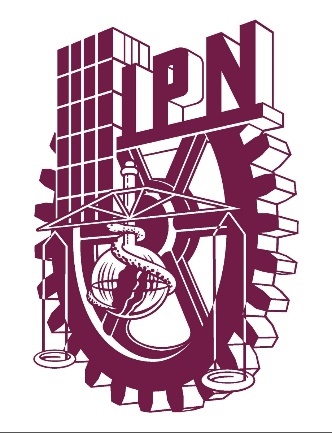
**INSTITUTO POLITÉCNICO NACIONAL**

**CENTRO DE BIOTECNOLOGÍA GENÓMICA**



**“CRIBADO VIRTUAL INTELIGENTE PARA IDENTIFICAR INHIBIDORES MULTI-BLANCO ENFOCADOS AL TRATAMIENTO DE LA**

**ASOCIACIÓN ALZHEIMER-DIABETES MELLITUS”**

ANTEPROYECTO DE INVESTIGACIÓN

**MAESTRÍA EN CIENCIAS EN BIOTECNOLOGÍA GENÓMICA**

PRESENTA

**ELIUD ULISES AGUILAR DURÁN**

CD. REYNOSA, TAMAULIPAS, MÉXICO JUNIO, 2024

**INSTITUTO POLITÉCNICO NACIONAL**

**CENTRO DE BIOTECNOLOGÍA GENÓMICA**



**“CRIBADO VIRTUAL INTELIGENTE PARA IDENTIFICAR INHIBIDORES MULTI-BLANCO ENFOCADOS AL TRATAMIENTO DE LA**

**ASOCIACIÓN ALZHEIMER-DIABETES MELLITUS”**

ANTEPROYECTO DE INVESTIGACIÓN

**MAESTRÍA EN CIENCIAS EN BIOTECNOLOGÍA GENÓMICA**

PRESENTA

**ELIUD ULISES AGUILAR DURÁN**

CD. REYNOSA, TAMAULIPAS, MÉXICO JUNIO, 2024

**ÍNDICE**

**Sección Página**

[LISTA DE CUADROS VI](#_Toc169120106)

[LISTA DE FIGURAS VII](#_Toc169120107)

[LISTA DE SÍMBOLOS Y NOMENCLATURA VIII](#_Toc169120108)

[1. INTRODUCCIÓN 1](#_Toc169120109)

[2. ANTECEDENTES 3](#_Toc169120110)

[2.1. La enfermedad de Alzheimer 3](#_Toc169120111)

[2.1.1 La demencia 3](#_Toc169120112)

[2.1.2. El Alzheimer 3](#_Toc169120113)

[2.1.3. Fisiopatología 4](#_Toc169120114)

[2.1.4. Patogénesis 7](#_Toc169120115)

[2.2. Diabetes mellitus 8](#_Toc169120116)

[2.2.1. La Diabetes Mellitus 8](#_Toc169120117)

[2.2.2. Fisiopatología y patogénesis 8](#_Toc169120118)

[2.3. Relación Alzheimer-Diabetes Mellitus 10](#_Toc169120119)

[2.3.1. Estudios cognitivos sobre pacientes con diabetes mellitus 10](#_Toc169120120)

[2.3.2. Vínculos moleculares entre EA y DM2 11](#_Toc169120121)

[2.4. Diseño de fármacos 14](#_Toc169120122)

[2.4.1. El diseño de fármacos 14](#_Toc169120123)

[2.4.2. La inteligencia artificial en el diseño de fármacos 15](#_Toc169120124)

[2.4.3 Machine Learning 16](#_Toc169120125)

[2.4.4. Machine Learning Supervisado 17](#_Toc169120126)

[2.4.5. Machine learning no supervisado 18](#_Toc169120127)

[2.4.6 Deep Learning 18](#_Toc169120128)

[2.5. Diseño de fármacos para la enfermedad de Alzheimer 19](#_Toc169120129)

[2.6.1. Fármacos de EA para una sola diana 21](#_Toc169120130)

[2.6.2. Fármacos de EA para modular múltiples dianas 22](#_Toc169120131)

[3. JUSTIFICACIÓN 28](#_Toc169120132)

[4. HIPÓTESIS 29](#_Toc169120133)

[5. OBJETIVOS 30](#_Toc169120134)

[6. MATERIALES Y MÉTODOS 31](#_Toc169120135)

[6.1. Recopilación y análisis de datos experimentales de inhibición 31](#_Toc169120136)

[6.1.1. Obtención de los datos experimentales de los inhibidores 32](#_Toc169120137)

[6.1.2. Tratamiento de datos 32](#_Toc169120138)

[6.1.3. Establecimiento del umbral de inhibición 33](#_Toc169120139)

[6.1.4. Cálculo de los descriptores moleculares 34](#_Toc169120140)

[6.1.5. Filtros de baja varianza y de alta correlación 35](#_Toc169120141)

[6.2. Entrenamiento de los modelos de aprendizaje automático 35](#_Toc169120142)

[6.2.1. División del conjunto de datos 35](#_Toc169120143)

[6.2.2. Validación interna 36](#_Toc169120144)

[6.2.3. Entrenamiento de los modelos y validación externa 38](#_Toc169120145)

[6.3. Determinación de potenciales inhibidores 38](#_Toc169120146)

[6.3.1. Obtención de las moléculas a partir de las bases de datos 38](#_Toc169120147)

[6.3.2. Tratamiento de los datos descargados 39](#_Toc169120148)

[6.3.3. Cálculo de descriptores moleculares 39](#_Toc169120149)

[6.3.4. Determinación de potenciales inhibidores multiblanco 40](#_Toc169120150)

[6.4. Acoplamiento molecular 40](#_Toc169120151)

[6.4.1. Preparación del receptor 40](#_Toc169120152)

[6.4.2. Preparación del ligando 41](#_Toc169120153)

[6.4.3. Acoplamiento molecular 41](#_Toc169120154)

[7. CRONOGRAMA DE ACTIVIDADES 42](#_Toc169120155)

[8. BIBLIOGRAFÍA 43](#_Toc169120156)

# LISTA DE CUADROS

**Cuadro Página**

[**Cuadro 1.** Tabla que muestra la elación entre las proteínas diana con la EA y la DM2…..25](#_Toc169118811)

# LISTA DE FIGURAS

**Figura Página**

[**Figura 1.** Procesamiento de la APP 5](#_Toc169119069)

[**Figura 2.** Nuevo paradigma del Machine Learning 16](#_Toc169119070)

[**Figura 3.** Clasificación de los modelos de Machine learning. 17](#_Toc169119071)

[**Figura 4.** Esquema general de la metodología. 31](#_Toc169119072)

[**Figura 5.** Esquema general del tratamiento de datos y selección de columnas de interés. 33](#_Toc169119073)

[**Figura 6.** Ejemplo de los DataFrames con las columnas filtradas y la columna de actividad añadida. 34](#_Toc169119074)

[**Figura 7.** Esquema ilustrativo de la creación de los conjuntos de entrenamiento y prueba. 36](#_Toc169119075)

[**Figura 8.** Árbol de decisión para la aprobación de créditos bancarios a clientes. 37](#_Toc169119076)

