

MÉTODO DE CRIBADO VIRTUAL DE INHIBIDORES ENZIMÁTICOS BASADO EN MODELAMIENTO FARMACÓFORO DINÁMICO Y APRENDIZAJE AUTOMÁTICO

5

CAMPO DE LA INVENCION

La presente invención se encuentra relacionada con el campo tecnológico de los procedimientos para el desarrollo de medicamentos “in silico”, en especial, de los procesos de análisis computacional de bases de datos de compuestos, dirigidos a
10 identificar y seleccionar un número limitado de candidatos que posean la actividad biológica deseada sobre un blanco terapéutico específico.

ANTECEDENTES DE LA INVENCION

El desarrollo de medicamentos “in silico” comprende diferentes tipos de procesos
15 que involucran una serie de herramientas computacionales para la toma de decisiones en relación con la selección de mejores opciones de moléculas novedosas.

En este contexto, los movimientos de las proteínas son cruciales en el estudio de sus interacciones con otras moléculas. Dos modelos significativos utilizados para
20 describir la dinámica de las proteínas son el modelo de ajuste inducido de ligandos y la selección conformacional (Galburt & Tomko, 2017).

El modelo de ajuste inducido propone que la unión de un ligando provoca cambios conformacionales en la proteína para adaptarse a la presencia del ligando. Por otro lado, el modelo de selección conformacional sugiere que las proteínas pueden
25 existir en múltiples conformaciones, y la unión de un ligando favorece selectivamente una conformación específica. Por lo tanto, es muy importante considerar la flexibilidad del sitio activo de las dianas biológicas (proteínas o enzima) en los procedimientos de cribado virtual (HTVS) (Anderson et al., 2001).

Un enfoque prometedor para tener en cuenta la flexibilidad del receptor (proteína o
30 enzima) es el “*ensemble docking*” (acoplamiento masivo. En este método, la dinámica de la proteína (o diana biológica) se caracteriza, bien sea mediante simulaciones computacionales denominadas dinámicas moleculares (MD por sus siglas en inglés) o a partir del empleo de varias estructuras cristalinas. En ambos casos, se extrae una serie de conformaciones del sitio activo (Kapoor et al., 2022).
35 Estas estructuras se utilizan para realizar varios cálculos de acoplamiento, en lugar de acoplarse a una única conformación.

Con este enfoque se pretende reproducir el mecanismo de selección conformacional, en el que los ligandos seleccionan conformaciones proteína/ligando(Evangelista Falcon et al., 2019).

En 2012 *Osguthorpe* y sus colaboradores llevaron a cabo un estudio centrado en la obtención de estructuras de receptores para realizar procedimientos de *docking* de conjuntos moleculares(Osguthorpe et al., 2012). Como primer paso, se extrajeron las estructuras del sitio activo de los cristales de proteasa del VIH (135 cristales) y CDK2 (92 cristales). Por otro lado, se realizó una simulación de dinámica molecular de 1 ns del receptor de andrógenos (AR) y se extrajeron 200 estructuras de tiempos de simulación equidistantes. Las estructuras cristalinas y las estructuras seleccionadas de dinámica molecular se sometieron a un proceso de agrupación basado en los volúmenes del sitio activo(Osguthorpe et al., 2012).

Inicialmente, los volúmenes del sitio activo se calcularon utilizando el programa *SiteMap* de *Schrodinger* y, a continuación, se realizó un solapamiento de volúmenes para cada par de estructuras, dando como resultado una matriz de solapamiento de volúmenes. Luego, se llevó a cabo una agrupación jerárquica y se eligió el centroide de cada agrupación como estructura representativa. Como resultado, se obtuvieron 4 conformaciones representativas para la proteasa del VIH, CDK2 y AR. Al utilizar estas 12 conformaciones en los procesos de acoplamiento molecular (*docking*), obtuvieron un aumento del 30% en los compuestos activos recuperados en comparación con el uso de una única conformación. Lo anterior indica que el uso de varias conformaciones en el acoplamiento molecular permite recuperar un mayor número de activos.

En otro estudio de *Falcon* y colaboradores(Evangelista Falcon et al., 2019)⁴, se realizó un ensamble de acoplamiento molecular (*docking* masivo), comenzando con la ejecución de 4 simulaciones de dinámica molecular de los receptores acoplados a proteínas G en la membrana. Como resultado, se obtuvo un porcentaje de recuperación entre el 70-99% de los ligandos conocidos.

Los anteriores ejemplos ilustran las ventajas de realizar un “*ensemble docking*” en comparación con el “*docking*” en una única estructura de receptor.

El ajuste inducido es un fenómeno crucial en la interacción entre la acetilcolinesterasa (AChE) y los ligandos. En una investigación de *Senapati* y colaboradores(Senapati et al., 2005), se realizaron simulaciones de dinámica molecular de mAChE (AChE de ratón) en complejo con el ligando (syn-TZ2PA6) se compararon con simulaciones de apo-mAChE (enzima sin ligando). Las simulaciones demostraron que la proteína sufría cambios conformacionales causados por el ligando revelando la importancia del ajuste inducido y del uso de métodos que permitan considerar la flexibilidad del sitio activo de la diana biológica. Estos cambios incluían la reducción de la flexibilidad en residuos específicos, el

estrechamiento de las vías de acceso y la reorientación de aromáticos, lo que condujo a una mayor estabilidad y optimización de las interacciones en el sitio activo.

En relación con la modelización de farmacóforos, dicho concepto se encuentra asociado a los fragmentos moleculares y sus propiedades. La IUPAC (*International Union of Pure and Applied Chemistry*) define el farmacóforo como una característica estérica o electrónica necesaria para que una molécula interactúe de forma óptima con una diana (Langer & Hoffmann, 2006). Esta característica viene determinada por las interacciones que se producen entre la molécula y el receptor. Las características farmacóforas incluyen donante de enlace de hidrógeno (HBD por sus siglas en inglés), aceptor de enlace de hidrógeno (HBA por sus siglas en inglés), aromaticidad, hidrofobicidad, catiónico, aniónico, entre otros (Qing et al., 2014). Un farmacóforo no representa una molécula o un grupo funcional concreto, sino que es un concepto abstracto que da cuenta de las capacidades de interacción de un compuesto hacia una diana biológica, adicionalmente un farmacóforo describe las características estéricas y electrónicas más esenciales para la interacción óptima de una molécula con una diana biológica.

Un modelo farmacóforo, también conocido como hipótesis farmacóforo, es un conjunto de características farmacóforas responsables de la afinidad que una molécula puede mostrar hacia una diana biológica. La afinidad de un ligando por una diana farmacológica suele medirse por: el cambio en la energía libre (ΔG), la constante de inhibición (KI) o la concentración inhibitoria semi-máxima (IC50) (Lu & Li, 2010). Generar modelos farmacóforos basados en ligandos activos es esencial para identificar las características estéricas y electrónicas más relevantes que debe poseer un ligando para interactuar óptimamente con una proteína.

El modelado farmacóforo puede clasificarse en tres tipos:

- modelado farmacóforo basado en el ligando, en el que se conocen las moléculas activas, pero se desconoce la diana biológica.
- farmacóforo basado en el complejo, en el que se conocen tanto las moléculas activas como la diana biológica.
- farmacóforo basado en el receptor, en el que sólo se dispone de datos sobre la diana biológica.

Los puntos farmacóforos son las regiones que representan los lugares donde se producen contactos favorables proteína-ligando (Ertl et al., 2011). Esto último es válido para los farmacóforos basados en complejos, que utilizan la estructura del receptor en complejo con el ligando para la creación del modelo. En el caso de los

modelos farmacóforos basados en el ligando o en el receptor, la hipótesis podría no captar lugares en los que se producen contactos de unión favorables.

Una vez construida la hipótesis farmacológica, es esencial validar el modelo comparándolo con una biblioteca de moléculas activas y señuelos (*decoys*). Un señuelo es una molécula estructuralmente similar a las moléculas activas, pero presuntamente inactiva(Schaller et al., 2020).

Durante el proceso de validación, los compuestos activos y señuelos de la biblioteca se asignan a una de las cuatro categorías basadas en el modelo farmacóforo:

- verdadero positivo (molécula activa identificada como activa)
- falso positivo (molécula inactiva identificada como activa)
- verdadero negativo (molécula inactiva identificada como inactiva)
- falso negativo (molécula activa identificada como inactiva).

Se han establecido diversas métricas para evaluar la calidad de los modelos farmacóforos(Schaller et al., 2020). La principal son las curvas de características operativas del receptor (ROC por sus siglas en inglés), que son una herramienta habitual para evaluar el rendimiento de los modelos farmacóforos. Estas curvas trazan la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos (especificidad) en diferentes umbrales de puntuaciones de predicción del modelo. La sensibilidad mide la proporción de compuestos activos verdaderos identificados correctamente por el modelo, mientras que la especificidad mide la proporción de compuestos inactivos verdaderos identificados incorrectamente por el modelo. El área bajo la curva ROC (AUC por sus siglas en inglés) es una métrica ampliamente utilizada para evaluar el rendimiento de un modelo farmacóforo. EL AUC oscila entre 0 y 1; un modelo perfecto tiene un AUC de 1, y un modelo aleatorio tiene un AUC de 0,5. Un AUC más alto indica una mejor capacidad del modelo para discriminar entre compuestos activos e inactivos(Schaller et al., 2020).

A su vez, Los inhibidores de la AChE actúan interrumpiendo la reacción de hidrólisis de la acetilcolina (ACh) por la enzima acetilcolinesterasa. Al inhibir la AChE, se impide la degradación de la ACh, lo que eleva los niveles de ACh en las personas que padecen la enfermedad de Alzheimer (EA). Este enfoque concuerda con la hipótesis colinérgica y ofrece un tratamiento potencial para la EA. En la actualidad se dispone de inhibidores de la AChE, como la tacrina, el donepezilo, la rivastigmina y la galantamina, que funcionan como tratamientos para la EA. Sin embargo, algunos de estos compuestos generan efectos secundarios como fatiga excesiva, somnolencia, torpeza o debilidad(Akhoon et al., 2020). En consecuencia, el diseño de nuevos inhibidores de la AChE que sean más específicos hacia esta enzima y

que por tanto tengan menos efectos secundarios, se ha convertido en un interesante campo de investigación.

En 2020, Akhoun y sus colaboradores descubrieron un inhibidor basado en donepezilo llamado D1. Este inhibidor es capaz de interactuar con PAS y CAS; por lo tanto, se clasifica como un inhibidor dual. Se recogieron cristales de proteínas correspondientes a complejos proteína-ligando con los inhibidores Huperzina A (PDB 4EY5), Galantamina (PDB 4EY6) y Donepezilo (PDB 4EY7). Se utilizaron para calcular 3 modelos farmacóforos basados en proteína-ligando, utilizando el software *Schrodinger's Phase*. Para la huperzina A y la galantamina se descubrieron 4 características, en contraste con el donepezilo, donde se obtuvieron 5 características. Se extrajeron 102.436 compuestos de varias bases de datos de productos naturales y se sometieron a un procedimiento de cribado virtual (HTVS por sus siglas en inglés) a través de las 3 hipótesis farmacóforas. Los 100 mejores resultados ordenados según la función de puntuación del modelo farmacóforo se sometieron a un procedimiento de acoplamiento molecular. A partir de los resultados del acoplamiento molecular y de la inhibición experimental se descubrió el compuesto denominado D1. Los cálculos de energía libre de unión proteína-ligando realizados con el método mecánica molecular/ área superficial de Born generalizada (MM/GBSA por sus siglas en inglés) y las simulaciones de dinámica molecular indicaron que el donepezilo (-82,89 kcal/mol) tiene una mayor afinidad por el sitio activo que el D1 (-59,68 kcal/mol). Esto se debe a que los grupos químicos del donepezilo permiten la generación de resonancia de dobles enlaces que promueven una mayor localización de la carga. Sin embargo, los estudios in vivo indicaron que el D1 tiene una eficacia superior a la del donepezilo en el aumento de los niveles de acetilcolina, quizá debido a los efectos de solubilidad del ligando.

En el estudio de Gao y colaboradores, los investigadores utilizaron modelos de relación cuantitativa estructura-actividad en 3 dimensiones (3D-QSAR por sus siglas en inglés), acoplamiento molecular y simulaciones de dinámica molecular para descubrir derivados de rivastigmina (un conocido inhibidor de la AChE) con mayor actividad contra la AChE (Gao et al., 2022). Recogieron 25 compuestos derivados de la rivastigmina de otros artículos que presentaban datos de IC₅₀ y los utilizaron como conjunto de entrenamiento para identificar las características más relevantes para la actividad contra la AChE. Los investigadores descubrieron que la presencia de pequeños grupos donadores de carga podía potenciar la actividad de la rivastigmina.

Además, se generaron dos compuestos denominados L-1 y L-2, que posteriormente se sometieron a un procedimiento de acoplamiento molecular utilizando rivastigmina como control (Gao et al., 2022). Las puntuaciones de acoplamiento de L-1, L-2 y rivastigmina fueron de -8,58, -10,35 y -8,39 kcal/mol respectivamente. Posteriormente, se llevaron a cabo simulaciones de dinámica molecular y cálculos

por el método mecánica molecular/área superficial de Poisson Boltzmann (MM/PBSA por sus siglas en inglés) encontrándose los residuos Trp-86 y Trp-286 como los principales contribuyentes a la energía libre de unión para L-2. El inhibidor L-2 mostró la mayor afinidad por el sitio activo en comparación con los demás, ya que ocupó todo el PAS, a diferencia de L-1 y rivastigmina¹.

