S, Wu Y, Lin T, Abel R, Redmann JP, Summa CM, et al. Lead optimization mapper: Automating free energy calculations for lead optimization. *J Comput Aided Mol Des* 2013;27:755–70. <https://doi.org/10.1007/S10822-013-9678-Y/TABLES/2>.
- [22] Rognan D. The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacol Ther* 2017;175:47–66. <https://doi.org/10.1016/J.PHARMTHERA.2017.02.034>.
- [23] Von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 1993 363:6428 1993;363:418–23. <https://doi.org/10.1038/363418a0>.
- [24] AL References Christensen ET, Chang AS, Biochem EJ, Jagt V, Cushman DW, Cheung HS, et al. CUSHMAN Design of Potent Competitive Inhibitors of Angiotensin-Converting Enzyme. Carboxyalkanoyl and Mercaptoalkanoyl Amino Acids'1". *Proc Natl Acad Sci USA* 1975;79:15.
- [25] Craig JC, Duncan IB, Hockley D, Grief C, Roberts NA, Mills JS. Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral Res* 1991;16:295–305. [https://doi.org/10.1016/0166-3542\(91\)90045-S](https://doi.org/10.1016/0166-3542(91)90045-S).
- [26] Pozzobon T, Goldoni G, Viola A, Molon B. CXCR4 signaling in health and disease. *Immunol Lett* 2016;177:6–15. <https://doi.org/10.1016/J.IMLET.2016.06.006>.
- [27] Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, et al. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* (1979) 2015;347:1117–22. [https://doi.org/10.1126/SCIENCE.1261064/SUPPL\\_FILE/QIN-SM.PDF](https://doi.org/10.1126/SCIENCE.1261064/SUPPL_FILE/QIN-SM.PDF).
- [28] de la Rosa-Trevín JM, Quintana A, del Cano L, Zaldívar A, Foche I, Gutiérrez J, et al. Scipion: A software framework toward integration, reproducibility and validation in 3D

- electron microscopy. J Struct Biol 2016;195:93–9.  
<https://doi.org/10.1016/j.jsb.2016.04.010>.
- [29] Green H, Durrant JD. DeepFrag: An Open-Source Browser App for Deep-Learning Lead Optimization. J Chem Inf Model 2021;61:2523–9.  
[https://doi.org/10.1021/ACS.JCIM.1C00103/SUPPL\\_FILE/CI1C00103\\_SI\\_001.TXT](https://doi.org/10.1021/ACS.JCIM.1C00103/SUPPL_FILE/CI1C00103_SI_001.TXT).
- [30] Green H, Koes DR, Durrant JD. DeepFrag: a deep convolutional neural network for fragment-based lead optimization. Chem Sci 2021;12:8036–47.  
<https://doi.org/10.1039/D1SC00163A>.
- [31] Spiegel JO, Durrant JD. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. Journal of Cheminformatics 2020 12:1 2020;12:1–16.  
<https://doi.org/10.1186/S13321-020-00429-4>.
- [32] Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 2017;13. <https://doi.org/10.1371/JOURNAL.PCBI.1005659>.
- [33] GitHub - openmm/pdbfixer: PDBFixer fixes problems in PDB files n.d.  
<https://github.com/openmm/pdbfixer> (accessed January 3, 2025).
- [34] Chintha N, Jupudi S, Palathoti N, Bharathi JJ, Justin A. In-silico docking and molecular dynamic introspective study of multiple targets of AChE with Rivastigmine and NMDA receptors with Riluzole for Alzheimer’s disease. J Biomol Struct Dyn 2023;41:12620–31. <https://doi.org/10.1080/07391102.2023.2167119>.
- [35] PyMOL | pymol.org n.d. <https://www.pymol.org/> (accessed January 3, 2025).
- [36] GitHub - Discngine/fpocket: fpocket is a very fast open source protein pocket detection algorithm based on Voronoi tessellation. The platform is suited for the scientific community willing to develop new scoring functions and extract pocket descriptors on a large scale level. fpocket is distributed as free open source software. If you are interested in integrating fpocket in an industrial setting and require official support, please contact Discngine (www.discngine.com). n.d.  
<https://github.com/Discngine/fpocket> (accessed January 13, 2025).