# LISTA DE SÍMBOLOS Y NOMENCLATURA

|  |  |
| --- | --- |
| % | Porciento. |
| α | Alfa. |
| β | Beta. |
| γ | Gamma. |
| IC50 | Concentración inhibitoria media. |
| nM | Nano molar. |
| μM | Micro molar. |
| ADAM | Una desintegrina y metaloproteinasapor, sus siglas en inglés “A Disintegrin And Metalloproteinase”. |
| APOE | Alipoproteína E. |
| APP | Proteína precursora amiloide, por sus siglas en inglés "Amyloid Precursor Protein". |
| Aβ | Proteína beta-amiloide. |
| BACE1 | Beta Site APP Cleaving Enzyme 1. |
| BD | Base de datos. |
| CADD | Diseño de fármacos asistido por computadora, por sus siglas en inglés “Computer, Aided Drug Design”. |
| DBSCAN | Agrupación espacial basada en densidad de aplicaciones con ruido, por sus siglas en inglés “Density-Based Spatial Clustering of Applications with Noise”. |
| DL | Aprendizaje profundo, por sus siglas en inglés “Deep Learning". |
| DM | Diabetes mellitus. |
| DM1 | Diabetes mellitus tipo 1. |
| DM2 | Diabetes mellitus tipo 2. |
| DT | Árboles de decisión, por sus siglas en inglés “Decision Trees”. |
| EA | Enfermedad de Alzheimer. |
| EGFR | Receptor del factor de crecimiento epidérmico, por sus siglas en inglés "Epidermal Growth Factor Receptor". |
| ESR1 | Receptor de estrógenos 1, por sus siglas en inglés "Estrogen Receptor 1". |
| FDA | Adminsitración de comida y fármacos, por sus siglas en inglés “Food and Drug Administration”. |
| GSK-3 | Glucógeno Sintasa Cinasa 3 |
| HLA | Antígeno leucocitario humano, por sus siglas en inglés "human leukocyte antigen". |
| HSP90AA1 | Proteína de choque térmico HSP 90-alfa. |
| IA | Inteligencia artificial. |
| IDE | Enzima degradadora de insulina, por sus siglas en inglés "Insulin Degrading Enzyme". |
| IGF-1 | Factor de crecimiento similar a la insulina, por sus siglas en inglés “Insuline Grow Factor”. |
| IRS1 | Sustrato del receptor de insulina 1, por sus siglas en inglés “Insulin Receptor Substrate 1". |
| k-NN | Vecinos más cercanos, por sus siglas en inglés “k-Nearest Neighboors". |
| LBDD | Diseño de fármacos basado en el ligando, por sus siglas en inglés “Ligand Based Drug Design”. |
| MAPK1 | MAP quinasa 1. |
| ML | Aprendizaje automático, por sus siglas en inglés “Machine Learning”. |
| NMDA | N-metil-D-aspartato. |
| PCA | Análisis de componentes principales, por sus siglas en inglés “Principal Component Analysis. |
| PDB | Protein Data Bank. |
| PIK3R1 | Fosfatidilinositol 3-quinasa. |
| PSEN1 | Presenilina 1. |
| PSEN2 | Presenilina 2. |
| QSAR | Relación Cuantitativa Estructura-Actividad, por sus siglas en inglés “Quantitative Structure-Activity Relationship”. |
| IR | Receptor de insulina, por sus siglas en inglés “Insuline Receptor”. |
| RMSD | Desviación cuadrada de raíz, por sus siglas en inglés “Root-Mean Square Deviation”. |
| SBDD | Diseño de fármacos basado en la estructura, por sus siglas en inglés “Structure Based Drug Design”. |
| SMILES | Especificación de entrada de línea de entrada molecular simplificada, por sus siglas en inglés “Simplified Molecular Input Line Entry Specification”. |
| SRC | Protooncogén tirosina-proteína quinasa SRC. |
| STAT3 | Transductor de señal y activador de la transcripción 3, por sus siglas en inglés “Signal Transducer and Activator of Transcription 3". |
| SVM | Máquinas de soporte vectorial, por sus siglas en inglés, “Support Vector Machines”. |
| t-SNE | Incrustación de vecinos estocásticos, por sus siglas en inglés “Stochastic Neighbor Embedding”. |
| UBC | Ubiquitina C. |
| UPR | Respuesta de proteínas desplegadas, por sus siglas en inglés “Unfolded Protein Respondes”. |

# 1. INTRODUCCIÓN

La demencia es actualmente la séptima causa de muerte a nivel mundial y una de las principales causas de discapacidad en personas mayores (OPS, 2023a). La enfermedad de Alzheimer (EA) es la forma más común de demencia, representando casi el 70% de los casos (WHO, 2023). En México, aproximadamente un millón 300 mil personas padecen Alzheimer (SSA México, 2021). Las personas afectadas por esta enfermedad experimentan un deterioro progresivo de habilidades cognitivas como la memoria, la orientación, el lenguaje y el aprendizaje (IMSS, 2015).

Hay disponibles en el mercado varios medicamentos aprobados por la para tratar la EA. Algunos de estos medicamentos son la galantamina, la rivastigmina y el donepezilo (Uddin *et al*., 2020). Sin embargo, estos medicamentos están orientados a disminuir los efectos de la EA en un determinada etapa de la enfermedad y tienen como efectos secundarios náuseas, vómitos, diarrea, reacciones alérgicas, pérdida de apetito, dolor de cabeza, confusión, mareo y caídas (NIH, 2023). Debido a esto, se han realizado muchas investigaciones para determinar potenciales inhibidores para distintos objetivos asociados con la EA. Sin embargo, Uno de los principales problemas del diseño de nuevos fármacos destinados a la EA es su baja tasa de éxito, pues alta tasa de deserción de los fármacos. Se encontró una tasa de éxito general de solo el 0.4% (Cummings *et al*., 2014).

En el mercado existen varios medicamentos aprobados por la Food and Drug Administration (FDA) para el tratamiento de la EA. Entre estos fármacos se encuentran la galantamina, la rivastigmina y el donepezilo (Uddin *et al*., 2020). Sin embargo, estos medicamentos están diseñados para mitigar los efectos de la EA en una etapa específica de la enfermedad y presentan efectos secundarios como náuseas, vómitos, diarrea, reacciones alérgicas, pérdida de apetito, dolor de cabeza, confusión, mareo y caídas (NIH, 2017). Debido a estos inconvenientes, se han llevado a cabo numerosas investigaciones para identificar potenciales inhibidores dirigidos a diversas dianas relacionadas con la EA. No obstante, se ha reportado que la tasa de éxito general para los nuevos fármacos destinados al tratamiento de la EA es de tan solo un 0.4% (Cummings *et al*., 2014).

Para abordar este problema, varios estudios han señalado que se deberían diseñar fármacos multidiana para el tratamiento de la EA, debido su naturaleza multifactorial (Lovestone *et al*., 2015). Existen múltiples factores que pueden ocasionar EA y recientemente se ha identificado la diabetes mellitus tipo 2 (DM2) como una de estas causas (Yang y Song, 2013). Esta enfermedad se caracteriza por niveles elevados de glucosa en la sangre, lo que, a largo plazo, incrementa el riesgo de daño a los ojos, riñones, páncreas, nervios y corazón. Múltiples estudios han confirmado la conexión entre la DM2 y la EA hasta el punto de que esta relación ha sido denominada diabetes mellitus tipo 3 (Brands *et al*., 2005; Kopf y Frölich, 2009; Strachan *et al*., 1997). Debido a esta interesante relación, en la última década los investigadores han probado diversos medicamentos destinados a la diabetes mellitus en pacientes con enfermedad de EA, observando que pueden mejorar la cognición y la memoria en estos pacientes (Michailidis *et al*., 2022). No obstante, hasta la fecha no existen estudios enfocados en el diseño de fármacos específicamente dirigidos a las dianas asociadas a la relación entre Alzheimer y diabetes mellitus tipo 2. Por ello, esta investigación propone emplear modelos de aprendizaje automático para identificar potenciales inhibidores dirigidos a blancos relacionados con esta asociación.

# 2. ANTECEDENTES

# 2.1. La enfermedad de Alzheimer

# 2.1.1 La demencia

La demencia es un término empleado para diversas enfermedades que son crónicas y progresivas y que resultan en deterioros cognitivos que interfieren en la capacidad de una persona de realizar actividades de la vida diaria (OPS, 2023a). Esta suele manifestarse en edades avanzadas, siendo muy poco común en personas menores de 60 años. (Medline, 2024). En la actualidad, esta es la séptima causa de defunción y una de las principales razones de discapacidad y dependencia entre la población de edad avanzada a nivel mundial.

# 2.1.2. El Alzheimer

La enfermedad de Alzheimer es la forma más prevalente de demencia, abarcando entre el 60% y el 70% de los casos de esta (WHO, 2023). Esta enfermedad fue descrita por primera vez en el año 1906 en Alemania por el psiquiatra Alois Alzheimer (Stelzmann *et al*., 1995). Esta tipo de demencia se caracteriza por el deterioro progresivo de las funciones mentales, afectando las neuronas, provocando su degeneración y muerte. Las personas afectadas experimentan un deterioro gradual en habilidades cognitivas como la memoria, la orientación, el lenguaje y el aprendizaje (IMSS, 2015). Se estima que en México aproximadamente un millón 300 mil personas padecen la enfermedad de Alzheimer, lo que representa entre el 60 y el 70 por ciento de los casos de demencia diagnosticados.