Se descubrieron nuevos inhibidores de la AChE a partir de estrategias de reposicionamiento farmacóforo basadas en el acoplamiento molecular, el modelado farmacóforo y la similitud estructural en un estudio de investigación realizado por Pérez-Sánchez y colaboradores (Pérez-Sánchez et al., 2021). Este estudio se dividió en dos enfoques: basado en el ligando y basado en el receptor. En el enfoque basado en el receptor, se extrajeron 11353 compuestos pertenecientes a una biblioteca de fármacos de Villapharma y se utilizaron para realizar un acoplamiento molecular en el sitio activo de la eeAChE (rayo eléctrico). Posteriormente, se generó un modelo farmacóforo basado en proteína-ligando a partir de la pose del inhibidor rivastigmina en el cristal PDB 1GQR. El modelo farmacóforo se utilizó para filtrar los mejores compuestos en el ranking de *docking*. De este modo, se obtuvieron 33 compuestos que fueron cribados utilizando la métrica de eficacia del ligando. Finalmente, se obtuvo un compuesto como resultado.

La acetilcolinesterasa (AChE) funciona como una enzima hidrolasa, facilitando la degradación de acetilcolina (ACh) a través de una reacción de hidrólisis. La inhibición de la AChE mantiene el potencial para mantener niveles elevados de ACh al impedir la descomposición de la ACh. En la actualidad, varios trabajos han demostrado la capacidad de elevar los niveles de ACh mediante la inhibición de la AChE que ofrece una vía potencial para reforzar la eficacia del tratamiento de la enfermedad de Alzheimer. Sin embargo, estos compuestos suelen tener efectos secundarios. En este sentido, la búsqueda de soluciones más selectivas y potentes de inhibidores se vuelve crucial para hacer frente a estos efectos secundarios, probablemente, derivados de la no interacciones específicas con otras enzimas.

En este contexto, en el enfoque basado en ligandos, se generaron modelos farmacóforos basados en ligandos empleando 1535 compuestos inhibidores de la AChE de la biblioteca del *Drugbank*. Al mismo tiempo, se realizaron cálculos de similitud entre los 11353 compuestos de *Villapharma* y los 1535 compuestos del *Drugbank* para encontrar entre los 11353 compuestos aquellos que estructuralmente más se parecían a las moléculas inhibidoras. La comprobación cruzada de los resultados del modelado farmacóforo y de los cálculos de similitud arrojó 108 compuestos. Combinando los resultados de los dos enfoques, se extrajeron 24 compuestos y se probaron experimentalmente para medir los datos de IC₅₀, obteniéndose 3 compuestos con IC₅₀ de 13,1, 16,02 y 6,22 μ M. Finalmente, estos resultados se explotaron para construir modelos QSAR, en los que se determinó que las características fisicoquímicas más relevantes para la

actividad eran un aceptor de enlace de hidrógeno, un grupo aromático y un grupo hidrofóbico(Pérez-Sánchez et al., 2021).

En un manuscrito de *Westhuizen* y colaboradores. se descubrieron nuevos inhibidores de la AChE a partir del cribado virtual, el cribado in vitro y las simulaciones MD(Van Der Westhuizen et al., 2022). Inicialmente, se seleccionaron 109 cristales de proteína tcAChE y 39 de huAChE. Se inspeccionaron visualmente las cavidades del sitio activo para identificar conformaciones únicas. Los autores mencionan un ejemplo: la comparación de los cristales de las proteínas 1ODC y 4W63. Presentaron un cambio conformacional en el bucle Trp-279-Ser-291 debido al ajuste inducido por el ligando, lo que demuestra la gran flexibilidad del sitio activo y la necesidad de considerar varias conformaciones para este receptor en los procedimientos de cribado virtual. Como consecuencia de esta inspección, se extrajeron 109 conformaciones.

Consecutivamente, se recuperaron 453 ligandos activos de la base de datos DUDE y se generaron 26250 señuelos. Las conformaciones, los activos y los señuelos se utilizaron en el cribado virtual y se generaron curvas ROC. Las 8 mejores conformaciones del receptor según el área bajo la curva ROC (AUC) se seleccionaron para la segunda etapa. En esta etapa, la biblioteca Biofocus de 20000 compuestos se acopló a los 8 receptores mediante acoplamiento molecular. Los 720 mejores compuestos según los resultados del *docking* se emplearon en experimentos de inhibición. Basándose en los experimentos de inhibición, se identificaron dos andamiajes denominados A y B. Los compuestos A4 y B18 mostraron la mayor potencia con un IC_{50} de 3,78 y 2,19 μ M respectivamente. Para identificar las características fisicoquímicas cruciales para la actividad de estos compuestos, se generaron varios descriptores moleculares para los activos. A continuación, se realizó un análisis PCA que reveló que las moléculas activas tienden a contener fragmentos aromáticos y polares. Por último, se realizaron simulaciones de dinámica molecular de A4 y B18 para identificar la importancia de los grupos cargados positivamente para la formación de interacciones catión- π con varios residuos del sitio activo.

Aunque estudios previos han explorado ampliamente las características estéricas y electrónicas de los inhibidores de la acetilcolinesterasa (AChE), sigue faltando un análisis exhaustivo de estos aspectos en las diferentes familias de inhibidores de la AChE, teniendo en cuenta la flexibilidad del sitio activo mediante simulaciones de dinámica molecular (MD) (Cheung et al., 2012; Langer & Hoffmann, 2006; Wilkinson et al., 2012). Al mismo tiempo, la plasticidad conformacional de las proteínas en las interacciones proteína-ligando desempeña un papel clave. Aunque algunos estudios han abordado esta cuestión mediante simulaciones con técnicas avanzadas (Cheung et al., 2012; Langer & Hoffmann, 2006; Wilkinson et al., 2012) , la

práctica de emplear conformaciones energéticamente desfavorables en la búsqueda de inhibidores de proteínas no es habitual, lo que presenta un aspecto intrigante que merece atención.

En el grupo de las técnicas de descubrimiento de fármacos, los modelos farmacóforos destacan como herramientas prometedoras(Cheung et al., 2012), ampliamente empleadas en la literatura(Cheng et al., 2017) Sin embargo, su potencial puede aprovecharse aún más mediante la introducción de un concepto novedoso: el modelado farmacóforo de conjunto.

La inteligencia artificial y los modelos de aprendizaje automático han cobrado importancia en el descubrimiento de fármacos(Branduardi et al., 2005; Kapoor et al., 2022)y pueden desempeñar un papel crucial en los procedimientos de cribado virtual.

En este contexto, la presente invención aporta una solución tecnológica que permite la adecuada selección de inhibidores enzimáticos a partir de la identificación de las características moleculares esenciales para inhibir selectivamente las dianas biológicas de interés, mediante un procedimiento de cribado virtual que combina modelos farmacóforos basados en ligandos, modelos farmacóforos de conjunto basados en complejos ligando-receptor junto con modelos de aprendizaje automático.

BREVE DESCRIPCIÓN DE LA INVENCION

En un único objeto la invención enseña un método de cribado virtual para la selección de inhibidores enzimáticos para dianas biológicas específicas, a partir de la identificación de las características moleculares esenciales para inhibir selectivamente la enzima (diana biológica) deseada, basado en modelamiento farmacóforo dinámico y aprendizaje automático.

El modelamiento farmacóforo dinámico se fundamenta en dos componentes: modelo basado en ligando y modelo basado en complejo. En una modalidad de realización de la invención, para el enfoque basado en ligando, se descargaron 4.643 inhibidores de AChE conocidos y se clasificaron en 70 grupos o familias según la estructura molecular. Se seleccionaron nueve familias representativas mediante parámetros estadísticos para un análisis posterior. Para el enfoque basado en complejo se descargaron 61 estructuras cristalinas de huAChE en formato PDB y se seleccionaron 18 según ciertos criterios.

Se realizó un estudio combinatorio de RMSD por pares en estas 18 estructuras, identificando 17 residuos flexibles. Luego, los dos enfoques se combinaron utilizando un enfoque de acoplamiento de ajuste inducido, en el que los ligandos

representativos (centroide) de cada una de las nueve familias se acoplaron al receptor de AChE.

Se realizaron nueve simulaciones MD (dinámica molecular) de 50 ns basadas en las poses acopladas con ajuste inducido que se obtuvieron en el paso anterior, junto con una simulación adicional del complejo AChE-galantamina. Las conformaciones de la proteína se extrajeron de estas simulaciones y se utilizaron en un enfoque de acoplamiento de conjunto, en el que todos los compuestos de cada una de las nueve familias se acoplaron con las conformaciones de la proteína extraídas.

Los compuestos que obtuvieron el mejor puntaje en la etapa de acoplamiento de conjunto, es decir, aquellos que obtuvieron un XP Gscore más negativo, se utilizaron para la generación de modelos de aprendizaje automático.

El enfoque basado en ligando utilizó para la selección de ligandos los puntajes de acoplamiento resultantes y los valores experimentales de IC50 para generar una clasificación de afinidad, identificando los compuestos más activos de cada familia. Estos compuestos se utilizaron para generar modelos de farmacóforos basados en ligando.

Para el enfoque basado en complejo, se utilizó el complejo receptor y compuesto más activo de cada familia para extraer conformaciones del sitio activo de las simulaciones MD de 500 ns, análisis RMSD y 41 cálculos fisicoquímicos TRAPP. Estas conformaciones se utilizaron para generar modelos de farmacóforos basados en complejos los cuales en conjunto forman los modelos de farmacóforos dinámicos.

En una modalidad de realización, la selección de los nuevos inhibidores mediante el procedimiento de cribado virtual objeto de la presente descripción, obtuvo la identificación de nueve moléculas con potencial inhibidor de la AChE, los cuales pueden emplearse como punto de partida en el diseño de nuevos fármacos para el tratamiento de Alzheimer. El procedimiento de cribado virtual objeto de la presente descripción puede emplearse en otras enfermedades.

El método de cribado virtual de inhibidores enzimáticos basado en modelamiento farmacóforo dinámico y aprendizaje automático emplea un protocolo in silico exhaustivo que incorpora: las etapas que se describen a continuación:

Búsqueda de diversidad estructural. Seleccionar ligandos representativos de inhibidores de la enzima objetivo (AChE en una modalidad particular de realización del método) por agrupación. Para esto se emplea un algoritmo de agrupación por similitud estructural de las moléculas. La búsqueda se realiza en una base de datos de literatura científica utilizando en una modalidad particular de realización la suite Schrodinger 2023.1 (Schrodinger, LCC).

Para realizar el análisis de agrupamiento, se emplean las siguientes configuraciones: precisión de 64 bits, tipo de huella digital radial, criterio de similitud de Tanimoto y método de vinculación promedio. Estas configuraciones se seleccionan para garantizar una alta precisión y eficiencia en el proceso de agrupación. Se utilizó el criterio de similitud de Tanimoto para medir el grado de similitud entre cada par de moléculas y el método de enlace promedio se utilizó para determinar la distancia entre grupos.

Exploración conformacional del receptor. Se introducen los ligandos seleccionados en complejo con la proteína y se analiza su dinámica molecular, combinación de resultados computacionales y experimentales, a partir de la métrica denominada *Yasser número uno* (YN1) (métrica creada por los autores). Empleando esta métrica se logra integrar los datos experimentales reportados en literatura y los resultados computacionales obtenidos. A continuación, se detalla la métrica:

$$LE_i = -\frac{XP\ GScore_i}{MW_i} \quad i = 1, 2, \dots, N$$

$$YN = \frac{LE_i - LE_{min}}{LE_{max} - LE_{min}} + \left(1 - \frac{\log_{10} IC50_i - \log_{10} IC50_{min}}{\log_{10} IC50_{max} - \log_{10} IC50_{min}}\right)$$

Donde LE_i es la eficiencia del ligando, MW_i es el peso molecular, $\log_{10} IC50_i$ es el log base 10 de IC50, max es el valor máximo dentro del *clúster*, min es el valor mínimo dentro del *clúster*, y N es el número de moléculas del *clúster*.

El método involucra, a su vez, etapas de combinación de múltiples modelos conformacionales del receptor (Modelo Dinámico), combinación de modelos de aprendizaje de máquina como filtro en la búsqueda de nuevos inhibidores y la selección diferencial de inhibidores para pruebas experimentales, a partir de la métrica denominada *Yasser número dos* (YN2) (métrica creada por los autores)..

El método objeto de divulgación emplea el aprendizaje de máquina ML como un “filtro” extra que permite reducir la dimensionalidad del problema. Por ejemplo, al realizar una búsqueda en una base de datos comercial, se logra pasar del análisis de aproximadamente 12 millones de estructuras a 1,2 millones, es decir, reducir el número de estructuras en un orden de magnitud, todo esto con un mínimo de pérdidas de información. Lo anterior implica reducir significativamente el tiempo y el costo computacional en la búsqueda de nuevos fármacos.

Selección diferencial de inhibidores para pruebas experimentales

Una vez se ha completado el flujo de las anteriores etapas del método, aún se cuenta con bastantes moléculas candidatas, lo cual, se traduce en mayores tiempos

de identificación de nuevos medicamentos y, por tanto, costos elevados. De hecho, con una gran cantidad de moléculas no es posible emplear herramientas “*in silico*” más precisas, pues estas requieren más tiempo y poder de cómputo.