- [37] GitHub - rdk/p2rank: P2Rank: Protein-ligand binding site prediction from protein structure based on machine learning. n.d. <https://github.com/rdk/p2rank> (accessed January 13, 2025).
- [38] Ravindranath PA, Sanner MF. AutoSite: an automated approach for pseudo-ligands prediction—from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* 2016;32:3142–9. <https://doi.org/10.1093/BIOINFORMATICS/BTW367>.
- [39] Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2023;51:D523–31. <https://doi.org/10.1093/NAR/GKAC1052>.
- [40] Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. *Biophys Rev* 2014;6:75–87. <https://doi.org/10.1007/S12551-013-0130-2/FIGURES/1>.
- [41] Schrödinger LLC. Schrödinger Release 2024-4: Desmond Molecular Dynamics System. New York, NY: 2024.
- [42] Schulz-Gasch T, Stahl M. Scoring functions for protein–ligand interactions: a critical perspective. *Drug Discov Today Technol* 2004;1:231–9. <https://doi.org/10.1016/J.DDTEC.2004.08.004>.
- [43] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;27:221–34. <https://doi.org/10.1007/S10822-013-9644-8/TABLES/9>.
- [44] Wescott MP, Kufareva I, Paes C, Goodman JR, Thaker Y, Puffer BA, et al. Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proc Natl Acad Sci U S A* 2016;113:9928–33. [https://doi.org/10.1073/PNAS.1601278113/SUPPL\\_FILE/PNAS.1601278113.SAPP.PDF](https://doi.org/10.1073/PNAS.1601278113/SUPPL_FILE/PNAS.1601278113.SAPP.PDF).
- [45] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* 2004 3:11 2004;3:935–49. <https://doi.org/10.1038/nrd1549>.

## Anexo I

**Tabla 10.** Lista de fármacos comercializados en los que el diseño asistido por computadora desempeñó un papel decisivo.

Compuesto	Estructura	Objetivo	Indicación	Papel del CADD
Captopril		Enzima convertidora de angiotensina	Hipertensión	Modelado del sitio activo por homología con la carboxipeptidasa A
Zanamivir		Neuraminidasa de la influenza	Gripe	El diseño <i>de novo</i> sugiere reemplazar el grupo hidroxilo de un análogo del estado de transición por una guanidina básica, mejorando potencia y selectividad
Dorzolamida		Anhidrasa carbónica	Glaucoma	Cálculos de energía sugieren agregar un grupo metilo en el anillo de 6 miembros para optimizar la complementariedad con la estructura cristalográfica del sitio activo de la enzima
Saquinavir		Proteasa del VIH-1	SIDA	Optimización basada en estructura de sustituyentes lipofílicos que llenan los 4 subbolsillos lipofílicos (S2', S1', S1, S2) del sitio de unión del sustrato
Tipranavir		Proteasa del VIH-1	SIDA	Sustitución de un anillo de cumarina (resultado de cribado HTS) por un 4-hidroxipirona e incorporación adicional de una sulfonamida en la posición meta para mejorar interacciones
Rilpivirina		Transcriptasa inversa del VIH	SIDA	Adaptación al sitio alostérico de unión no nucleósido y direccionamiento hacia el Trp299 conservado
Aliskiren		Renina	Hipertensión	Optimización basada en estructura del tamaño del grupo hidrofóbico que llena la gran cavidad S1-S3. La cadena lateral metoxialcoxi del inhibidor es esencial para una fuerte unión
Nilotinib		Abl-quinasa	Leucemia mieloide crónica	Sustitución de la N-metil piperazina de imatinib por un metilindol que se ajusta óptimamente a un subbolsillo de la estructura inactiva de la quinasa