La Food and Drug Administration de los Estados Unidos (FDA) ha aprobado varios medicamentos para el tratamiento de la enfermedad de Alzheimer. Estos fármacos están diseñados para controlar los síntomas y tratar la enfermedad, siendo más efectivos en personas en etapas tempranas o intermedias de esta. Sin embargo, no existe actualmente ningún medicamento que cure o revierta los efectos de la EA. Entre los medicamentos aprobados se encuentran la galantamina, la rivastigmina y el donepezilo, todos ellos inhibidores de la colinesterasa. Estos medicamentos pueden ayudar a reducir o controlar algunos síntomas cognitivos y conductuales. Otro medicamento es la mematina, un antagonista del *N*-metil-aspartato (NMDA). Este podría permitir al paciente mantener funciones diarias por un poco más de tiempo del que lo harían sin el medicamento (Uddin *et al*., 2020). Sin embargo, todos estos medicamentos se recetan en dosis bajas para reducir la probabilidad de los efectos secundarios. Estos efectos secundarios incluyen náuseas, vómitos, diarrea, reacciones alérgicas, pérdida de apetito, dolor de cabeza, confusión, mareo y caídas (NIH, 2023).

Las características principales de la EA son la formación de placas amiloides y la acumulación de ovillos neurofibrilares. Las placas amiloides, compuestas principalmente por la proteína beta-amiloide (Aβ), se generan a partir del procesamiento de la proteína precursora amiloide (APP, por sus siglas en inglés "Amyloid Precursor Protein") y se acumulan entre las neuronas. Estas placas interfieren en la comunicación entre las células nerviosas provocando inflamación y activan una respuesta inmunitaria que contribuye a la muerte neuronal (Sun *et al*., 2012). Por otro lado, los ovillos neurofibrilares son formados por la proteína tau. Tau es un grupo de proteínas asociadas a los microtúbulos neuronales que se forman a través de un proceso de empalme alternativo del ARNm. Estos ovillos neurofibrilares se acumulan en el cerebro durante el desarrollo de la enfermedad de Alzheimer (Johnson y Stoothoff, 2004). La acumulación de tau altera el transporte de nutrientes y otras moléculas esenciales dentro de las neuronas, lo que finalmente conduce a la disfunción y muerte de las células neuronales (Zhang *et al*., 2021). Existen diversas teorías acerca de la fisiopatología del EA. Entre ellas, se encuentran la hipótesis amiloide, la hipótesis colinérgica, el papel de la proteína Tau y la implicación del estrés oxidativo y el calcio (Sanabria-Castro *et al*., 2017).

# 2.1.3. Fisiopatología

La hipótesis de la proteína amiloide postula que la enfermedad de Alzheimer se debe a anomalías en el procesamiento de la APP. En este contexto, el desequilibrio entre la producción y la eliminación de beta-amiloide es responsable de la formación anormal de placas amiloides, las cuales provocan la muerte de las células neuronales (Cummings *et al*., 2007). El péptido beta-amiloide es un péptido compuesto por entre 37 y 43 aminoácidos. Las isoformas más comunes son Aβ1-40 y Aβ1-42 (Deane *et al*., 2009), siendo la isoforma Aβ1-42 la más hidrofóbica y la que tiene la mayor toxicidad (Mohandas *et al*., 2009).

La proteína beta-amiloide se origina a través del procesamiento de la APP en la membrana plasmática, como se ilustra en la Figura 1 (Sanabria-Castro *et al*., 2017). En el proceso normal, también conocido como vía no amiloidogénica, la APP es descompuesta principalmente por enzimas con actividad α-secretasa, que pertenecen a las familias de desintegrina y metaloproteasa (ADAM, por sus siglas en inglés “A Disintegrin And Metalloproteinase”). La escisión de la APP por las ADAM resulta en la formación y liberación de APPsα, una forma soluble bajo ciertas condiciones (Tanzi y Bertram, 2005). Durante el proceso de producción de beta-amiloide, también conocido como vía amilodogénica, la primera escisión de APP genera un fragmento amino terminal más corto conocido como APPsβ. Esta división es realizada por la enzima BACE1, una proteasa transmembrana con actividad de β-secretasa. Posteriormente, la γ-secretasa, un complejo compuesto por cuatro subunidades: presenilinas, nicastrina, PEN-2 y APH-1, lleva a cabo otra escisión que libera la beta amiloide, que se acumula en tretrámeros para posteriormente acumularse en forma de placas amiloides (Sanabria-Castro *et al*., 2017).

Gráfico, Gráfico en cascada

Descripción generada automáticamente

**Figura 1.** Procesamiento de la APP (Sanabria-Castro et al., 2017)

La proteína tau es el componente principal de los filamentos helicoidales que conforman los ovillos neurofibrilares en el cerebro de personas con EA, y en estos, esta proteína se encuentra anormalmente fosforilada (Grundke-Iqbal *et al*., 1986). Tau es una fosfoproteína, y su fosforilación anormal impide su capacidad para estimular el ensamblaje de microtúbulos. La proteína tau es una proteína presenta hasta seis isoformas distintas. Contiene 80 residuos de serina o treonina y 5 residuos de tirosina, lo que significa que casi el 20% de la proteína tiene el potencial de ser fosforilada por proteínas quinasas (Goedert *et al*., 1989). La proteína tau probablemente desempeña un papel en la regulación de la extensión de las neuritas, ya que estudios han demostrado que la supresión de su expresión inhibe el crecimiento de estas. Además, está implicada en la regulación del transporte axonal (Ishihara *et al*., 1999). Sin embargo, la fosforilación inapropiada de la proteína Tau está presente en enfermedades neurodegenerativas como la EA, donde resulta en la incapacidad de esta proteína para ensamblar microtúbulos, lo que conduce a la formación de ovillos neurofibrilares (Johnson y Stoothoff, 2004).

La hipótesis colinérgica propone que la disfunción de las neuronas que contienen acetilcolina contribuye al deterioro cognitivo observado tanto en el envejecimiento normal como en la enfermedad de Alzheimer (Terry y Buccafusco, 2003). Los receptores muscarínicos son un tipo de receptor de acetilcolina presentes en las células del sistema nervioso y en diversos tejidos corporales. Los antagonistas de estos receptores, como la escopolamina, bloquean la acción de la acetilcolina en estos receptores, lo que afecta la vía de señalización de la acetilcolina y, por ende, la memoria y otras funciones cognitivas (Liskowsky y Schliebs, 2006). Además, se ha comprobado que tanto los agonistas muscarínicos selectivos como los no selectivos mejoran el aprendizaje y la memoria. Se han identificado agonistas selectivos para los receptores muscarínicos M1, los cuales han demostrado reducir los niveles de beta-amiloide en el líquido cefalorraquídeo (Fisher, 2008). Actualmente, esta hipótesis ha sido la base de varios tratamientos, y existe un consenso en que la relación entre el deterioro cognitivo y la reducción de la transmisión colinérgica en el cerebro desempeña un papel crucial en la enfermedad de Alzheimer (Pena *et al*., 2006).

# 2.1.4. Patogénesis

La EA es una enfermedad multifactorial que involucra tanto factores genéticos como ambientales en su patogenia. Entre estos factores se encuentran principalmente la edad y los antecedentes familiares positivos de demencia (Kandimalla *et al*., 2017). Los tipos de Alzheimer se dividen en dos grandes grupos según el momento en que aparecen los primeros signos. En el caso del Alzheimer de inicio temprano, los síntomas surgen por primera vez entre los 30 y los 65 años, mientras que, en el Alzheimer de inicio tardío, los signos comienzan a manifestarse alrededor de los 65 años (NIH, 2017). Existen cuatro subtipos de la enfermedad de Alzheimer, clasificados según el gen afectado (Bird, 2008). El primer subtipo, conocido como EA1, implica mutaciones en el gen de la APP. Estas mutaciones pueden llevar a una producción excesiva del péptido beta-amiloide, lo que contribuye a la formación de placas amiloides en el cerebro (Weggen y Beher, 2012). La apolipoproteína está implicada en el subtipo 2. La apolipoproteína E (APOE) es una proteína de unión a lípidos y constituye la principal proteína de transporte de colesterol en el cerebro. Aunque su función principal es el transporte, también interactúa con la beta amiloide, los betaglucanos y los ácidos lipoteicoicos (Mahley y Rall, 2000). En los seres humanos, existen tres isoformas de APOE, conocidas como ε2, ε3 y ε4. Las personas portadoras del alelo ε2 tienen un menor riesgo de desarrollar la EA mientras que aquellas con el alelo ε4 tienen un riesgo de 10 a 15 veces mayor. La presencia del alelo ε4 se considera el factor de riesgo genético más fuerte para la EA (Troutwine *et al*., 2022). En el subtipo EA3, el gen implicado es el PSEN1, también conocido como STM2, como una de las causas del Alzheimer. Este es el subtipo de EA más común. Este gen humano codifica la proteína presenilina 1, una subunidad de la γ-secretasa. Se ha identificado que una mutación puntual en el gen PSEN1 resulta en un cambio de isoleucina por una asparagina. Estas mutaciones pueden alterar la actividad de la γ-secretasa, una enzima que participa en el procesamiento de la APP, lo que puede llevar a un aumento en la producción de beta-amiloide y a la formación de placas amiloides características de la enfermedad de Alzheimer (Levy-Lahad *et al*., 1995). El subtipo 4 de la enfermedad de Alzheimer (EA) está relacionado con mutaciones en el gen PSEN2. Al igual que las mutaciones en PSEN1, estas mutaciones también afectan la función de la γ-secretasa. Sin embargo, este subtipo es menos común que el subtipo EA3 (Rogaev *et al*., 1995). Sin embargo, la causa de la enfermedad de Alzheimer de inicio tardío sigue siendo desconocida, aunque se han identificado muchos factores de riesgo. Entre estos, la diabetes mellitus 2 se destaca como un fuerte factor de riesgo para el desarrollo de la EA (Yang y Song, 2013).