Por lo anterior, el método objeto de descripción aprovecha la información recolectada en la misma y con ella se ha desarrollado una nueva métrica que, dentro del contexto de la presente memoria descriptiva se denominará *Yasser número dos* (YN2) en la cual se combinan los resultados generales del *software Schrodinger* con datos propios de la dinámica del sistema biológico. Empleando YN2 se produce una reducción significativa de tiempos y costos computacionales.

10

$$\hat{C}_{unstable_i} = \frac{C_{unstable_i}}{C_{unstable\ total}} \quad \hat{C}_{stable_i} = \frac{C_{stable_i}}{C_{stable\ total}} \quad i = 1, 2, \dots, N$$

$$if\ Ph_{score_i} \geq Ph_{score\ min} \quad then \quad \widehat{Ph}_{score} = \frac{Ph_{score_i} - Ph_{score\ min}}{Ph_{score\ max} - Ph_{score\ min}} \quad and$$

$$YN2_i = \frac{(\hat{C}_{unstable}) + (\hat{C}_{stable}) + (\widehat{Ph}_{score})}{3}$$

$$if\ Ph_{score_i} < Ph_{score\ min} \quad then \quad \widehat{Ph}_{score} = 0 \quad and \quad YN2_i = 0$$

15 Donde N es el número de moléculas en el procedimiento de validación o cribado virtual, $C_{unstable_i}$ es el número de modelos farmacóforos derivados de conformaciones inestables que se ajustan a la molécula i , $C_{unstable\ total}$ es el número total de modelos farmacóforos derivados de conformaciones inestables, C_{stable_i} es el número de modelos farmacóforos derivados de conformaciones estables que se ajustan a la molécula i , $C_{stable\ total}$ es el número total de modelos farmacóforos a partir de conformaciones estables, Ph_{score_i} es el valor de puntuación de fase del conjunto para la molécula i , $Ph_{score\ min}$ es el valor mínimo de puntuación de fase del conjunto entre las moléculas activas conocidas, y $Ph_{score\ max}$ es el valor máximo de puntuación de fase del conjunto entre las moléculas activas conocidas.

25 En la modalidad preferida de realización de la invención, de las librerías de una base de datos de literatura científica se descargaron 4.643 inhibidores conocidos de la enzima AChE y se clasificaron en 70 grupos o familias en función de su estructura molecular. De estas familias, se seleccionaron mediante criterios estadísticos nueve para su posterior análisis. Los ligandos representativos de cada una de las nueve familias se acoplaron al receptor de la AChE empleando el método *Induced-fit docking* (acoplamiento de ajuste inducido)

30

Luego, se realizaron nueve simulaciones MD de 50 ns basadas en las poses acopladas, junto con una simulación adicional del complejo AChE-galantamina. Las conformaciones de la proteína se extrajeron de estas simulaciones y se utilizaron

en un enfoque de acoplamiento conjunto, en el que los compuestos de cada una de las nueve familias se acoplaron por ajuste inducido. Las puntuaciones de acoplamiento resultantes y los valores experimentales de IC50 se utilizaron para generar un ranking de afinidad, identificando los compuestos más activos de cada familia con la nueva métrica YN1.

Estos compuestos activos se utilizaron para generar modelos de aprendizaje automático ML y modelos farmacóforos basados en ligandos. Para el enfoque basado en el receptor, el compuesto más activo de cada familia se utilizó para extraer conformaciones del sitio activo mediante simulaciones MD de 500 ns.

Estas conformaciones del sitio activo se utilizaron para generar modelos farmacóforos basados en complejos, y los modelos farmacóforos se combinaron para crear un modelo de farmacóforos conjuntos.

Los objetos anteriormente descritos, al igual que los objetos adicionales a que hubiere lugar, serán expuestos al detalle y con la suficiencia necesaria en el capítulo descriptivo que se divulga a continuación, el cual constituirá el fundamento del capítulo reivindicatorio.

BREVE DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Representación simplificada del método objeto de la invención con ilustraciones. En una modalidad de realización, se descargaron 4.643 inhibidores de la AChE conocidos y se clasificaron en 70 grupos o familias en función de su estructura molecular. De estas familias, se seleccionaron nueve para su posterior análisis. Los ligandos representativos de cada una de las nueve familias se acoplaron al receptor de la AChE empleando el método Induced-fit *docking*. Se realizaron nueve simulaciones MD de 50 ns basadas en las poses acopladas, junto con una simulación adicional del complejo AChE-galantamina. Las conformaciones de la proteína se extrajeron de estas simulaciones y se utilizaron en un enfoque de acoplamiento conjunto, en el que los compuestos de cada una de las nueve familias se acoplaron por separado. Las puntuaciones de acoplamiento resultantes y los valores experimentales de IC50 se utilizaron para generar un ranking de afinidad, identificando los compuestos más activos de cada familia. Estos compuestos activos se utilizaron para generar modelos de aprendizaje automático y modelos farmacóforos basados en ligandos. Para el enfoque basado en el receptor, el compuesto más activo de cada familia se utilizó para extraer conformaciones del sitio activo mediante simulaciones MD de 500 ns. Estas conformaciones del sitio activo se utilizaron para generar modelos farmacóforos basados en complejos, y los modelos farmacóforos se combinaron para crear un conjunto de modelos farmacóforos.

Figura 2. Representación simplificada de la metodología de cribado virtual (*virtual screening* -VS). Se utilizó una base de datos de moléculas comercial (ZINC2217), que se filtró secuencialmente utilizando modelos de aprendizaje automático, modelos farmacóforos basados en ligandos y modelos farmacóforos de conjunto basados en complejos ligando-receptor. Las moléculas obtenidas son candidatas a inhibidores.

Figura 3. (A) RMSD Media móvil normalizado para los 9 conjuntos de datos (RMSD de átomos pesados para los 6 residuos del sitio activo, RMSD de la columna vertebral para toda la proteína, RMSD de la columna vertebral para los 17 residuos del sitio activo y RMSD de átomos pesados del ligando. (B) Media de RMSD media móvil normalizado para los 9 conjuntos de datos. (C) Derivada numérica de la media móvil normalizada de RMSD. Los gráficos pertenecen a C4.

Figura 4. Tiempos de simulación en los que la media móvil normalizado RMSD y la derivada numérica son superiores a 0,6. Verde (media móvil RMSD < 0,6), azul (media móvil RMSD => 0,6) y rojo (derivada RMSD => 0,6). Los gráficos pertenecen a C31.

Figura 5. Resumen de los conjuntos de modelos de farmacóforos para cada grupo. Los contactos hidrofóbicos se muestran en verde, donantes de enlace H en azul claro, aceptores de enlace H en rojo, interacciones aromáticas en beige y cargados positivamente grupos en azul oscuro. Los residuos cruciales de cada característica se muestran junto a la característica correspondiente. S indica conformación estable y U indica conformación inestable. Los porcentajes mostrados corresponden al número de veces que aparece una característica en el total de conformaciones estables e inestables. TP: verdadero positivo, TN: verdadero negativo, FP: falso positivo, FN: falso negativo, PC: precisión, RC: recuperación, SP: especificidad, BACC: precisión equilibrada, ROC-AUC: área bajo ROC curva.

Figura 6. Esquema del procedimiento utilizado para el cribado virtual de las moléculas de las bases de datos y los resultados. Consta de tres etapas: modelos de aprendizaje automático, modelos de farmacóforos basados en ligandos y un conjunto de modelos de farmacóforos basados en complejos. Inicialmente, las moléculas de la base de datos se someten a un análisis utilizando 9 modelos de aprendizaje automático, cada uno diseñado para familias 31, 4, 19, 50, 35, 20, 36, 42 y 23. Posteriormente, se combinan los activos identificados de cada modelo y se eliminan los duplicados. Luego, la librería resultante se somete a una selección utilizando modelos de farmacóforos basados en ligandos diseñados para las familias 31, 4, 19, 50, 35, 20, 36, 42 y 23. Los activos identificados luego proceden a los modelos de conjunto de farmacóforos correspondientes a sus respectivas familias.

Figura 7. Esquema de los mejores modelos de farmacóforos basados en ligandos. Aceptor de enlace H (A) en rojo, donante de enlace H (D) en luz azul, punto positivo (P) en azul oscuro, aromático (R) en naranja e hidrofóbico (H) en verde. TP: verdadero positivo, TN: verdadero negativo, FP: falso positivo, FN: falso negativo, RC: recuperación, SP: especificidad, BACC: precisión equilibrada, ROC-AUC: área bajo la curva ROC.

DESCRIPCIÓN DETALLADA DE LA INVENCION

La presente invención nace como respuesta a la necesidad de desarrollar nuevos métodos de identificación de nuevos inhibidores enzimáticos incluyendo, pero sin limitar de la AChE, con mayor selectividad para minimizar los efectos secundarios, mediante el análisis detallado de la dinámica proteína-ligando farmacoforado.

La descripción de la realización de la presente invención no pretende limitar su alcance, sino servir como un ejemplo particular de la misma. Se espera que una persona versada en la materia comprenda que las modalidades equivalentes no se apartan del espíritu y alcance de la presente invención en su forma más amplia.

En las estructuras de ligando se realizó una minimización de energía electrónica utilizando el método cuántico RHF/6-31G(d). Las geometrías minimizadas se utilizaron para un cálculo de potencial electrostático restringido (RESP) utilizando un esquema Merz-Singh-Kollman (MK). Todos los cálculos mecánicos cuánticos se realizaron en la fase gaseosa, con multiplicidad singlete. Además, se utilizó el script acpype.py para convertir los archivos al formato Gromacs3 Software Gaussian 16.

El cálculo RESP se utilizó para la parametrización del ligando empleando el campo de fuerza GAFF a través de AmberTools22. Los átomos de proteína fueron parametrizados utilizando el campo de fuerza ff14SB y los átomos de ligando con el campo de fuerza GAFF.

Dentro del contexto de la presente descripción, debe tenerse en cuenta que el procedimiento general de simulaciones de dinámica molecular MD fue consistente, con solo cambios realizados en el tiempo de producción. Se extrajeron las posiciones de ligando contenidas en los complejos proteína-ligando.

En las modalidades recomendadas de realización, los complejos proteína-ligando fueron solvatados utilizando una caja cúbica con una distancia de 1,5 nm. Las moléculas de agua fueron parametrizadas utilizando el modelo TIP3P. La carga neta del sistema fue neutralizada utilizando iones de sodio y cloruro con una concentración de NaCl de 0,15 M. El algoritmo integrador utilizado fue leap-frog. La preparación del sistema fue desarrollada en tres pasos principales: minimización de energía, calentamiento y equilibración. La minimización de energía se realizó en dos

etapas: i) minimización de átomos de solvente, iones y átomos de hidrógeno, y ii) minimización de todos los átomos. Para la minimización, se utilizó el algoritmo de descenso más empinado con un paso de 0,01 nm y restricción con una constante de fuerza de 10 kcal/molÅ²

5 El calentamiento se realizó en tres etapas:

i) NVT de -273 a -173 C que restringe los átomos pesados de la cadena principal y el ligando con una constante de fuerza de 10 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 63 ps.

10 ii) NVT de -173 C a -73 C que restringe los átomos pesados de la cadena principal y el ligando con una constante de fuerza de 10 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 63 ps.

iii) NVT de -73 a 37 C que restringe los átomos pesados de la cadena principal y el ligando con una constante de fuerza de 5 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 63 ps.

15 Las velocidades atómicas aleatorias iniciales se obtuvieron utilizando la distribución de Maxwell-Boltzmann con una temperatura de 37 C. El sistema fue equilibrado en seis pasos:

i) NPT a 37 C con estructura de contención y ligando con una constante de fuerza de 5 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 2 ns.

20 ii) NPT a 37 C con estructura de contención con una constante de fuerza de 3 kcal/molÅ y ligando con 5 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 2 ns.

25 iii) NPT a 37 C con estructura de contención con una constante de fuerza de 1 kcal/molÅ y ligando con 5 kcal/molÅ . Paso de tiempo de 1 fs y un tiempo de simulación de 2 ns. kcal/molÅ .

iv) NPT a 37 C restricción de la cadena principal con una constante de fuerza de 1 kcal/molÅ² y ligando con 3 kcal/molÅ². Paso de tiempo de 1 fs y un tiempo de simulación de 2 ns.

30 v) NPT a 37 C con una estructura principal de contención con una constante de fuerza de 1 kcal/molÅ² y ligando con 1 kcal/molÅ². Paso de tiempo de 1 fs y un tiempo de simulación de 2 ns.

vi) NPT a 37 C con una estructura principal de contención y ligando con una constante de fuerza de 0,1 kcal/molÅ². Paso de tiempo de 2 fs y un tiempo de simulación de 20 ns.

Las interacciones electrostáticas se modelaron con el método *Particle Mesh Ewald* (PME) utilizando un espacio de Fourier de minimización de 0,125 con un paso de 0,01 nm. Finalmente, la corrida de producción de los sistemas fue realizada utilizando el termostato modificado de Berendsen y el barostato modificado de Parrinello-Rahman. Las simulaciones descritas fueron corridas en el cluster HYPATIA de la Universidad de Los Andes y la estación de trabajo TITAN.

El modelo farmacóforo conjunto del método de la presente invención captura las interacciones proteína-ligando, incluyendo en una modalidad de realización el π -catión con Trp-86 y varias interacciones π - π con los residuos Tyr-341, Tyr-337, Tyr-124 y Tyr-72, incluyendo su validación mediante el logro de métricas de alto rendimiento en modelos predictivos.