## Índice de figuras

Figura 1. Esquema del desarrollo de fármacos.....	9
Figura 2. Diseño de fármacos asistido por ordenador.....	11
Figura 3. Superficie de CXCR4 destacando hendidura TMV-TMVI, residuos de unión a CXCL12 y cavidad identificada. ....	12
Figura 4. Scipion-chem integra programas organizados por etapas del flujo de trabajo VDS. Los programas específicos para objetivos están marcados con círculos rojos, y los específicos para ligandos con triángulos azules, además de incluir herramientas de visualización y utilidades adicionales.....	14
Figura 5. Protocolo OpenMM para la preparación del receptor.....	18
Figura 6. Protocolo fpocket para la identificación de pockets. ....	19
Figura 7. Protocolo p2rank para la identificación de pockets. ....	20
Figura 8. Protocolo AutoSite para la identificación de pockets.....	20
Figura 9. Protocolo para realizar un consenso de la información obtenida al identificar pockets en la proteína.....	21
Figura 10. Estructura de la proteína y pockets identificados. Las helices 5 y 6 se colorean de azul y naranja respectivamente para identificar los pockets que se encuentran entre ellos (esferas rojas), que se corresponden con el pocket 3 y el pocket 19.....	22
Figura 11. Moléculas iniciales. En verde aparecen sombreadas las diferencias entre las cadenas laterales. ....	24
Figura 12. Protocolo RDKit para la preparación de los ligandos.....	24
Figura 13. Protocolo autodock para realizar el docking con AutoDock-GPU. ....	25
Figura 14. Protocolo lephar para realizar el docking con LeDock.....	26
Figura 15. Protocolo autodock para realizar el docking con Vina. ....	26
Figura 16. Protocolo de ODDT para reevaluar las puntuaciones de los acomplamientos.....	27
Figura 17. Protocolo para filtrar los acomplamientos. ....	27

Figura 18. Protocolo consenso.....	28
Figura 19. Protocolo DeepFrag para lead optimization.....	29
Figura 20. Molécula 137 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína. ....	30
Figura 21. Molécula 135 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína. ....	31
Figura 22. Molécula 131 para optimizar. El átomo marcado es el punto de unión para la optimización y la línea amarilla discontinua marcan las interacciones polares entre la molécula y la proteína. ....	31
Figura 23. Estructura de la proteína y pockets identificados. Las helices 5 y 6 se colorean de azul y naranja respectivamente para identificar los pockets que se encuentran entre ellos (esferas rojas), que se corresponden con el pocket 3 y el pocket 19.....	34
Figura 24. Moléculas tras optimizar con Deepfrag.....	38
Figura 25. Moléculas tras optimizar con AutoGrow4. ....	39
Figura 26. Comparación de la molécula 131 tras la optimización con Deepfrag. ....	41
Figura 27. Comparación de la molécula 135 tras la optimización con Deepfrag. ....	42
Figura 28. Comparación de la molécula 137 tras la optimización con Deepfrag. ....	42
Figura 29. Comparación de la molécula 131 tras la optimización con AutoGrow4.....	43
Figura 30. Comparación de la molécula 137 tras la optimización con AutoGrow4.....	44



## Índice de tablas

Tabla 1. Atributos de cada pocket. ....	23
Tabla 2. Moléculas seleccionadas antes de optimizar.....	28
Tabla 3. Moléculas seleccionadas después de optimizar. ....	30
Tabla 4. Moléculas seleccionadas después de optimizar con AutoGrow4.....	33
Tabla 5. Propiedades de los pockets situados entre las hélices 5 y 6.....	35
Tabla 6. Energías de unión y puntuaciones ODDT de los ligandos seleccionados tras el docking molecular. ....	36
Tabla 7. Comparativa de energías y puntuaciones de los ligandos originales y optimizados tras la optimización con DeepFrag.....	37
Tabla 8. Moléculas seleccionadas después de optimizar con AutoGrow4.....	39
Tabla 9. Comparación de las modificaciones por cada programa. ....	45
Tabla 10. Lista de fármacos comercializados en los que el diseño asistido por computadora desempeñó un papel decisivo. ....	53