# 2.2. Diabetes mellitus

# 2.2.1. La Diabetes Mellitus

La diabetes mellitus (DM) es un grupo de enfermedades metabólicas caracterizadas por niveles elevados de glucosa en la sangre, lo que resulta de defectos en la producción o acción de la insulina. Esta se divide en dos tipos principales, la diabetes mellitus tipo 1 (DM1) y la diabetes mellitus tipo 2 (DM2). La diabetes mellitus tipo 1 se caracteriza principalmente por la destrucción autoinmune de las células beta del páncreas, las cuales son responsables de la producción de insulina, mientras que la diabetes tipo 2 se distingue principalmente por dos defectos fisiológicos: la resistencia a la insulina y la disfunción progresiva de las células beta del páncreas (Medline, 2020). Se estima que 62 millones de personas en el continente americano viven con diabetes tipo 2. Además, se calcula que este número se ha triplicado desde 1980 y podría alcanzar los 109 millones de pacientes para el año 2040. En 2020, la DM2 fue la sexta causa principal de muerte en la región, con aproximadamente 244,084 muertes atribuidas directamente a esta enfermedad (OPS, 2023b).

# 2.2.2. Fisiopatología y patogénesis

El páncreas es un órgano complejo que desempeña dos roles principales: por un lado, sus células exocrinas participan en la digestión, mientras que, por otro, sus células endocrinas regulan la homeostasis de la glucosa en el cuerpo. Dentro del páncreas, los islotes de Langerhans albergan células β especializadas en la producción de insulina, un componente que regula los niveles de glucosa en la sangre (Docherty y Russ, 2019). Después de una ingesta de alimentos, se elevan los niveles de glucosa y las células β pancreáticas responden liberando insulina en la sangre. La insulina permite la utilización de la glucosa como fuente de energía en las células. Además, promueve el almacenamiento de glucosa en forma de glucógeno en el hígado y los músculos (Gillespie, 2006).

En la diabetes tipo 1, las células β productoras de insulina en el páncreas son atacadas y destruidas por el sistema inmunológico. Este tipo de enfermedades son denominadas enfermedades autoinmunes, en las cuales el sistema inmunitario responde de manera anómala a antígenos propios, lo que resulta en daño celular y tisular (Erf, 2014). Del mismo modo que otras enfermedades autoinmunes, la diabetes tipo 1 está asociada con el antígeno leucocitario humano (HLA, por sus siglas en inglés “human leukocyte antigen”). El HLA es un conjunto de genes que codifican proteínas presentes en la superficie de las células del cuerpo, permitiendo que el sistema inmunológico las reconozca como propias y no las ataque. Se ha demostrado que el HLA está asociado con la diabetes tipo 1 y se estima que contribuye a la mitad de la predisposición genética a esta enfermedad (Speight y Pouwer, 2023). Los haplotipos del gen HLA DR4-DQ8 y DR3-DQ2 son especialmente importantes, ya que están presentes en el 90% de los casos de diabetes tipo 1 en niños. Estos alelos pueden predisponer a un individuo a una respuesta autoinmune contra las células β pancreáticas, provocando la DM1 (Devendra y Eisenbarth, 2003). Asimismo, en las personas en riesgo, se cree que los virus, los factores ambientales (incluidos los dietéticos) y otros factores estresantes pueden desencadenar la destrucción autoinmune de las células β. Sin embargo, aunque se han identificado varios factores de riesgo y se comprenden algunos de los mecanismos, las causas de la diabetes tipo 1 aún no se conocen completamente (Krischer *et al*., 2022).

En la diabetes tipo 2 los factores son más complejos. Su desarrollo se debe a una combinación de factores genéticos y ambientales, y se caracteriza principalmente por la resistencia a la insulina y la disfunción de las células beta del páncreas. Esta resistencia ocurre cuando las células del cuerpo no responden adecuadamente a la insulina, Este hecho conduce a que la glucosa se acumule en el torrente sanguíneo en lugar de emplearse como energía. Inicialmente, la resistencia a la insulina es compensada con un aumento en la secreción de insulina, lo que permite mantener los niveles de glucosa dentro de un rango normal. Sin embargo, a medida que la enfermedad progresa, las células beta del páncreas comienzan a fallar y la secreción de insulina ya no es suficiente para mantener la homeostasis de la glucosa, resultando en hiperglucemia (Henson *et al*., 2023). La resistencia a la insulina puede causar estrés en ciertos orgánulos de las células beta, como el retículo endoplásmico, debido a la sobreproducción de insulina. Este estrés activa la respuesta de proteínas desplegadas (UPR, por sus siglas en inglés “Unfolded Protein Respondes”), el cual es un mecanismo que intenta restaurar la función normal del RE. Sin embargo, si el estrés persiste, la UPR puede inducir la apoptosis en las células beta (de Boer *et al*., 2024).

Diversos factores pueden contribuir al desarrollo de la diabetes mellitus tipo 2. Entre estos se encuentran los altos niveles de ácido úrico sérico, la calidad y cantidad del sueño, el tabaquismo, la depresión, las enfermedades cardiovasculares, la dislipidemia, la hipertensión, el envejecimiento, el origen étnico, los antecedentes familiares de diabetes, la inactividad física y la obesidad (Ismail *et al*., 2021).

# 2.3. Relación Alzheimer-Diabetes Mellitus

# 2.3.1. Estudios cognitivos sobre pacientes con diabetes mellitus

En las últimas décadas, numerosos estudios han identificado vínculos comunes entre la enfermedad de Alzheimer y la diabetes mellitus tipo 2. La mayoría de estos estudios utilizan pruebas estandarizadas, como el Mini-Mental State Examination (MMSE), para evaluar el rendimiento cognitivo tanto en pacientes diabéticos como en pacientes no diabéticos. Estos estudios comparan los resultados de las pruebas para observar las diferencias en el desempeño cognitivo entre ambos grupos. Los resultados de estos estudios han revelado vínculos significativos entre la EA y la DM2. Incluso, se ha propuesto el término "diabetes tipo 3" para describir la forma específica en que la diabetes está relacionada con la EA (Kandimalla *et al*., 2017).