El método identificó 18 moléculas nuevas de la base de datos ZINC con valores de energía de enlace que oscilan entre -62 a -115 kJ/mol, lo que indica un gran potencial como inhibidores enzimáticos, particularmente, de la AChE, sin que dicha modalidad de realización configure limitación para su aplicación a enzimas diana en general, logrando el desarrollo de nuevos inhibidores.

En una modalidad de realización, los inhibidores de (AChE) se descargaron y clasificaron en 70 grupos o familias en función de su estructura molecular. De estas familias, se seleccionaron nueve para un análisis más detallado. Los ligandos representativos de cada una de las nueve familias se acoplaron al receptor AChE empleando acoplamiento de ajuste inducido.

Se realizaron nueve simulaciones MD de 50 ns basadas en las posturas acopladas, junto con una simulación adicional del complejo AChE-galantamina.

La proteína se extrajo de estas simulaciones y se utilizó en un enfoque de acoplamiento de conjuntos, en el que los compuestos de cada una de las nueve familias se acoplaron a la proteína. Las puntuaciones de acoplamiento resultantes y los valores experimentales de IC50 se utilizaron para generar un ranking de afinidad, identificando los compuestos más activos de cada familia.

Los compuestos se utilizaron para generar modelos de aprendizaje automatizado (Machine Learning) y modelos de farmacóforos basados en ligandos. Para el enfoque basado en receptores, se utilizó el compuesto más activo de cada familia para extraer las conformaciones de la sustancia activa a través de simulaciones MD de 500 ns.

Los modelos de farmacóforos basados en ligandos y los modelos de farmacóforos basados en complejos ligando-receptor se combinan para crear un modelado de farmacóforos conjunto.

La figura 1 muestra una representación ilustrativa del método de cribado virtual. Se utilizó una base de datos de literatura científica (binding data base) empleo secuencialmente para generar los modelos de aprendizaje automático, modelo de farmacóforo basado en ligandos, modelamiento basados en complejos ligando-receptor y la combinación de estos en modelos de farmacóforos conjuntos para, finalmente, obtener un modelo robusto que permita identificar las moléculas candidatas a ser inhibidoras.

El modelamiento de farmacóforos por conjuntos aportado por el método objeto de la presente descripción integra múltiples modelos de farmacóforos basados en complejos, cada uno de los cuales representa un marco específico de trayectoria dinámica, para superar las limitaciones del modelo individual, teniendo en cuenta el sitio activo con flexibilidad, generando un efecto técnico sorprendente en el cribado virtual para el desarrollo de nuevos tratamientos de enfermedades, particularmente, de la enfermedad de Alzheimer, siendo susceptible de aplicación a otras enzimas, allanando el camino para la identificación de nuevos candidatos a fármacos contra diversas enfermedades.

En una modalidad de realización, la presente invención desarrolla e implementa un método dirigido a identificar nuevos inhibidores de la acetilcolinesterasa al dilucidar las características moleculares esenciales dirigiéndose selectivamente a la enzima acetilcolinesterasa humana (huAChE).

Este método involucra un extenso protocolo in-silico que abarca la base de datos de gestión, agrupamiento de ligandos, cálculos RMSD, acoplamiento de ajuste inducido, simulaciones de dinámica molecular, análisis fisicoquímicos, acoplamiento de conjuntos, modelado de farmacóforos y técnicas de aprendizaje de máquina.

El método de la invención incluye la búsqueda de la diversidad estructural de nuevos inhibidores, la exploración del espacio conformacional de los receptores, la amalgama de resultados experimentales, la integración de modelos de aprendizaje automático en la búsqueda de nuevos inhibidores, y la curación selectiva de inhibidores para pruebas experimentales. El método integra diferentes modelos de estructuras biológicas de ligandos diana (complejos) y utiliza modelos extraídos de la dinámica de conformaciones del complejo. Esto dota al método objeto de descripción de la capacidad para dar cuenta de la dinámica intrínseca de los sistemas biológicos.

En una modalidad de realización, el método logra la obtención de nueve familias distintas de inhibidores, cada una asociada con un modelo predictivo de aprendizaje automático, un modelo de farmacóforo basado en ligandos y un modelamiento conjunto de farmacóforo derivado de complejos ligando-receptor. Estos modelos facilitaron el cribado virtual contra una base de datos comercial (ZINC22), lo que resultó en la identificación de 18 moléculas con potencial inhibitorio contra la AChE.

Búsqueda de diversidad estructural

En este punto lo tradicional es centrarse exclusivamente en una familia de compuestos o librería química en la cual se realiza modificaciones pequeñas en el andamio de la molécula (mínima estructura común a toda la librería) para posteriormente mediante un modelo de relación estructura actividad cuantitativo (QSAR por sus siglas en inglés) identificar las modificaciones más prometedoras. Estas modificaciones estructurales en las moléculas a su vez se convertirán en un nuevo punto de partida para generar nuevas librerías. Iterando una vez más esta nueva familia de moléculas con lo cual se busca lograr una mejora sustancial en los inhibidores. Este proceso se repite hasta lograr convergencia en los resultados.

El enfoque aportado por el método objeto de descripción es radicalmente diferente, pues, primero, no requiere la iteración experimental y, segundo, las moléculas se identifican de un agrupamiento y selección de familias mediante similitud estructural y métricas estadísticas tomando la información inicial de moléculas reportadas en bases de datos.

Agrupamiento por similitud de las estructuras de los inhibidores.

Inhibidores con un $IC_{50} 1 \leq 99000$ nM contra el huAChE se extrajeron de la Base de datos (BD) (<https://www.bindingdb.org/>). Estos inhibidores fueron ingresados en la herramienta LigPrep de la suite Schrödinger 2023.1 (Schrödinger, LCC) para generar 3D estructuras con un pH de $7,4 \pm 0,229$.

Se lleva a cabo un agrupamiento de similitud estructural de las moléculas del BD utilizando el Canvas dentro de la suite Schrödinger 2023.1 (Schrödinger, LCC). El análisis de agrupamiento fue ejecutado empleando configuraciones específicas que incluyen precisión de 64 bits, tipo de huella dactilar radial, el criterio de similitud y el método de vinculación promedio. Posteriormente, utilizando los resultados de Canvas y el valor de penalización de Kelley, se determina el número óptimo de clústeres, equilibrando entre el exceso de agrupamiento y la penalidad del número de clusters para garantizar la importancia de los clústeres resultantes.

Se obtuvieron un total de 70 grupos de ligandos del BD. Para cada grupo, se calculan métricas estadísticas que incluyen el promedio IC_{50} , la desviación estándar de IC_{50} , el coeficiente de variación de IC_{50} y el tamaño del clúster. Estos parámetros estadísticos, junto con otros criterios de selección, se utilizaron para seleccionar 9 grupos representativos para un análisis más detallado.

Acoplamiento de ajuste inducido de moléculas representativas de grupos de inhibidores

Para generar la rejilla de deslizamiento para la caja de acoplamiento de ajuste inducido, los 17 residuos resultantes de los cálculos emparejados de RMSD fueron

empleados. Los ficheros PDB correspondientes al código UNIPROT P22303 de huAChE fueron descargados, lo que da como resultado un total de 61 estructuras. Sin embargo, las estructuras que no eran de tipo nativo ~~salvaje~~, carecían de complejos proteína-ligando con ligandos no covalentes, o tenían una resolución superior a 3,2 Å fueron excluidos. Este proceso de filtrado dejó un total de 18 estructuras con los siguientes códigos PDB: 4BDT, 4EY5, 4EY6, 4EY7, 4M0E, 4M0F, 5FOQ, 6CQU, 6F25, 6NEA, 6O4W, 6O4X, 6O5R, 6O5V, 6O50, 6O52, 6U3P y 6U34.

Empleando los residuos del sitio activo, los cálculos pareados de RMSD facilitaron la identificación de los residuos clave que componen el sitio activo, entre ellos Tyr-72, Asp-74, Trp-86, Gly-120, Gly-121, Tyr-124, Tyr-133, Glu-202, Ser-203, Trp-286, Phe-295, Phe-297, Tyr-337, Phe-338, Tyr-341, His-447, y Gly-448.

La validación del acoplamiento del ajuste inducido se llevó a cabo mediante cálculos para el ligando de proteína huAChE (sistema que involucra moléculas de acetilcolina y galantamina). Se empleó el cristal de proteína con el código PDB 4EY6 y abordó los residuos faltantes utilizando el módulo Prime dentro de la suite Schrödinger 2023.1 (Schrödinger, LCC).

Los estados de protonación se asignaron mediante PROPKA a un pH de 7.4 . Se llevó a cabo una minimización energética de las estructuras cristalinas, restringiendo el RMSD máximo de átomos pesados a 0,18 Å. Se aplicaron protocolos estándar con configuraciones específicas: re-acoplamiento dentro de 30 kcal/mol y 20 estructuras en total, empleando precisión XP, utilizando el campo de fuerza OPLS3, ajustando la escala del receptor van der Waals a 0,5, el ligando van der Waals escalando a 0,5, refinando los residuos dentro de 5 Å de las posiciones del ligando, lo que permite la optimización del lado cadenas, estableciendo el número máximo de poses en 20, y muestreando conformaciones de anillos dentro de una Ventana energética de 2,5 kcal/mol. Tras la selección de moléculas representativas, basada en la estructura molecular (centroide de grupo), de las nueve familias de conglomerados.

Simulación de complejos proteína-ligando de moléculas representativas de grupos de inhibidores de la acetilcolinesterasa humana y muestreo.

Las posturas correspondientes a las 9 estructuras moleculares representativas asociadas a las familias C31, C4, C19, C50, C35, C20, C36, C42 y C23 (C significa cluster y el número identifica el mismo) se extrajeron de los cálculos de ajuste inducido.

Además, también se extrajo la pose experimental del ligando galantamina del cristal 4EY6. Estas 10 complejos se utilizaron para llevar a cabo simulaciones MD con una duración de fase de producción de 50 ns.

Las 10 trayectorias obtenidas de las simulaciones de dinámica molecular de 50 ns se utilizaron para extraer conformaciones del sitio activo. El valor RMSD de la espina dorsal de la enzima (backbone) se monitoreó para el sitio activo y Trp-236, Glu-334, Ala-204, Gly-122 a lo largo de toda la trayectoria.

- 5 Los valores RMSD de: promedio xRMSD, promedio más y menos el doble de la desviación estándar $xRMSD \pm 2\delta$ se calcularon. Este proceso arrojó un total de 7 conformaciones por trayectoria.

Las 7 conformaciones se analizaron utilizando los cálculos pareados RMSD. En los cálculos, se seleccionaron 5 conformaciones de las 7 conformaciones originales
10 para cada trayectoria. En total, se extrajeron 50 conformaciones clave para su posterior análisis.

Exploración conformacional del receptor

Mediante una combinación de *docking* y dinámicas moleculares con cada representante de las familias previamente mencionadas, se exploran diferentes
15 conformaciones estructurales del receptor (diana biológica). Esto permite tener mejores estructuras conformacionales sobre la cuales buscar los mejores inhibidores para estas. Usualmente esta búsqueda se lleva a cabo empleando la información cristalográfica de las dianas biológicas correspondientes o empleando algunas dinámicas con un par de los ligandos más potentes.

- 20 Este enfoque innovador integra múltiples modelos farmacóforos basados en complejos, cada uno de los cuales representa un marco específico de una trayectoria de dinámica molecular, para superar las limitaciones de los modelos individuales y tener en cuenta la flexibilidad del sitio activo.

Combinación de resultados computacionales y experimentales

- 25 En la gran mayoría de investigaciones se incluyen únicamente los resultados computacionales, ocasionalmente se contrastan con datos experimentales, estos datos experimentales provienen mayoritariamente del grupo de investigación y/o colaboraciones particulares. La razón de considerar solo estos datos experimentales es evitar la gran variación de estos, cuando se consideran los
30 valores reportados en bases de datos.

Las 50 conformaciones de receptor se utilizaron como entrada para un acoplamiento de conjunto. Se utilizó el módulo Epik para indicar. Los estados de protonación de la enzima y el ligando.

- Los valores de XP GScore obtenidos a partir del acoplamiento del conjunto y los
35 valores experimentales de inhibición (IC₅₀) de los compuestos en grupos clave se utilizaron para diseñar una puntuación de afinidad para clasificar los compuestos en

cada grupo, según su inhibición experimental y mediante predicción por computadora.

El método de cribado virtual objeto de descripción comprende una métrica Yasser número uno (YN1). Empleando esta métrica se logra integrar los datos experimentales reportados en literatura y los resultados computacionales obtenidos. Esto, a su vez, permite tener una mejor certeza en la selección de moléculas que son usualmente empleadas para modelos de búsqueda de nuevos inhibidores. A continuación, se detalla la métrica:

$$LE_i = -\frac{XP\ GScore_i}{MW_i} \quad i = 1, 2, \dots, N$$

$$YN = \frac{LE_i - LE_{min}}{LE_{max} - LE_{min}} + \left(1 - \frac{\log_{10} IC50_i - \log_{10} IC50_{min}}{\log_{10} IC50_{max} - \log_{10} IC50_{min}}\right)$$

Donde LE_i es la eficiencia del ligando, MW_i es el peso molecular, $\log_{10} IC50_i$ es el log base 10 de IC50, max es el valor máximo dentro del *clúster*, min es el valor mínimo dentro del *clúster*, y N es el número de moléculas del *clúster*.