Desde la última década del siglo pasado, se han llevado a cabo investigaciones sobre el estado cognitivo de pacientes con diabetes mellitus tipo 2. Strachan y sus colaboradores realizaron una revisión de la literatura donde tuvieran como objetivo estudiar el estado cognitivo de pacientes con diabetes mellitus. Identificaron 19 estudios de este tipo y en 13 de estos estudios, se demostró que los individuos diabéticos tenían un desempeño más pobre en al menos un aspecto de la función cognitiva, siendo la memoria verbal la capacidad más afectada (Strachan *et al*., 1997). En otro estudio de revisión de la literatura, se identificaron un total de 33 estudios. Se comparó el grupo de pacientes con diabetes mellitus con un grupo de control de sujetos no diabéticos, y se encontró que el grupo diabético mostró un rendimiento significativamente menor en varias áreas cognitivas: inteligencia, velocidad de procesamiento de la información, eficiencia psicomotora, atención visual y atención sostenida, flexibilidad cognitiva y percepción visual. Por lo tanto, en pacientes con diabetes mellitus, se observa una disminución en la velocidad mental y la flexibilidad cognitiva, con efectos que pueden variar de leves a moderados (Brands *et al*., 2005). Asimismo, en otro estudio se buscaron informes publicados de estudios longitudinales que describieran la evaluación de la función cognitiva en pacientes con diabetes mellitus. Se identificaron 25 estudios en los que se aplicaron las pruebas Mini-Mental State Examination y Digit Symbol Substitution Test. Los resultados indicaron que los individuos con diabetes mellitus experimentaron un deterioro de la función cognitiva entre 1.2 y 1.5 veces mayor que aquellos sin diabetes. Por lo tanto, se concluye que, en comparación con las personas sin diabetes, las personas con diabetes tienen una mayor tasa de deterioro de la función cognitiva y un mayor riesgo de desarrollar deterioro cognitivo (Cukierman *et al*., 2005).

# 2.3.2. Vínculos moleculares entre EA y DM2

Asimismo, diversos estudios han intentado explicar la relación molceular entre la EA y la DM2. Uno de los vínculos fisiopatológicos que relacionan la EA con DM2 es la resistencia a la insulina con la β-amiloide. Esta implica una serie de interacciones entre la señalización de insulina y los depósitos de β-amiloide. La resistencia a la insulina y los niveles reducidos del factor de crecimiento similar a la insulina (IGF-1, por sus siglas en inglés “Insuline Grow Factor”) pueden contribuir a la acumulación de β-amiloide en el cerebro. Esta acumulación, a su vez, puede antagonizar la unión de la insulina y el receptor IGF-1, desencadenando una cascada de eventos que incluyen la secreción de agentes inflamatorios y la aparición de resistencia a la insulina. Este ciclo de retroalimentación patológica conduce tanto a la resistencia a la insulina como la neurotoxicidad asociada a la β-amiloide (Mullins *et al*., 2017). Del mismo modo, la desregulación de la señalización de la insulina puede afectar la función de la proteína precursora amiloide, lo que conduce a la acumulación de β-amiloide en las células.(Tumminia *et al*., 2018). Además, se ha demostrado que la insulina y el receptor de insulina (IR, por sus siglas en inglés “Insuline Receptor”) desempeñan un papel crucial en el aprendizaje asociativo. De hecho, la interrupción de la producción de insulina y de la actividad del RI provoca déficits en el aprendizaje asociativo. En pacientes con EA se han observado niveles anormales de insulina y del RI, y varios estudios han demostrado que la administración de insulina en estos pacientes mejora significativamente su rendimiento cognitivo (Zhao y Alkon, 2001). Por otro lado, la enzima degradadora de insulina (IDE, por sus siglas en inglés "Insulin Degrading Enzyme") es la principal peptidasa responsable de la degradación tanto de la insulina como de la β-amiloide. El papel fundamental de la IDE en la resistencia a la insulina está relacionado con su capacidad para regular los niveles de insulina (Kurochkin *et al*., 2018). La IDE degrada la insulina, lo que ayuda a mantener su equilibrio en el cuerpo. En situaciones de resistencia a la insulina, los niveles elevados de insulina pueden sobrecargar la capacidad de la IDE para degradarla, lo que conduce a un exceso de insulina circulante. Además, la IDE también degrada la β-amiloide, una proteína que se acumula en el cerebro de los pacientes con Alzheimer. Cuando la IDE está ocupada degradando cantidades excesivas de insulina, su capacidad para descomponer la β-amiloide puede verse comprometida (De Felice, 2013). Finalmente, estudios en ratones han demostrado que la resistencia a la insulina incrementa la actividad de la γ-secretasa, lo que resulta en un aumento de la producción de β-amiloide en el cerebro. Además, se observa una disminución en la actividad de la enzima degradadora de insulina, lo que contribuye a la acumulación de β-amiloide (Starks *et al*., 2015).

Se han identificado vínculos moleculares entre la resistencia a la insulina y la proteína tau. La proteína tau no solo se expresa en el cerebro, sino también en otros órganos, como el endotelio pulmonar. Además, estudios han demostrado que tanto la proteína tau como la β-amiloide se expresan en el páncreas. Un análisis del tejido pancreático realizado en 21 casos de autopsias de pacientes con DM2 confirmó la presencia de estas proteínas (Balczon *et al*., 2024). Una investigación reveló la importancia de la tau pancreática en la regulación de la secreción de insulina y la homeostasis de la glucosa. Se observó que los niveles de tau estaban elevados en las células de los islotes β de pacientes con DM2, y la pérdida de tau resultó en un aumento en la secreción de insulina en líneas celulares, drosófila y ratones (Balczon *et al*., 2024). La tasa de agregación de la proteína tau está fuertemente influenciada por su proceso de fosforilación. De acuerdo con investigaciones realizadas en humanos y animales, se ha observado que la enzima fosfatasa PP2A juega un papel crucial en este proceso. Esta enzima está implicada tanto en la fosforilación de la tau en enfermedades como la EA como en la DM2, Además, se ha descubierto que la expresión de la fosfatasa PP2A disminuye cuando se administra insulina al paciente, lo que sugiere una conexión entre la regulación de la glucosa y la fosforilación de la tau (Clodfelder-Miller *et al*., 2006).

Se cree que la neuroinflamación desempeña un papel crucial como vínculo entre la EA y la DM2. Esta inflamación surge en las etapas iniciales de la EA y contribuye a su patología mediante el daño oxidativo, la hiperfosforilación de la proteína tau, la acumulación de β-amiloide y la disfunción del sistema colinérgico. (Kandimalla *et al*., 2017). La neuroinflamación es reconocida por numerosos estudios como una de las principales causas de la resistencia a la insulina y al IGF-1 observadas en el cerebro de pacientes con EA. La resistencia a la insulina inducida por DM2 y la obesidad genera altos niveles de macromoléculas lipídicas citotóxicas, las cuales, al atravesar la barrera hematoencefálica, desencadenan inflamación y resistencia a la insulina en el cerebro. Un proceso fundamental en esta neuroinflamación es la liberación de diversas células microgliales por el β-amiloide, lo que conduce a la secreción de citoquinas inflamatorias como la interleucina-6. Se destaca que ambos fenómenos median la relación entre la DM2 y la EA (de la Monte, 2009). Mientras que por otro lado se ha demostrado que estas moléculas también provocan una disminución en la función sináptica y otros efectos, lo que respalda la teoría de que la DM2 posiblemente actúa como precursora al activar sustancias inflamatorias que conducen a la neuroinflamación y, en última instancia, a la EA (Rosenberg, 2005).

Otro proceso que establece un vínculo entre la EA y la DM2 es el estrés oxidativo. Esta se refiere a un desequilibrio entre la producción de especies reactivas de oxígeno y la defensa antioxidante (Sies, 2019). Se ha demostrado que, en condiciones de resistencia a la insulina, se observa un aumento de las reacciones oxidativas en el plasma sanguíneo (Henriksen *et al*., 2011). Esto es perjudicial para el cerebro humano, ya que este es particularmente vulnerable al daño por el estrés oxidativo (Cobley *et al*., 2018). Asimismo, algunos estudios han propuesto que el estrés oxidativo y la resistencia a la insulina pueden conducir a la acumulación de la proteína β-amiloide y tau, lo que implica que haya una fuerte correlación con la patología de la EA (Gu *et al*., 2008).

# 2.4. Diseño de fármacos

# 2.4.1. El diseño de fármacos

El diseño de fármacos es el proceso de crear moléculas que tengan propiedades específicas para interactuar con objetivos biológicos, como proteínas o enzimas, con el fin de modular su actividad y producir efectos terapéuticos deseados. El proceso de descubrimiento y desarrollo de fármacos está compuesto de varias etapas. Estas comprenden desde la identificación del objetivo farmacológico hasta la optimización de compuestos líderes, pasando por investigaciones preclínicas y ensayos clínicos y finalizando con la puesta en el mercado del nuevo fármaco (Salazar y Gormley, 2017).

La etapa inicial del desarrollo de un fármaco, conocida como el descubrimiento de fármacos, se centra en la búsqueda de nuevos compuestos con propiedades químicas específicas destinadas al tratamiento de enfermedades (Carracedo-Reboredo *et al*., 2021). Este proceso inicia con una investigación sobre las causas de una enfermedad, en algunos casos conduciendo a la identificación de una o varias dianas moleculares asociadas con dicha enfermedad. Posteriormente, se procede a la identificación de compuestos activos que interactúan con la diana molecular y a la optimización de su actividad biológica. La evaluación de estos compuestos activos comprende diversos pasos experimentales, que incluyen ensayos en líneas celulares, estudios en modelos animales, y ensayos clínicos en humanos (FDA, 2019). Finalmente, los compuestos que pasan satisfactoriamente por todas las etapas son aprobadas por un agente regulatorio (Salazar y Gormley, 2017).

Dado el alto costo económico y de tiempo requeridos, junto con la baja eficacia intrínseca de los métodos convencionales, desde finales del siglo pasado se han desarrollado herramientas computacionales para el diseño de fármacos. A los enfoques que emplean estas herramientas computacionales se conocen hoy en día como diseño de fármacos asistido por computadora (CADD, por sus siglas en inglés “Computer, Aided Drug Design”). Estos métodos permiten un descubrimiento de fármacos más preciso, sino que también se destacan por su eficiencia mejorada, constituyendo así una alternativa más efectiva en el proceso de desarrollo de fármacos (Coumar, 2021).

El diseño de fármacos asistido por computadora abarca una diversidad de enfoques teóricos y computacionales que constituyen elementos esenciales en el proceso contemporáneo de descubrimiento de fármacos (Prieto-Martínez *et al*., 2019). El CADD puede ser dividido en dos enfoques de acuerdo con la disponibilidad de la estructura de la proteína diana: diseño de fármacos basado en la estructura (SBDD, por sus siglas en inglés “Structure Based Drug Design”) y diseño de fármacos basado en el ligando (LBDD, por sus siglas en inglés “Ligand Based Drug Design”) (Coumar, 2021).

# 2.4.2. La inteligencia artificial en el diseño de fármacos

En la última década, la inteligencia artificial (IA) ha experimentado un significativo avance que la ha colocado como el centro de atención de la sociedad. Su rápido desarrollo ha despertado el interés en diversos sectores, convirtiéndola en un tema común de conversación a nivel global y en un nuevo recurso cotidiano para muchas personas. La IA nace en los años 1950 como un “esfuerzo de automatizar tareas normalmente realizadas por humanos” (Chollet, 2018). En la actualidad, la IA es definida como el estudio y la creación de máquinas inteligentes que se asemejan a los humanos, o incluso que superen su inteligencia (Singh y Khanna, 2023). El campo de la IA abarca el Machine Learning (ML) y el Machine Learning a su vez tiene un subcampo llamado Deep Learning (DL) (Batool *et al*., 2019).

Las tecnologías que integran el aprendizaje automático y la IA se han convertido en herramientas versátiles que pueden ser aplicadas de manera generalizada en diversas etapas del proceso de desarrollo de fármacos. Esto incluye la identificación y validación de dianas farmacológicas, el diseño de nuevos medicamentos, la reutilización de fármacos existentes, la mejora de la eficiencia en la investigación y desarrollo, la recopilación y análisis de información biomédica, así como la optimización del proceso de toma de decisiones para la selección de pacientes en ensayos clínicos (Huang *et al*., 2017). Además, se han desarrollado métodos basados en la IA para predecir la estructura tridimensional de una proteína muy prometedores que mejoran la calidad estructural acercándose a la precisión experimental (Bordin *et al*., 2023). Algunos ejemplos de ello son el software AlphaFold (Jumper *et al*., 2021), RoseTTafold (Baek *et al*., 2021) y CollabFold (Mirdita *et al*., 2022).

# 2.4.3 Machine Learning

El Machine Learning (ML) representa un paradigma nuevo en comparación con la programación clásica. En esta última, se proporcionan una serie de reglas y datos a un programa, y las respuestas se generan conforme a esas instrucciones, esto se ilustra en la Figura 2. En cambio, en el Machine Learning, se presentan tanto los datos como las respuestas al programa, con la expectativa de que aprenda "reglas" que luego puedan aplicarse a nuevos datos para obtener resultados (Chollet, 2018).

Diagrama

Descripción generada automáticamente

**Figura 2.** Nuevo paradigma del Machine Learning, diagrama adaptado de Chollet, 2018.

El aprendizaje automático se puede categorizar de acuerdo con la inclusión o no de la variable dependiente (y) en los datos utilizados para entrenar el modelo como se ilustra en la Figura 3. Cuando el modelo es entrenado con datos que contienen la variable dependiente (y), se le llama aprendizaje supervisado; mientras que, si carecen de ellos, se le conoce como aprendizaje no supervisado. Además, ambas categorías pueden subdividirse según la naturaleza de los datos, ya sean continuos o categóricos (Batool *et al*., 2019). El aprendizaje supervisado se emplea principalmente para tareas de regresión y clasificación, permitiendo la predicción de propiedades farmacocinéticas y toxicidad (Pillai *et al*., 2022), mientras que el aprendizaje no supervisado, en el contexto del cribado de fármacos, se puede emplear el aprendizaje no supervisado para analizar la similitud de los compuestos a través de métodos de agrupamiento, lo que ayuda a agrupar los compuestos según sus similitudes (Voicu *et al*., 2020).

Diagrama

Descripción generada automáticamente

**Figura 3.** Clasificación de los modelos de Machine learning.

# 2.4.4. Machine Learning Supervisado

En el Machine Learning supervisado, los datos deben ir acompañados de su respectivo resultado y. Un modelo de ML supervisado tiene como objetivo determinar una función f(x) = y, a partir de una lista de pares de entrenamiento (x1, y1), (x2, y2), …, (xn, yn) (Raghunathan y Priyakumar, 2022). El ML supervisado se subdivide según la naturaleza de los datos que procesa, ya sea continua o categórica. Para datos continuos, se emplean algoritmos de regresión, mientras que, para datos categóricos, se recurre a algoritmos de clasificación. Aunque estas son las clasificaciones principales, hay otros algoritmos que no se ajustan solo a una de estas categorías y pueden aplicarse para ambos tipos de datos. Los modelos de Machine Learning supervisado más comúnmente utilizados en el descubrimiento de nuevos fármacos, tanto para clasificación como para regresión, incluyen la regresión lineal, la regresión logística, los árboles de decisión (DT, por sus siglas en inglés “Decision Trees”), los bosques aleatorios (conocidos en inglés como “random forests”), las máquinas de soporte vectorial (SVM, por sus siglas en inglés, “Support Vector Machines”), el algoritmo de vecinos más cercanos (k-NN, por sus siglas en inglés “k-Nearest Neighboors), XGBoost y el clasificador Naive Bayes (Schneider *et al*., 2019).

# 2.4.5. Machine learning no supervisado

Los enfoques de Machine Learning no supervisado buscan identificar patrones en conjuntos de datos que carecen de etiquetas y variables de salida (Raghunathan y Priyakumar, 2022). Al igual que en el ML supervisado, este tipo de enfoque puede clasificarse según la naturaleza de los datos: categóricos o continuos (Schneider y Xhafa, 2022). Los algoritmos de ML no supervisado más utilizados en el descubrimiento de fármacos incluyen K-means clustering, clustering jerárquico, DBSCAN (por sus siglas en inglés “Density-Based Spatial Clustering of Applications with Noise”), análisis de componentes principales (PCA, por sus siglas en inglés “Principal Component Analysis”) y t-SNE (por sus siglas en inglés “Stochastic Neighbor Embedding”). Estos algoritmos permiten reducir la dimensionalidad de los datos, identificar patrones y seleccionar candidatos prometedores para pruebas adicionales. Aplicaciones como estas mejoran la eficiencia en la construcción de modelos QSAR y en la identificación de nuevas moléculas con potencial terapéutico.

# 2.4.6 Deep Learning

El Deep Learning constituye un subcampo del Machine Learning que transforma el paradigma de adquisición de conocimientos. Mientras que en el Machine Learning se busca obtener una representación de los datos, el Deep Learning va más allá. En este nuevo paradigma, se obtienen capas sucesivas de representaciones para los datos. La profundidad en el Deep Learning no implica una comprensión más exhaustiva de los "datos", sino que se refiere a la presencia de múltiples capas "profundas" de representaciones. Por ende, una denominación más apropiada para el Deep Learning sería "aprendizaje de representaciones en capas" (Chollet, 2018).