En YN hay dos términos, el primero se refiere al valor normalizado de la eficacia del ligando, y el segundo al valor normalizado de $\log_{10} IC50$. El segundo término calcula el valor normalizado invertido de $\log_{10} IC50$, donde el valor normalizado se resta de 1 para invertir la escala. Esto se debe a que un valor más cercano a 0 indica un compuesto más activo, mientras que un valor más cercano a 1 indica un compuesto menos activo. La puntuación oscila entre 0 y 2, y una puntuación más cercana a 2 indica una mayor capacidad inhibidora.

Todas las moléculas con un valor de XP GScore del acoplamiento de conjuntos se consideraron moléculas activas en esta etapa. Así, se extrajeron moléculas activas de los grupos 31, 4, 19, 50, 35, 20, 36, 42 y 23. Estas moléculas activas se dividieron aleatoriamente en dos grupos: 20% para las pruebas y el 80% para entrenamiento. Los conjuntos de entrenamiento y prueba de moléculas activas se utilizaron por separado para, aproximadamente, 50 señuelos por molécula activa. En consecuencia, se obtuvieron dos bibliotecas, cada una con aproximadamente 50 señuelos por cada molécula activa.

Entrenamiento de modelos de aprendizaje automático

Se tomaron los activos y señuelos del conjunto de datos de entrenamiento balanceado, los descriptores moleculares se calcularon mediante RDKit 3.4 y se realizó un análisis de correlación. Se determinaron los hiperparámetros, mediante entrenamiento de algoritmos utilizando el conjunto de datos.

Prueba de modelos de aprendizaje automático

Se tomaron los activos y señuelos del conjunto de datos de prueba desequilibrados, y se descriptores moleculares se calcularon utilizando RDKit. Se eliminaron las variables correlacionadas encontradas durante el entrenamiento. A continuación, se probaron los algoritmos de clasificación con las moléculas de prueba.

En la generación de un modelo de farmacóforo basado en ligandos, tanto las moléculas activas como los señuelos son componentes esenciales. Las moléculas activas, identificadas por una puntuación YN1 igual o superior a 1, fueron seleccionadas. Los señuelos correspondientes a las moléculas activas identificadas se generaron utilizando el servidor DUDE (Base de datos de señuelos útiles: mejorado), lo que resultó en la producción de 50 señuelos por compuesto activo.

El modelado de farmacóforos basado en ligandos se configuró para identificar la mejor alineación y las características más comunes. Se estableció un objetivo de 50 conformadores. El porcentaje de umbral de activos necesarios para coincidir con una hipótesis se estableció en el 50%. La función de puntuación empleada fue el Phase Hypo Score. Los cálculos se realizaron utilizando el módulo Phase de la suite Schrödinger 2023.1 (Schrödinger, LCC).

Modelado de farmacóforos basado en complejos de conjuntos.

Para cada una de las nueve trayectorias de dinámica molecular (MD), que abarcan 500 nanosegundos se identificaron las conformaciones estables e inestables del sitio activo. Posteriormente se utilizaron conformaciones para construir modelos de farmacóforos basados en complejos, empleando la herramienta 'Develop Pharmacophore Model' dentro del módulo Schrödinger Phase. Inicialmente se aplicaron las configuraciones predeterminadas, se realizaron ajustes en el número máximo de funciones y características obligatorias para refinar y optimizar la zona de características de funcionamiento del receptor (ROC-AUC).

Agrupamiento conformacional de las posturas de los ligandos a partir del acoplamiento de conjuntos

Se seleccionaron los compuestos con mayor puntuación de YN1 para cada familia, totalizando 9 compuestos. Las poses de estos compuestos seleccionados, adquiridas a través del acoplamiento de conjuntos, fueron empleados en el agrupamiento conformacional utilizando la herramienta Conformer Cluster de Schrödinger 2023.1. El agrupamiento utilizó el método de vinculación promedio, RMSD atómico y regiones de comparación especificadas involucrando átomos pesados + OH, SH. Se identificó la postura observada con mayor frecuencia entre las varias posturas obtenidas del acoplamiento del conjunto para cada uno de los 9 ligandos. Se utilizaron estas 9 posturas para llevar a cabo simulaciones MD con una duración de fase de producción de 500 ns.

Cálculos TRAPP y extracción de conformaciones de emplazamientos activos a partir de 500 ns MD

Se calcularon los valores de RMSF para los átomos pesados (columna vertebral y cadena lateral) para los residuos en el sitio activo. Estos valores RMSF se ordenaron de mayor a menor. De cada simulación, se seleccionaron para el análisis los 6 residuos con los valores más altos de RMSF.

Los átomos pesados RMSD para los 6 residuos individuales del sitio activo, la columna vertebral RMSD para toda la proteína, la RMSD de la columna vertebral para los residuos del sitio activo y la RMSD para los átomos pesados del ligando calculado para todas las simulaciones MD. Una media móvil exponencial (EMA) con un periodo de 5000 se calculó para los 9 conjuntos de datos RMSD, cada uno de los cuales consta de 50000 fotogramas.

El período de EMA se alineó con el número de fotogramas, manteniendo una relación de 1:10.

A continuación, se normalizó para garantizar que los valores estuvieran dentro del rango de 0 a 1.

Para lograr esto, el valor mínimo de la variable que se está normalizando se restó de la variable. A continuación, se calculó la diferencia entre los valores mínimo y máximo y se dividió por este último.

La media móvil normalizada se utilizó para calcular la derivada numérica estimando el valor absoluto de la derivada en un punto como la diferencia entre el valor de la función en ese punto y un punto un poco más grande, dividido por la distancia entre los dos puntos. Los gráficos de mapas de calor fueron generados utilizando la media móvil normalizada y la derivada numérica.

Además, los gráficos se generaron utilizando la media móvil normalizada del RMSD y derivada numérica para enfatizar las regiones que exhiben una media móvil y una derivada numérica superior a 0,6. La utilización simultánea de la media móvil y la derivada numérica facilitó la identificación de intervalos de tiempo.

Se utilizó la herramienta de agrupamiento TRAPP-Analysis para extraer fotogramas de rangos de tiempo de conformaciones estables. Por otro lado, se utilizó la herramienta TRAPP-Pocket para extraer fotogramas de conformaciones inestables. Para el análisis TRAPP, el parámetro "cluster" se estableció en 3 Å, "kmeans" se estableció en 3 Å, "zmax" se estableció en 8 Å y "bb" se configuró en "on". Para TRAPP-Pocket, el parámetro "radio" se fijó en 5 Å, los residuos del sitio activo se introdujeron manualmente utilizando los 17 residuos identificados previamente, y los demás ajustes se mantuvieron igual que en el ejemplo de AR-TRAPP que se encuentra en las carpetas de ejemplo de TRAPP.

Los scripts de python TRAPP-pocket se editaron para habilitar el cálculo de 50000 fotogramas. Los fotogramas obtenidos de la agrupación con TRAPP-Análisis de rangos de tiempo estables se llamaron conformaciones estables. Por otro lado, las gráficas de los descriptores de las propiedades del sitio activo se realizaron a partir de los resultados de TRAPP-pocket, y los fotogramas. Se extrajeron los valores atípicos de los descriptores. A estos marcos se les llamó conformaciones inestables.

Modelización de farmacóforos basada en complejos.

Cálculos TRAPP y extracción de conformaciones de emplazamientos activos a partir de 500 ns MD

Para cada una de las 9 trayectorias de dinámica molecular (MD) de 500 ns, los marcos de conformaciones inestables del sitio activo se emplearon para construir modelos de farmacóforos utilizando la herramienta Develop Pharmacophore Model en Schrödinger 2023.1 suite (Schrödinger, LLC) Módulo de fase.

Se utilizaron los ajustes predeterminados y se realizaron variaciones al número máximo de características y características obligatorias dentro de los rangos de 6 a 5 y de 3 a 2, respectivamente. Las hipótesis farmacofóricas se combinaron, formando lo que se denomina dentro del contexto de la presente descripción un conjunto de hipótesis o conjunto de modelos de farmacóforos.

En este contexto, una molécula se considera activa si puede coincidir con al menos, un modelo de farmacóforo del conjunto. Por el contrario, una molécula se considera inactiva si no puede coincidir con ningún modelo de farmacóforo del conjunto. Dentro del alcance de la presente descripción se define el término "puntuación de fase de conjunto" como la puntuación de fase más alta obtenida por molécula entre todas las puntuaciones de fase obtenidas para cada modelo de farmacóforo en el conjunto. Si una molécula no coincide con ningún modelo, tendrá una puntuación de cero. Para ejecutar este proceso, se desarrolló un conjunto de scripts de Python.

Para validar los 9 conjuntos, se extrajeron activos y señuelos de cada clúster. Compuestos con una puntuación YN1 mayor o igual a 1 se consideraron activos para esta etapa. Los señuelos fueron obtenidos del servidor DUDE (Database of Useful Decoys: Enhanced)(Mysinger et al., 2012) , generando 50 señuelos por activo. Se generaron curvas ROC, BACC, recuerdo y valores de especificidad para cada validación utilizando la puntuación de fase de conjunto y la herramienta de cálculo de enriquecimiento de la suite Schrödinger 2023.1.

A continuación, se selecciona la configuración del número máximo de características y las características obligatorias que exhibió el valor más alto de ROC-AUC para el conjunto.

Validación de conjuntos de hipótesis.

El método objeto de la presente descripción metodológica aprovecha la información recolectada, desarrollando una nueva métrica que se denomina *Yasser número* dos (YN2) que se describe a continuación, la cual, combina los resultados generales de
5 de la suite Schrödinger 2023.1. con datos propios de la dinámica del sistema biológico. Empleando este enfoque, se produce una reducción significativa de tiempos y costos computacionales.

$$\hat{C}_{unstable_i} = \frac{C_{unstable_i}}{C_{unstable\ total}} \quad \hat{C}_{stable_i} = \frac{C_{stable_i}}{C_{stable\ total}} \quad i = 1, 2, \dots, N$$

$$if\ Ph_{score_i} \geq Ph_{score\ min} \quad then \quad \widehat{Ph}_{score} = \frac{Ph_{score_i} - Ph_{score\ min}}{Ph_{score\ max} - Ph_{score\ min}} \quad and$$

$$10 \quad YN2_i = \frac{(\hat{C}_{unstable}) + (\hat{C}_{stable}) + (\widehat{Ph}_{score})}{3}$$

$$if\ Ph_{score_i} < Ph_{score\ min} \quad then \quad \widehat{Ph}_{score} = 0 \quad and \quad YN2_i = 0$$

Donde N es el número de moléculas en el procedimiento de validación o cribado virtual, $C_{unstable_i}$ es el número de modelos farmacóforos derivados de conformaciones inestables que se ajustan a la molécula i , $C_{unstable\ total}$ es el número
15 total de modelos farmacóforos derivados de conformaciones inestables, C_{stable_i} es el número de modelos farmacóforos derivados de conformaciones estables que se ajustan a la molécula i , $C_{stable\ total}$ es el número total de modelos farmacóforos a partir de conformaciones estables, Ph_{score_i} es el valor de puntuación de fase del conjunto para la molécula i , $Ph_{score\ min}$ es el valor mínimo de puntuación de fase del
20 conjunto entre las moléculas activas conocidas, y $Ph_{score\ max}$ es el valor máximo de puntuación de fase del conjunto entre las moléculas activas conocidas.

YN2 combina los valores de la puntuación de fase del conjunto, el número de conformaciones estables coincidentes, y el número de conformaciones inestables coincidentes. Las moléculas con pocas coincidencias con los modelos de
25 farmacóforo son penalizadas.

En una modalidad de realización YN2 se emplea en otra ronda de validación. Inicialmente, el cálculo de YN2 se realizó para todas las moléculas. Para que una molécula sea considerada activa, debe tener un valor mayor o igual a un umbral YN2; Por el contrario, para que una molécula se considere inactiva, debe tener un
30 valor inferior que un umbral YN2. Para determinar el umbral YN2, se genera una función objetivo que es la suma de recall y especificidad. Se optimiza la función objetivo, obteniendo el valor YN2 que produjo el valor máximo para la función objetivo.

Por lo tanto, de cada validación de conjunto, se extrae el umbral YN2, el valor de Puntuación de Phscore máximo y Puntuación de Phscore min. Estos tres valores, junto con el número de conformaciones estables e inestables, sirven como parámetros para una posterior ejecución de cribado virtual.

5 Cribado virtual de la base de datos.

De la base de datos comercial (ZINC), se descargaron en stock las moléculas con un peso molecular igual o inferior a 500 Da, lo que resultó en la recuperación de 11.012.710 moléculas.

Las moléculas se sometieron a un protocolo de detección virtual que consta de tres etapas: modelos de aprendizaje de máquina, modelos de farmacóforos basados en ligando- y un conjunto de modelos de farmacóforos basados en complejos.

Inicialmente, las moléculas de la base de datos ZINC se sometieron a un cribado utilizando 9 modelos de aprendizaje automático, cada uno diseñado para las Familias 31, 4, 19, 50, 35, 20, 36, 42 y 23. Posteriormente, se identificaron los activos de cada modelo, se combinaron y se eliminaron los duplicados.

A continuación, la biblioteca resultante se sometió a un cribado, empleando modelos de farmacóforos basados en ligandos diseñados para grupos clave. Los activos identificados procedieron a filtrar con los modelos de conjunto de farmacóforos correspondientes a sus respectivas familias. Finalmente, se seleccionaron los 500 ligandos con los YN2 de cada clúster y se sometieron a un procedimiento de acoplamiento conjunto, incorporando las funciones de clasificación HTVS, SP, XP y MMGBSA.