Los algoritmos de Deep Learning emplean redes neuronales artificiales para lograr su cometido. Las redes neuronales artificiales (ANN, por sus siglas en inglés “Artificial Neural Network”) son algoritmos que reciben múltiples señales de entrada, procesan la información calculando una suma ponderada de estas entradas mediante una función de activación no lineal, y luego transmiten la señal resultante a las neuronas conectadas en la siguiente capa hasta finalmente dar una respuesta (van Gerven y Bohte, 2017). Si bien, se les llama redes neuronales, no hay prueba alguna de que el cerebro funcione de esta manera (Chollet, 2018).

Las redes neuronales se categorizan según su arquitectura y las conexiones internas que poseen. Entre los ejemplos de estas redes se encuentran las redes neuronales recurrentes y las redes neuronales convolucionales (Paul *et al*., 2021). Las redes neuronales recurrentes son una red neuronal artificial especializada diseñada específicamente para procesar datos de series temporales o información que incluye secuencias de datos. Estas incorporan el concepto de "memoria", lo que les permite retener estados o información de las entradas previas para construir la salida subsiguiente de la secuencia. Las redes neuronales convolucionales se emplean para analizar imágenes. Utiliza una técnica llamada convolución, la cual es una operación matemática entre dos funciones que produce una tercera función, describiendo cómo la forma de una afecta la forma de la otra (Batool *et al*., 2019).

# 2.5. Diseño de fármacos para la enfermedad de Alzheimer

Con los avances tecnológicos de las últimas décadas, ha aumentado la cantidad de datos farmacéuticos disponibles que pueden ser utilizados por la IA para el diseño de fármacos. La aplicación del Machine Learning en el diseño de fármacos se está generalizando rápidamente en la última década. El Machine Learning permite extraer características estructurales moleculares, analizar interacciones fármaco-diana y establecer vínculos entre fármacos, enfermedades y proteínas. Este enfoque facilita el desarrollo de sistemas de predicción altamente precisos en diversas etapas del proceso, minimizando la incertidumbre en cada fase de la investigación (Geng *et al*., 2024). Gracias a los avances en los algoritmos de IA, las soluciones impulsadas por estos tienen un potencial sin precedentes para acelerar el proceso de descubrimiento de fármacos, mejorando significativamente la tasa de éxito (Vatansever *et al*., 2021), siendo este uno de los principales problemas en el desarrollo de fármacos para EA, pues alta tasa de deserción de los fármacos en desarrollo. Durante el período 2002-2012, se encontró una tasa de éxito general de solo el 0.4%. Esta alta tasa de deserción es una de las razones más importantes para enfocar el desarrollo de fármacos en este ámbito (Cummings *et al*., 2014).

El principal reto en el desarrollo de nuevos fármacos yace en caracterizar las interacciones fármaco-diana y luego aplicar estas caracterizaciones para evaluar la probabilidad de interacción entre una molécula de fármaco y una diana específica (Ye *et al*., 2021). Las investigaciones actuales en el diseño de fármacos para la EA abordan diversos aspectos, como la cantidad y tipo de dianas terapéuticas, con estudios que se enfocan tanto en dianas únicas como múltiples. Se utilizan diferentes metodologías para identificar nuevos inhibidores potenciales, siendo los modelos QSAR (Relación Cuantitativa Estructura-Actividad, por sus siglas en inglés “Quantitative Structure-Activity Relationship”) una herramienta común (Roy *et al*., 2015). Estos modelos determinan la relación entre la estructura de una molécula y su actividad biológica o reactividad química. Algunos investigadores utilizan algoritmos de Machine Learning para desarrollar modelos QSAR, mientras que otros emplean métodos estadísticos convencionales. Las dianas terapéuticas comunes en las investigaciones de descubrimiento de fármacos contemporáneos para la EA incluyen principalmente inhibidores de la acetilcolinesterasa, que aumentan los niveles de acetilcolina en el cerebro, mejorando la transmisión sináptica. También se investigan inhibidores dirigidos contra la beta-amiloide y antagonistas del receptor *N*-metil-D-aspartato, que modulan la actividad del glutamato al antagonizar estos receptores, aliviando así los síntomas de la enfermedad (Parvathy Dharshini *et al*., 2022).

# 2.6.1. Fármacos de EA para una sola diana

Existen diversos estudios enfocados en descubrir nuevos fármacos para el Alzheimer EA centrados en una sola diana terapéutica. En 2022, Mukerjee y su equipo realizaron un estudio para identificar posibles fármacos dirigidos contra Keap1, una proteína implicada en el estrés oxidativo relacionado con la enfermedad de Alzheimer. Descargaron datos experimentales de inhibición de 239 compuestos y calcularon descriptores moleculares utilizando el software PaDeL, así como huellas dactilares de PubChem. Entrenaron modelos de aprendizaje automático supervisado a partir de estas huellas dactilares y los valores de IC50 obtenidos de los datos de actividad de ChEMBL. La concentración de IC50 mide la concentración necesaria para inhibir un proceso biológico en un 50% (Ugwu y Conradie, 2023). Evaluaron un total de 41 modelos de aprendizaje automático, identificando al modelo Random Forest como el mejor, con un valor de R² cercano a 1 y un RMSD cercano a 0. Posteriormente, utilizaron este modelo Random Forest para realizar un cribado virtual con moléculas descargadas de la base de datos DrugBank, identificando cinco inhibidores potenciales: DB06841, DB04310, DB11784, DB12730 y DB12677, con valores de IC50 de 81 nM, 125 nM, 158 nM, 170 nM y 175 nM, respectivamente. Realizaron un análisis de acoplamiento molecular, determinando que el mejor puntaje fue de –8.941 kcal/mol para DB12677. Finalmente, llevaron a cabo análisis de dinámica molecular, donde DB06841 mostró el valor de RMSD más bajo a los 20 ns. Con base en estos resultados, concluyeron que DB06841 es el inhibidor más prometedor para Keap1 (Mukerjee *et al*., 2022). En 2022, Lee y colaboradores se centraron en desarrollar varios modelos de machine learning para identificar inhibidores de S100A9, una proteína relacionada con la enfermedad de Alzheimer debido a su papel en la inflamación y el estrés oxidativo. Obtuvieron datos experimentales de inhibidores a partir de tres patentes y calcularon descriptores moleculares 2D utilizando PaDeL. Consideraron cada bit de las huellas dactilares como un descriptor molecular independiente. Entrenaron y optimizaron varios modelos, incluyendo Random Forest, Decision Tree y Naive Bayes, determinando que el mejor modelo fue Random Forest, con un valor de AUCROC de 1. Utilizando este modelo, realizaron un cribado virtual de moléculas de la base de datos eMolecules. Tras evaluar aproximadamente 6 millones de moléculas, identificaron 46 posibles inhibidores de S100A9 (Lee *et al*., 2019). En ese mismo año, Cavas y sus colaboradores se centraron en la síntesis y evaluación de nuevos derivados de oxazolona con estructuras de carbazol como inhibidores de la acetilcolinesterasa, una enzima clave en enfermedades neurodegenerativas como el Alzheimer. Tres compuestos específicos fueron sintetizados y probados por su capacidad de inhibir AChE in vitro. Los resultados experimentales se modelaron utilizando redes neuronales artificiales, mostrando una alta correlación entre los datos experimentales y los predichos. Este estudio sugiere que estos nuevos inhibidores de oxazolona con carbazol tienen un potencial significativo y destaca el uso de ANN como una herramienta efectiva para modelar la inhibición enzimática, recomendando futuras investigaciones in vivo para confirmar estos hallazgos preliminares (Cavas *et al*., 2019).

# 2.6.2. Fármacos de EA para modular múltiples dianas

Debido a la naturaleza multifactorial del Alzheimer, los tratamientos con fármacos dirigidos a una sola diana pueden no ser exitosos. En 2015, Lovestone y colaboradores realizaron un ensayo de fase II para evaluar la efectividad de tideglusib en EA. Este fármaco es un inhibidor de la glucógeno sintasa cinasa 3 (GSK-3). Se administró tideglusib o un placebo a grupos de pacientes con EA, evaluando criterios como la atrofia cerebral en RM y los niveles de tau, amiloide-β y BACE1 en el líquido cefalorraquídeo. Los resultados no mostraron diferencias significativas entre los grupos, sugiriendo que centrarse en una sola diana terapéutica podría haber sido una limitación (Lovestone *et al*., 2015).