Combinación de múltiples modelos conformacionales del receptor

El método objeto de divulgación no solamente combina diferentes modelos de estructuras diana biológica - ligando (complejo), sino que también emplea modelos que se extraen de las dinámicas en conformaciones de alta labilidad del complejo, brindándole la capacidad de considerar la dinámica propia de los sistemas biológicos. Lo anterior permite considerar estructuras novedosas del complejo que puedan tener una muy buena interacción y que de otra forma es poco probable que sean identificadas. Adicionalmente, el uso del software TRAPP en dos formas diferentes, primero para la generalización de zonas estables en las cuáles no hay mayor variación estructural y segundo para la identificación de zonas en las cuales se aprecian cambios estructurales significativos, permite arribar a la selección de zonas con un elevado interés farmacológico.

Para la selección de estructuras se empleó el criterio desviación cuadrática media (RMSD por sus siglas en ingles). En el contexto de la presente invención, la referencia al cálculo de RMSD pareada debe entenderse como un cálculo que

calcula el RMSD correspondiente a un residuo de interés para varios pares de cristales. El resultado es un mapa de calor en el que los cristales están representados en el eje x y el eje y, y sus respectivos valores de RMSD se representan en el mapa de calor.

- 5 Para realizar un cálculo pareado de RMSD, inicialmente se buscan cristales de proteína que cumplan los siguientes criterios: i) resolución de menos de 3,2 Å, ii) enzima de tipo salvaje, y iii) que contuvieran ligandos co-cristalizados no covalentes.

Se seleccionaron un total de 18 cristales de proteína huAChE, identificados específicamente por sus códigos PDB: 4BDT, 4EY5, 4EY6, 4EY7, 4M0E, 4M0F,
10 5FOQ, 6CQU, 6F25, 6NEA, 6O4W, 6O4X, 6O5R, 6O5V, 6O50, 6O52, 6U3P y 6U34.

Para realizar un cálculo de RMSD pareado, se establecen tres etapas: (i) alinear todos los cristales de proteína en relación con un cristal de referencia, (ii) extraer residuos ubicados a 5 Å del centro de masa del ligando para cada estructura cristalina y obtener residuos con una ocupación superior al 60%, y (iii) calcular el
15 RMSD para residuos específicos (ocupación superior al 60%) encontrados en el segundo paso para cada combinación de pares de cristales de proteína.

El número de combinaciones posibles se determinó utilizando la fórmula $C(n, 2) = n! / (2! * (n-2)!)$, donde n es el número de cristales. Para facilitar este proceso, desarrollamos cuatro scripts de Python. Se empleó el módulo Pymol para Python y
20 el script Pymol RmsdByResidue4 para realizar los cálculos.

Para la primera etapa, los 18 cristales de proteína se alinearon con el cristal PDB 4BDT (seleccionado al azar) como estructura de referencia. En cuanto a la segunda etapa, para comparar los residuos ubicados a 5 Å del ligando, utilizamos el término "ocupación", que representa el porcentaje de veces que se encuentra un residuo a
25 5 Å del ligando en todos los cristales de proteína. Identificamos los residuos ubicados a 5 Å de los ligandos para los 18 cristales y luego se calculó la ocupación para cada residuo. Se encontraron 17 residuos con una ocupación del 60% o superior: Tyr-72, Asp-74, Trp-86, Gly-120, Gly-121, Tyr-124, Tyr-133, Glu-202, Ser-203, Trp-286, Phe-295, Phe-297, Tyr-337, Phe-338, Tyr-341, His-447 y Gly-448.
30 Adicionalmente, se identificaron 4 residuos con una ocupación del 100%: Asp-74, Tyr-124, Tyr-337 y Tyr-341.

Los 17 residuos identificados en la segunda etapa se emplearon para definir los residuos que constituyen el sitio activo. Por otra parte, para la tercera etapa, se emplearon los mismos 17 residuos para calcular los mapas de calor RMSD.

35 **Ranking de afinidad basado en el dinamismo del sistema biológico y experimental de datos.**

Dentro de la exploración de posibles compuestos que puedan inhibir huAChE, el método objeto de la presente descripción obtiene la recuperación 4643 inhibidores de la BD. Para examinar la composición química de estos inhibidores y descubrir posibles patrones moleculares, se realiza un análisis de agrupamiento de sus estructuras moleculares.

El análisis permitió clasificar los inhibidores en grupos separados en función de su valor molecular.

El análisis de conglomerados (clusters) indicó que el número óptimo de conglomerados era de 383, acompañados de por una penalización de Kelley de 1372. La penalización de Kelley sirve como métrica para evaluar el rendimiento de la agrupación, donde un valor más bajo significa un valor más favorable. También se calcula los valores de penalización de Kelley para tamaños de clúster de 50, 70 y 100, lo que arrojó 1780, 1570 y 1540 de penalización de Kelley, respectivamente. Para lograr un equilibrio entre la eficiencia computacional y fiabilidad, se eligió proceder con 70 clústeres. En consecuencia, los 4643 inhibidores fueron agrupados en 70 grupos distintos, sinónimos de 70 familias de compuestos, en función de su estructura.

Posteriormente, se disminuyó el número de clústeres de 70 por dos razones cruciales: en primer lugar, para reducir la carga de trabajo computacional y, en segundo lugar, para centrarse en familias con una mayor concentración de moléculas. Para lograr este objetivo, aprovechando el hecho de que cada molécula procedente de la BD posee un valor IC₅₀ indicativo de su capacidad inhibitoria, se calculó el coeficiente IC₅₀ promedio, desviación estándar de IC₅₀ de variación de IC₅₀ y tamaño de conglomerado para cada familia. Además, se asignó una ID de clúster para Facilitar la identificación de cada familia.

A continuación, se identificaron los conglomerados dentro del conjunto inicial de 70 que presentaban los niveles más bajos de valores medios de IC₅₀ (que indican un alto potencial inhibitorio), desviación estándar mínima de IC₅₀ (lo que sugiere una menor variabilidad en el IC₅₀ dentro del conglomerado), un coeficiente de variación modesto (lo que significa que datos más uniformes y consistentes) y una gran población (que ofrece un conjunto más grande de compuestos para su examen).

Para lograr este objetivo, se establece un umbral basado en los valores medios de Promedio de IC₅₀, desviación estándar de IC₅₀, coeficiente de variación y población calculados en todos los Familias. Los valores de corte determinados fueron un promedio de IC₅₀ de 16268 nM, un promedio de IC₅₀ una desviación de 14901 nM, un coeficiente de variación promedio de 1.05 y una población promedio de 66 Compuestos.

A continuación, se emplearon estos umbrales como base para crear filtros de clúster, refinándolos manualmente para maximizar la población de clústeres. Los criterios establecidos fueron: (i) promedio de IC50 por debajo de 10000 nM, ii) desviación típica IC50 inferior a 14900 nM, iii) población superior a 58 moléculas, y
5 (iv) coeficiente de variación menor a 1.7.

Un total de 17 clústeres cumplieron las condiciones (41, 1, 25, 51, 22, 56, 60, 46, 36, 31, 4, 19, 50, 35, 20, 42, 23). Se optó por mantener 3 clústeres que cumplieron con las 4 condiciones, junto con los 6 conglomerados que cumplieron las condiciones y tuvieron las mayores Poblaciones.

- 10 Además, se recuperaron los compuestos asociados a la centroide de cada clúster, que teóricamente debería representar un valor molecular medio estructura para cada familia respectiva.

Se requirieron las poses iniciales de las 9 moléculas de centroide en el sitio activo de AChE. Para ello, se empleó el método de acoplamiento de ajuste inducido. Para
15 validar la eficacia del ajuste inducido acoplamiento, se realizaron cálculos de acoplamiento de ajuste inducido para el ligando de proteína huAChE (PDB 4EY6) utilizando moléculas de acetilcolina y galantamina.

El acoplamiento de ajuste inducido reprodujo con éxito la pose auténtica de la galantamina, lo que produce un valor RMSD de 0,93 entre las poses predichas y las
20 reales. En el caso de la acetilcolina, el acoplamiento de ajuste inducido replicó con precisión la postura documentada de la literatura. Después de la validación, las estructuras de los centroides se sometieron con éxito a un acoplamiento inducido y el ligando resultante. Se utilizaron poses para realizar 9 simulaciones MD de 50 ns. Además, se realizó una simulación MD de 50 ns para el cristal 4EY6, que contiene
25 el ligando de galantamina co-cristalizado.

La energía, el volumen, la temperatura y la presión se crearon durante la fase de producción de MD para confirmar el correcto funcionamiento de los 10 sistemas simulados. Los gráficos cinéticos y de energía total demostraron fluctuación
30 alrededor de un valor estable, mientras que los gráficos de volumen exhibieron valores constantes.

El valor general de RMSD de la red troncal se monitoreó para 21 residuos de sitios activos a lo largo de toda la trayectoria. Estos incluyeron 17 residuos (bolsillo para llaves) identificados en el RMSD cálculos pareados y 4 residuos elegidos por su
35 papel crucial en el comportamiento del subsitio (Trp-236, Glu-334, Ala-204 y Gly-122)¹⁶. Los valores RMSD obtenidos se utilizaron para extraer diversas conformaciones del sitio activo de la proteína AChE.

Para ello, se realizó un análisis estadístico de los datos de la RMSD mediante el cálculo de $\bar{x}$ RMSD, x RMSD $\pm 2\delta$, IRMSD, $I + 1$ RMSD, hRMSD y $h - 1$ RMSD. El

proceso produjo un total de 7 conformaciones extraídas de cada simulación MD, que luego se repetido para los 10 MD, lo que da como resultado un total de 70 conformaciones. Las conformaciones obtenidas del paso anterior se trataron como si fueran cristales de proteínas diferentes y fueron sometida a los cálculos emparejados de RMSD. Sin embargo, los pasos 1 y 2 no se realizaron. Para el paso 3, los cálculos pareados de RMSD se llevaron a cabo utilizando todos los átomos de 21 residuos mencionados anteriormente.

Cinco de los 21 residuos que mostraron el promedio más alto de RMSD pareado fueron seleccionados. Mapas de calor RMSD se utilizaron para extraer 5 conformaciones de las 7 conformaciones iniciales para lograr la mayor diversidad conformacional. Comparando los resultados, se observó que los residuos Asp-74, Glu-334, Tyr-341, His-447, Phe-295 y Phe-338 mostraron un movimiento significativo en común entre las 7 conformaciones, revelando cambios tanto en la columna vertebral como en la cadena lateral.

Finalmente, se obtuvieron e implementaron 50 conformaciones en el procedimiento de acoplamiento de conjuntos. El proceso de acoplamiento del conjunto utilizó 50 conformaciones de sitio activo obtenidas a partir de simulaciones 50 ns MD como entrada, junto con todos los compuestos de las familias (grupos clave).

El resultado fue una clasificación de afinidad basada en las puntuaciones de acoplamiento Glide XP GScore, que abarca todas las conformaciones que emparejó cada ligando e identificó la conformación que mejor se adaptaba a cada ligando.

Posteriormente, se desarrolló la métrica (YN1) para clasificar los compuestos en función de sus puntuaciones de acoplamiento y capacidad de inhibición experimental (IC50). YN1 se compone de dos términos: El primer término representa el valor normalizado de la eficiencia del ligando, mientras que el segundo término implica el valor normalizado invertido de logIC50. El valor logIC50 invertido se calcula restando el valor normalizado de 1 para invertir la escala, donde un valor más cercano a 0 indica una mayor actividad, y más cerca de 1 indica menor actividad. La puntuación YN1 va de 0 a 2, donde una puntuación cercana a 2 significa un mayor potencial inhibitorio computacional y experimental.

Todos los ligandos de cada familia se clasificaron utilizando YN1 como puntuación de afinidad, lo que dio como resultado 9 clasificaciones, uno para cada familia.

Cluster ID	Tamaño inicial Cluster	moléculas recuperadas del acoplamiento de conjuntos	Porcentaje de recuperación moléculas recuperadas del acoplamiento (%)	Moléculas con YN1 =>1	Porcentaje de recuperación de moléculas YN1 (%)
4	59	58	98	25	43
19	64	59	92	30	51
20	261	156	60	66	42
23	1455	291	20	95	33
31	48	41	85	11	27
35	68	62	91	44	71

36	47	47	100	20	43
42	1038	241	23	139	58
50	68	46	68	16	35

Tabla 1. Resultados del acoplamiento de conjuntos y la implementación de YN1. Tamaño inicial del grupo: el número de moléculas perteneciente a la familia antes del acoplamiento del conjunto. Moléculas recuperadas del acoplamiento de conjuntos: el número de moléculas reportado por Glide con un valor XP Gscore. Porcentaje de recuperación de moléculas recuperadas de Ensamble Acoplamiento (%): El porcentaje de moléculas reportado por Glide con un valor XP Gscore en relación con el número inicial de moléculas. Moléculas con $YN1 \geq 1$: El número de moléculas con un valor YN1 mayor o igual que 1. Recuperación Porcentaje de moléculas con YN1 (%): El porcentaje de moléculas con un valor de YN1 mayor o igual a 1 sobre el número de moléculas recuperadas del acoplamiento del conjunto

Además, se utilizaron gráficos de logIC50 vs YN1 y XP Gscore vs YN1 generado, demostrando una correlación. Un valor menor de logIC50 correspondió a una puntuación YN1 más alta, lo que significa una mayor actividad inhibitoria experimental del compuesto.

Del mismo modo, una puntuación Gscore de XP más baja dio como resultado una puntuación YN1 más alta, lo que indica una mejora computacional actividad inhibitoria. Estos gráficos ilustraron que YN1 facilitó la identificación de compuestos con actividad inhibitoria computacional y experimental superior. Un resumen de los resultados del acoplamiento del conjunto se presenta en Tabla 1. No todas las moléculas que se sometieron a un acoplamiento de conjunto fueron necesariamente informados por Glide en la salida, ya que podrían tener un GPscore de XP pobre o que no hayan podido caber en el sitio activo de los receptores. Por esta razón, la Tabla 1 muestra el porcentaje de recuperación de moléculas recuperadas del conjunto acoplamiento (%), que representa el porcentaje de moléculas que se seleccionaron del conjunto.

El proceso de acoplamiento del conjunto y la posterior puntuación de YN1 permitieron identificar los más inhibidores prometedores de huAChE. Establecemos un umbral de puntuación YN1 de 1 o superior, lo que significa que solo los compuestos con una puntuación de al menos 1 se clasificaron como moléculas activas.

Dado que la puntuación YN1 oscila entre 0 y 2, y que los términos relacionados con la inhibición experimental y computacional varían de 0 a 1, se eligió este umbral para garantizar una evaluación equilibrada. Este umbral permitió identificar compuestos que demostraron un rendimiento promedio tanto en experimentos como en evaluaciones computacionales, exhibieron un rendimiento experimental

deficiente o mostraron un rendimiento experimental fuerte pero un rendimiento computacional débil.

El uso de los valores de IC50 complementa los valores moleculares, acoplando las clasificaciones con evidencia empírica. Los valores de IC50 se derivan de la biología in vitro, imitando el entorno fisiológico de la proteína, midiendo directamente una capacidad del compuesto para inhibir la proteína.

Esta medida directa ofrece una correlación más fuerte con actividad biológica *in vivo*. El acoplamiento molecular, sin embargo, es una técnica in-silico que predice unión de proteínas compuestas a través de simulaciones computacionales, que proporcionan información sobre interacciones y posibles sitios de unión, pero no garantiza la actividad biológica.

Finalmente, de cada ranking se extrajo el ligando con mejor YN1, resultando 9 ligandos presentando el mejor equilibrio entre la puntuación de acoplamiento y los valores de IC50. Entre los compuestos seleccionados, inhibidores con valores de IC50 inferiores a 420 nM y XP GScore inferiores a - se obtuvieron 9,4 kcal/mol, demostrando la interesante capacidad inhibitoria de los ligandos.

Modelado de aprendizaje automático.

Para esta etapa, tomamos las moléculas obtenidas de la salida de cada uno de los 9 conjuntos acoplados a los que se consideraron como activos. Se generaron 9 modelos de aprendizaje automático ML con capacidad de clasificación de moléculas como inactivas o activas con respecto a cada familia de inhibidores identificada.

El modelo se centra en la identificación con éxito las moléculas que pertenecen a cada familia, en lugar de centrarse en la identificación de las moléculas más potentes.

Las moléculas recuperadas del acoplamiento del conjunto, como se informa en la Tabla 1, se dividieron en librerías asignando una porción del 20% para pruebas y una porción del 80% para entrenamiento. Además, se generaron alrededor de 50 señuelos para cada compuesto activo para proporcionar resultados positivos y negativos.

En consecuencia, se formaron dos librerías desequilibradas, que comprenden instancias positivas y negativas, y asignaciones, por separado, para fines de capacitación y pruebas. Para garantizar un entrenamiento equilibrado, es imperativo establecer una librería equilibrada. Para lograr esto, se implementó un protocolo que involucra agrupamiento basado únicamente en los descriptores moleculares de las instancias negativas.

Este protocolo permitió la extracción de los más diversos señuelos, dando como resultado instancias negativas con un amplio rango de los valores de los descriptores moleculares, lo que facilita el entrenamiento óptimo del modelo.

Se procedió a entrenar cuatro algoritmos de clasificación: regresión logística, soporte máquina de vectores, árboles de decisión y bosques aleatorios. Durante esta fase de formación, se emplearon técnicas de optimización de hiperparámetros para maximizar el rendimiento de cada algoritmo utilizando los datos disponibles. Posteriormente, los valores medios de precisión, memoria, especificidad, y el AUC de ROC se calcularon a lo largo de múltiples iteraciones de validación cruzada.

Familia	Entrenamiento		Pruebas	
	No. de actividades	No de señuelos	No. de actividades	No de señuelos
C4	47	2450	11	637
C19	48	2351	11	539
C20	125	6135	31	1519
C23	233	13261	58	3649
C31	33	1617	8	392
C35	50	3220	12	931
C36	38	3234	9	833
C42	193	11502	48	2695
C50	37	1813	9	441

Tabla 2. División de datos para entrenamiento y pruebas.

Después de la fase de entrenamiento, los modelos se sometieron a pruebas utilizando la librería de pruebas designada.

Es fundamental subrayar que a lo largo de la fase de entrenamiento, los modelos de aprendizaje automático (ML) nunca se encontraron con la librería de pruebas, lo que garantiza su capacidad para predecir clasificaciones para datos desconocidos. Durante las pruebas, se calcularon los valores de precisión, recuperación, especificidad y ROC-AUC.

Además, se realizaron cuatro pruebas complementarias en los cuatro modelos de ML. Estas pruebas incluída la evaluación de la capacidad de los modelos para identificar los señuelos excluidos durante la generación de la librería de entrenamiento equilibrada como inactiva. Además, la capacidad de los modelos para clasificar los 4643 iniciales fueron evaluados.

Se examinaron las moléculas de la base de datos de unión como activas o inactivas, la aptitud para clasificar todas las moléculas pertenecientes al grupo/familia como activas o inactivas, así como su capacidad para clasificar moléculas de la base de datos de unión que no pertenecen al clúster/familia.

Al seleccionar el algoritmo óptimo para cada una de las nueve familias de inhibidores, se emplea una función objetivo compuesta por la suma de: valores de especificidad, la proporción de negativos verdaderos identificados dentro de los señuelos excluidos del entrenamiento, el porcentaje de compuestos activos

identificados de la familia, y el porcentaje de compuestos de la base de datos de unión no asociados a la familia. Esta elección fue impulsada por la importancia crítica de la alta especificidad, donde los modelos de aprendizaje automático (ML) actúan como el filtro principal y deben distinguir de manera efectiva compuestos inactivos. La identificación de negativos verdaderos dentro de señuelos excluidos del entrenamiento Sirve como validación adicional de la capacidad del modelo para reconocer compuestos inactivos.

Dado que estas moléculas no se utilizaron en el proceso de entrenamiento del modelo, proporcionan un medio para evaluar la capacidad del modelo para identificar moléculas inactivas desconocidas. La proporción de activos compuestos identificados de la familia ofrece una confirmación adicional de la competencia del modelo en reconocimiento de compuestos activos.

Por último, teniendo en cuenta el porcentaje de compuestos que no pertenecen a la familia, el modelo debe identificar, adecuadamente, una alta proporción de estos compuestos como inactivos. Es imperativo aclarar que el término inactivo, en el contexto de la presente descripción, denota la ausencia de capacidad inhibitoria de la molécula frente a la AChE.

Modelado de farmacóforos basado en ligandos.

Se utilizaron las moléculas derivadas de los 9 acoplamientos del conjunto, restringiendo la consideración a los acoplamientos activos (compuestos con una puntuación YN1 de 1 o superior). Los modelos de farmacóforos se generaron utilizando el Módulo de fase de Schrödinger 2023.1. Sin embargo, para validar estos modelos, los señuelos fueron esenciales.

Empleamos el paquete DUDE(Mysinger et al., 2012) para generar señuelos en una proporción de aproximadamente 50 señuelos por compuesto activo. Por lo tanto, producimos 9 librerías que comprenden moléculas con una puntuación YN1 mayor igual o igual a 1 junto con sus respectivos señuelos. El cálculo de la fase dio como resultado la creación de múltiples modelos de farmacóforos basados en ligandos para cada uno de los nueve grupos. Sin embargo, solo se utilizaron los modelos con el área más alta bajo la curva ROC (ROC AUC).

El examen de las métricas de rendimiento, en particular de los valores de AUC de ROC, indica que la mayoría de los modelos exhiben un alto poder predictivo. Por ejemplo, modelos como C-31 (AHHR-1) y C-4 (ARR-3) alcanzar valores de AUC ROC de 0,90 y 0,92, respectivamente, que son indicativos de una fuerte capacidad de discriminación. Estos altos valores de AUC se complementan con otras métricas como el recuerdo (RC) y la especificidad (SP), con algunos modelos que logran un

recuerdo perfecto (por ejemplo, C-4 con un RC de 1), destacando su capacidad para identificar correctamente todos los compuestos activos.

Modelado de farmacóforos basado en complejos de conjuntos.

Los 9 ligandos con el mejor YN1 de las 9 familias se emplearon para realizar nueve MD de 500 ns simulaciones. Para lograr esto, las poses de los ligandos se extrajeron de la acoplamiento de conjuntos. Se realizó un agrupamiento de la geometría del ligando, donde las poses asociadas con las 50 conformaciones de los receptores se agruparon para extraer la pose más frecuente. La pose con mayor frecuencia se definió como el centroide del cúmulo más poblado y se extrajo como la posición inicial para el grupo molecular de 500 ns.

Se definió una interacción hidrofóbica si la distancia entre el ligando y el residuo era menor igual o igual a 3,9 angstroms, mientras que la existencia de una interacción de ocupación de enlace de hidrógeno del ligando H se definió si la distancia entre el ligando y el aceptor de enlaces de hidrógeno o el residuo donante fue inferior a 3,35 angstroms.

Los siguientes residuos obtuvieron un porcentaje igual o superior al 80% en términos de contactos hidrofóbicos para varias simulaciones de MD, con el número total de apariciones en paréntesis:

Trp-86(9), Tyr-341(9), Tyr-124(8), Tyr-337(8), Gly-121(6), Trp-286(6), Tyr-72(6), Asp-74(4), Gly-448(4), Hid-447(4), Met-85(4), Phe-297(4), Phe-338(4), Thr-75(4), Thr-83(4), Glu-292(3), Ser-125(3), Val-73(3), Asn-87(2), Leu-76(2), Phe-80(2), Ser-293(2), Tyr-449(2), Tyr-77(2), Val-294(2), Ala-343(1), Arg-296(1), Gln-71(1), Gly-122(1), Gly-342(1), Gly-362(1), Gly-366(1), Gly-82(1), Ile-451(1), Leu-339(1), Leu-398(1), Lys-348(1), Phe-295(1), Pro-344(1), Ser-438(1), Trp-439(1), Tyr-133(1), Val-132(1), Val-365(1). Entre estos, los residuos que aparecieron en 4 o más simulaciones MD fueron: Trp-86, Tyr-341, Tyr-124, Tyr-337, Gly-121, Trp-286, Tyr-72, Asp-74, Gly-448, Hid-447, Met-85, Phe-297, Phe-338, Thr-75 y Thr-83.

Las 9 simulaciones MD se emplearon para extraer conformaciones de proteínas para generar modelos de farmacóforos basados en ligandos. Para ello, los valores RMSF calculados para átomos pesados (columna vertebral) y cadena lateral). Los valores RMSF para Tyr-72, Asp-74, Trp-86, Gly-120, Gly-121, Tyr-124, Tyr-133, Glu-202, Ser-203, Trp-286, Phe-295, Phe-297, Tyr-337, Phe-338, Tyr-341, His-447 y Gly-448 fueron seleccionados y clasificados de mayor a menor, dando como resultado 6 residuos siendo elegidos de cada una de las 9 simulaciones (Cuadro S17). Estos residuos presentaron los mayores variabilidad conformacional a lo largo de toda la simulación.

Se utilizaron conformaciones estables e inestables para construir farmacóforos basados en modelos complejos utilizando el módulo Fase de Schrödinger 2023-1.

Estos modelos se combinaron en un conjunto. Esta combinación de modelos de farmacóforos basados en complejos se denomina dentro del contexto de la presente descripción como "Conjunto de modelos de farmacóforos". Dentro de este modelo, las moléculas se someten a un cribado virtual contra todos los modelos de farmacóforos derivados de conformaciones de modelos estables e inestables.

Posteriormente, los ligandos son identificados como activos o inactivos por cada modelo farmacofórico individual. Una molécula se identifica inicialmente como activa si el modelo de farmacóforo lo informa en la salida del cálculo de cribado virtual, y se considera inicialmente una molécula inactiva si no se encuentra en la salida del cálculo de cribado virtual.

Al combinar todos los modelos de farmacóforos en el conjunto, para determinar si una molécula es activa o inactiva, se realiza una comparación de los resultados individuales de cada modelo de farmacóforo utilizando el operador lógico OR. Esto significa que para que una molécula sea activa, basta con un modelo individual para clasificarla como activa, y para que una molécula sea inactiva, no debería serlo clasificados como activos por cualquier modelo.

Por último, se definió la "puntuación de fase de conjunto" como la más alta puntuación de fase obtenida por una molécula entre todas las puntuaciones de fase obtenidas para cada farmacóforo modelado en el conjunto. Esto representa la primera forma de un conjunto de modelos de farmacóforos.

Para mejorar los resultados y el comportamiento de los conjuntos, se emplea la métrica Yasser número dos (YN2), que combina los valores de la puntuación de fase del conjunto, el número de estables coincidentes conformaciones, y el número de conformaciones inestables coincidentes. Esta métrica promedia los valores normalizados de la puntuación de fase del conjunto, el número de conformaciones estables coincidentes y el número de conformaciones inestables coincidentes. Un valor alto de YN2 indica que la molécula en cuestión tiene una excelente puntuación de fase de conjunto y encaja con la mayoría de las conformaciones estables e inestables.

Para que una molécula se considere activa empleando YN2, debe tener un valor mayor o igual a un umbral YN2; Por el contrario, para que una molécula se considere inactiva, debe tener un valor inferior a un umbral YN2. Para realizar el cribado virtual utilizando el modelo de farmacóforo de conjuntos acoplados con YN2, se requieren 5 parámetros: el número de Conformaciones, puntuación de $Phscore_{min}$ El valor mínimo de la puntuación de fase de conjunto entre los activos conocidos; Puntuación de $Phscore_{max}$ el valor máximo de la puntuación de fase del conjunto entre las moléculas activas conocidas y el umbral YN2. El número de conformaciones es fijo y conocido por el experimentador, mientras que los demás parámetros deben obtenerse en el proceso de validación del conjunto.

Para determinar el mejor umbral de YN2, generamos una función objetivo que es la suma de la recordación y la especificidad. Posteriormente, llevamos a cabo una optimización de una función objetivo correspondiente a la suma de la especificidad y el recuerdo variando el valor de YN2. De esta manera, se obtuvo el valor YN2 que produjo el valor máximo para la función objetivo. Además, los valores del número de estables y conformaciones inestables, Puntuación de $Phscore_{min}$ (el valor mínimo de la puntuación de fase de conjunto entre las moléculas activas), $Phscore_{max}$ (el valor máximo de puntuación de fase de conjunto entre activos conocidos).

Mediante el empleo de ensamble de modelos de farmacóforos, se obtuvo información valiosa sobre las diversas proteínas conformaciones a las que un ligando podría unirse, mejorando la precisión predictiva de la actividad inhibitoria.

En este contexto, resulta fundamental vincular los modelos de farmacóforos correspondientes a varias conformaciones de sitios activos utilizando el OR lógico y YN2 en los activos identificados de cada modelo individualmente. Este enfoque permite la flexibilidad del sitio activo, en contraste con el uso de un único modelo de farmacóforo. Además, se destaca la importancia de los volúmenes de exclusión en modelos de base compleja.

Estos volúmenes imitan los efectos estéricos del sitio activo, evitando eficazmente el reconocimiento de ligandos con geometrías desfavorables como activas. Los volúmenes de exclusión tienen como objetivo principal aumentar la especificidad y selectividad de los modelos de farmacóforos mediante la reducción del espacio de búsqueda, minimizando así los positivos resultantes de un posicionamiento incorrecto del ligando.

Respecto a la búsqueda de fármacos selectivos, resulta crucial abordar de forma individual conformaciones inestables. Por ejemplo, dirigirse a conformaciones de proteínas específicas puede producir más fármacos eficaces con menos efectos secundarios al reducir la probabilidad de interferencia con el funcionamiento de otras enzimas.

Cribado virtual de la base de datos.

El efecto técnico sorprendente aportado por el método objeto de la presente descripción consiste en implementar modelos predictivos para identificar moléculas que podrían funcionar, particularmente, como nuevos inhibidores de la AChE.

Para lograr este objetivo, se realizó un uso secuencial de modelos de aprendizaje automatizado ML, modelos de farmacóforos basados en ligandos y modelos de farmacóforos basados en complejos.

Como ejemplo de aplicación del método, se extrajeron 11.012.710 moléculas de la base de datos ZINC. Estas moléculas se sometieron a un cribado virtual con los

modelos de 9 ML, y las moléculas activas por cada modelo se agregaron en una biblioteca combinada, lo que resultó en un total de 1.697.608 moléculas recuperadas, lo que representa el 15,41% de la librería original de ZINC.

Posteriormente, estas moléculas se sometieron a cribado con 9 modelos de farmacóforos basados en ligandos y 9 conjuntos de modelos de farmacóforos basados en complejos. En la Figura 6 se presenta un esquema que resume los resultados. De cada lista final de moléculas, las 500 moléculas principales con las puntuaciones YN2 más altas fueron seleccionadas y sometidas a cálculos de acoplamiento de conjuntos y cálculos de energía libre MMGBSA.

- Los resultados del cribado virtual subrayan la eficacia del método objeto de la presente descripción que combina datos computacionales y experimentales a través del desarrollo y la aplicación de aprendizaje automático y modelado de farmacóforos.

El orden de aplicación del filtro fue específicamente diseñado para progresar de menor a mayor la precisión en la búsqueda de moléculas. Inicialmente, los modelos de ML fueron entrenados para identificar moléculas de cada familia, a diferencia de los basados en ligandos y complejos modelos de farmacóforos, que se entrenaron con las moléculas que presentaban la mayor actividad YN1 de cada familia.

Los modelos de ML resultaron muy útiles, ya que requerían menores costos computacionales en comparación con el uso directo de modelos de farmacóforos. Esto permitió una gestión exitosa de la inicial de ZINC de 11.012.710 moléculas, reduciéndola hasta en un 15% con potencia computacional y tiempos de cálculo.

Posteriormente, las moléculas se sometieron a un modelado de farmacóforos basado en ligandos y a Modelado de conjuntos de farmacóforos basado en el mismo. El modelado de fármacos es una herramienta poderosa para, pero su eficacia depende en gran medida de la disponibilidad tanto de moléculas activas como del objetivo biológico. Sin embargo, cuando sólo se conoce el objetivo biológico, este enfoque se topa con varios desafíos, lo que dificulta la identificación de candidatos a fármacos prometedores.

Si bien es particularmente cierto que en los farmacóforos basados en complejos, que utilizan la estructura del receptor en complejo con el ligando para la creación de modelos, los puntos farmacofóricos representan regiones donde ocurren contactos favorables proteína-ligando, no obstante, en el caso de los modelos de farmacóforos basados en ligandos, la hipótesis podría no capturar ubicaciones donde se producen contactos de unión.

Por lo anterior, en el método objeto de divulgación se posicionan los modelos basados en ligandos en segundo lugar, ya que los modelos complejos proporcionan información más precisa para identificar moléculas activas. Esto es evidente al

comparar las métricas superiores del conjunto de farmacoforos de base compleja con los de los modelos basados en ligandos.

Además, otra razón para priorizar el modelo basado en ligando, antes que los modelos basados en complejos, es el costo computacional, ya que los conjuntos de farmacóforos requieren más recursos. Por lo tanto, el método objeto de la presente descripción minimiza el número de moléculas, optimizando el tiempo y los recursos computacionales.

Aunque la presente invención ha quedado descrita con las realizaciones preferentes mostradas, queda entendido que las modificaciones y variaciones que conserven el espíritu y el alcance de esta invención se entienden dentro del alcance de las reivindicaciones adjuntas.

REFERENCIAS

- 5 Akhoun, B. A., Choudhary, S., Tiwari, H., Kumar, A., Barik, M. R., Rathor, L., Pandey, R., & Nargotra, A. (2020). Discovery of a new donepezil-like acetylcholinesterase inhibitor for targeting Alzheimer's disease: Computational studies with biological validation. *Journal of Chemical Information and Modeling*, 60(10), 4717–4729. https://doi.org/10.1021/ACS.JCIM.0C00496/SUPPL_FILE/CIOC00496_SI_001.PDF
- 10 Anderson, A. C., O'Neil, R. H., Surti, T. S., & Stroud, R. M. (2001). Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. *Chemistry & Biology*, 8(5), 445–457. [https://doi.org/10.1016/S1074-5521\(01\)00023-0](https://doi.org/10.1016/S1074-5521(01)00023-0)
- 15 Branduardi, D., Gervasio, F. L., Cavalli, A., Recanatini, M., & Parrinello, M. (2005). The role of the peripheral anionic site and cation-pi interactions in the ligand penetration of the human AChE gorge. *Journal of the American Chemical Society*, 127(25), 9147–9155. <https://doi.org/10.1021/JA0512780>
- Cheng, S., Song, W., Yuan, X., & Xu, Y. (2017). Gorge Motions of Acetylcholinesterase Revealed by Microsecond Molecular Dynamics Simulations. *Scientific Reports 2017 7:1*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-03088-y>
- 20 Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M. S., Gary, E. N., Love, J., Franklin, M. C., & Height, J. J. (2012). Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *Journal of Medicinal Chemistry*, 55(22), 10282–10286. https://doi.org/10.1021/JM300871X/SUPPL_FILE/JM300871X_SI_001.PDF
- Ertl, P., Schuffenhauer, A., & Renner, S. (2011). Chemoinformatics and Computational Chemical Biology. *Life Sciences*, 672(1), 588. <https://doi.org/10.1007/978-1-60761-839-3>
- 25 Evangelista Falcon, W., Ellingson, S. R., Smith, J. C., & Baudry, J. (2019). Ensemble Docking in Drug Discovery: How Many Protein Configurations from Molecular Dynamics Simulations are Needed To Reproduce Known Ligand Binding? *The Journal of Physical Chemistry. B*, 123(25), 5189–5195. <https://doi.org/10.1021/ACS.JPCB.8B11491>
- 30 Galburt, E. A., & Tomko, E. J. (2017). Conformational selection and induced fit as a useful framework for molecular motor mechanisms. *Biophysical Chemistry*, 223, 11–16. <https://doi.org/10.1016/J.BPC.2017.01.004>
- 35 Gao, W., Ma, X., Yang, H., Luan, Y., & Ai, H. (2022). Molecular engineering and activity improvement of acetylcholinesterase inhibitors: Insights from 3D-QSAR, docking, and molecular dynamics simulation studies. *Journal of Molecular Graphics and Modelling*, 116, 108239. <https://doi.org/10.1016/J.JMGM.2022.108239>
- Kapoor, K., Thangapandian, S., & Tajkhorshid, E. (2022). Extended-ensemble docking to probe dynamic variation of ligand binding sites during large-scale structural changes of proteins. *Chemical Science*, 13(14), 4150–4169. <https://doi.org/10.1039/D2SC00841F>
- 40 Langer, T., & Hoffmann, R. D. (2006). Pharmacophores and Pharmacophore Searches. In T. Langer & R. D. Hoffmann (Eds.), *Pharmacophores and Pharmacophore Searches* (1ra ed., Vol. 32). Wiley. <https://doi.org/10.1002/3527609164>
- Lu, C., & Li, A. P. (2010). Enzyme Inhibition in Drug Discovery and Development: The Good and the Bad. *Enzyme Inhibition in Drug Discovery and Development: The Good and the Bad*. <https://doi.org/10.1002/9780470538951>

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/SUPPL_FILE/JM300687E_SI_004.TXT

- 5 Osguthorpe, D. J., Sherman, W., & Hagler, A. T. (2012). Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chemical Biology & Drug Design*, 80(2), 182–193. <https://doi.org/10.1111/J.1747-0285.2012.01396.X>

- Pérez-Sánchez, H., Den Haan, H., Pérez-Garrido, A., Penã-García, J., Chakraborty, S., Erdogan Orhan, I., Senol Deniz, F. S., & Villalgordo, J. M. (2021). Combined Structure and Ligand-Based Design of Selective Acetylcholinesterase Inhibitors. *Journal of Chemical Information and Modeling*, 61(1), 467–480.
10 https://doi.org/10.1021/ACS.JCIM.0C00463/ASSET/IMAGES/MEDIUM/CI0C00463_0016.GIF

- Qing, X., Lee, X. Y., De Raeymaeker, J., Tame, J. R., Zhang, K. Y., De Maeyer, M., & Voet, A. R. (2014). Pharmacophore modeling: Advances, Limitations, And current utility in drug
15 discovery. *Journal of Receptor, Ligand and Channel Research*, 7, 81–92.
<https://doi.org/10.2147/JRLCR.S46843>

- Schaller, D., Šribar, D., Noonan, T., Deng, L., Nguyen, T. N., Pach, S., Machalz, D., Bermudez, M., & Wolber, G. (2020). Next generation 3D pharmacophore modeling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(4), e1468.
20 <https://doi.org/10.1002/WCMS.1468>

Senapati, S., Bui, J. M., & McCammon, J. A. (2005). Induced fit in mouse acetylcholinesterase upon binding a femtomolar inhibitor: a molecular dynamics study. *Journal of Medicinal Chemistry*, 48(26), 8155–8162. <https://doi.org/10.1021/JM050669M>

- Van Der Westhuizen, C. J., Stander, A., Riley, D. L., & Panayides, J. L. (2022). Discovery of Novel Acetylcholinesterase Inhibitors by Virtual Screening, in Vitro Screening, and Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, 62(6), 1550–1572.
25 https://doi.org/10.1021/ACS.JCIM.1C01443/SUPPL_FILE/CI1C01443_SI_001.PDF

- Wilkinson, D. G., Francis, P. T., Schwam, E., & Payne-Parrish, J. (2012). Cholinesterase Inhibitors Used in the Treatment of Alzheimer’s Disease. *Drugs & Aging* 2004 21:7, 21(7), 453–478.
30 <https://doi.org/10.2165/00002512-200421070-00004>

5

10