Debido a la complejidad y los múltiples mecanismos implicados en la EA, es crucial descubrir fármacos multidiana capaces de abordar las diversas vías involucradas en su desarrollo (Vatansever *et al*., 2021). Un fármaco multidiana es aquél que puede actuar en el tratamiento de enfermedades complejas mediante la modulación de múltiples dianas (Lavecchia y Cerchia, 2016). Un estudio señala que debe enfatizarse que el desarrollo y la aparición de la EA es el resultado de los efectos colectivos de múltiples factores, no de un único factor específico. Este hecho justifica la necesidad de dirigir los tratamientos hacia múltiples mecanismos y vías simultáneamente. Probablemente, esta es la razón por la cual algunos de los fármacos desarrollados basándose en una sola diana terapéutica, no han tenido éxito (Gong *et al*., 2018). El desarrollo de fármacos multidiana ha ganado mucha aceptación en las últimas décadas y ha sido aplicado principalmente en el cáncer. En el año 2019, se desarrollaron fármacos multidiana para tratar el cáncer de pulmón utilizando un enfoque de farmacología de sistemas que incluyó cribado fenotípico, química, fosfoproteómica y secuenciación de ARN. Se seleccionaron los inhibidores de multiquinasa foretinib y cabozantinib, identificando sus dianas moleculares y mecanismos de acción. La validación mediante estudios bioquímicos y celulares confirmó estas dianas. Finalmente, desarrollaron una combinación sinérgica de foretinib con barasertib, produciendo un inhibidor más potente de AURKB, para el cáncer de pulmón de células pequeñas amplificado por MYC (Sumi *et al*., 2019).

Se han realizado múltiples estudios para descubrir inhibidores multidiana que puedan tratar la enfermedad de Alzheimer. Algunos emplean técnicas de aprendizaje automático, mientras que otros utilizan enfoques más convencionales. En 2023, Kumar y sus colaboradores desarrollaron modelos QSAR multidiana para identificar nuevos inhibidores contra la enfermedad de Alzheimer (EA). Trabajaron con diversas dianas terapéuticas, incluyendo AChE, BuChE, BACE1, β-amiloide, 5-HT6, CDK-5, gamma-secretasa, Glutaminil Cicloasa, GSK-3β, MAO-B, NMDA y PDE10A, obteniendo datos de actividad molecular contra estas dianas. Utilizando regresión por mínimos cuadrados parciales y regresión lineal múltiple, desarrollaron modelos QSAR para cada diana. Posteriormente, aplicaron estos modelos para predecir la actividad inhibitoria de aproximadamente 19 millones de compuestos de bases de datos químicas como ZINC12, Asinex, NCI e InterBioscreen. Identificaron varios compuestos con actividad dual inhibitoria, incluyendo combinaciones como AChE y BACE1, AChE y β-amiloide, y AChE y BuChE. Además, implementaron análisis de docking molecular para validar las interacciones entre los compuestos activos y las dianas, observando que los compuestos formaban enlaces de hidrógeno y interacciones hidrofóbicas con los residuos aminoácidos activos, corroborando las características observadas en los modelos QSAR. Los resultados de esta investigación proporcionan herramientas valiosas para guiar a los investigadores en la predicción de la actividad inhibitoria de nuevos compuestos y en el diseño de tratamientos de próxima generación para la EA (Kumar *et al*., 2023). En el año 2020, Concu y sus colaboradores desarrollaron un modelo multidiana para predecir la actividad de los fármacos monooxidasa A y B. La actividad de las enzimas monooxidasa están relacionadas la EA. Para llevar a cabo el modelo, crearon un modelo QSAR basado en análisis discriminante lineal que se desarrolló en base a 5759 compuestos de una base de datos. El modelo resultante fue capaz de clasificar correctamente a todos esos compuestos por lo que se considera un modelo robusto y reproducible (Concu *et al*., 2020). Aunque en los estudios previos se crearon modelos QSAR tradicionales, los modelos QSAR basados en aprendizaje automático se consideran superiores debido a su mayor precisión y robustez. Un estudio demostró que los modelos de aprendizaje automático son más precisos y robustos en comparación con los modelos QSAR tradicionales, lo que los hace más efectivos para predecir la actividad biológica de compuestos en el descubrimiento de fármacos (Bosc *et al*., 2019). Dhamodharan y Mohan desarrollaron modelos de aprendizaje automático para predecir la actividad de inhibidores duales de AChE y BACE1 para el tratamiento de la EA Utilizaron técnicas de machine learning, incluyendo máquinas de vectores de soporte (SVM) y redes neuronales artificiales (ANN), para cada proteína diana. Los mejores modelos mostraron alta precisión predictiva en validaciones internas y externas, con un coeficiente de determinación (R²) de 0.8688 para AChE y 0.8177 para BACE1 (Dhamodharan y Mohan, 2022). En un estudio llevado a cabo en 2023, se diseñaron fragmentos químicos para descubrir inhibidores duales de AChE y BACE1. Utilizaron datos de 3,524 compuestos con mediciones de AChE y BACE1 obtenidos de la base de datos ChEMBL para entrenar árboles de decisión y construir modelos QSAR. Los modelos mostraron buenos rendimientos, con precisiones globales de entrenamiento/validación externa de 0.85/0.80 para AChE y 0.83/0.81 para BACE1. A partir de estos árboles, se identificaron reglas para descubrir inhibidores duales, resultando en la predicción de más de 250 nuevos inhibidores basados en fragmentos activos. Los compuestos predichos fueron luego validados mediante simulaciones de docking molecular, revelando que los fragmentos activos formaban interacciones estables y específicas con los residuos clave en los sitios activos de AChE y BACE1​ (Bao *et al*., 2023).

Actualmente, no se han realizado investigaciones sobre el descubrimiento de nuevos fármacos para la EA. La mayoría de estas investigaciones se han centrado en evaluar los efectos de los fármacos antidiabéticos en la EA, aprovechando la interrelación entre ambas enfermedades para identificar tratamientos potenciales. En una revisión sobre la aplicación de fármacos antidiabéticos en el tratamiento de la EA realizada por Michaidilis, se determinó que los investigadores están probando medicamentos como la insulina intranasal, la metformina, las incretinas y las tiazolidinedionas. Varios estudios han concluido que estos medicamentos pueden mejorar la cognición y la memoria en pacientes con EA (Michailidis *et al*., 2022). Una investigación reciente exploró los vínculos moleculares entre la EA y la DM2 utilizando redes de interacción proteína-proteína. En este estudio, se identificaron diez proteínas clave que podrían servir como biomarcadores y objetivos terapéuticos para ambas enfermedades. Estas proteínas destacan por su potencial en el desarrollo de tratamientos combinados, ya que abordan los mecanismos subyacentes compartidos entre EA y DM2. Este enfoque podría conducir a terapias más efectivas y específicas, mejorando los resultados clínicos para pacientes con ambas condiciones. Las proteínas propuestas fueron STAT3, EGFR, IRS1, MAPK1, SRC, HSP90AA1, PIK3R1, UBC, MAPK3 y ESR1 (Castillo-Velázquez *et al*., 2023). En la presente investigación, se propone utilizar estas proteínas como dianas para estudiar la asociación entre la enfermedad de Alzheimer y la diabetes mellitus tipo 2. En la Tabla 1 se muestra una breve descripción de cada una de estas proteínas, así como su posible relación con la EA y con la DM2.

**Cuadro 1.** Tabla que muestra la elación entre las proteínas diana con la EA y la DM2.

|  |  |  |  |
| --- | --- | --- | --- |
| PROTEÍNA | DESCRIPCIÓN | RELACIÓN CON LA EA | RELACIÓN CON LA DM2 |
| STAT3 | Proteína de la familia de las STAT involucrada en la señalización celular de procesos inflamatorios y de crecimiento celular (H. Q. Wang *et al*., 2022). | Implicada en la neuroinflamación y muerte neuronal inducida por beta-amiloide. Activada por la acumulación de Aβ, contribuye a la inflamación crónica en el cerebro (Mehla *et al*., 2021). | También implicada en inflamación, sensibilización a la insulina y reprogramación celular. Wakabayashi demostró que inhibir STAT3 facilita la reprogramación de células alfa a beta, crucial en la DM2 donde las células beta están destruidas (WAKABAYASHI *et al*., 2019) |
| EGFR | Es una proteína quinasa receptora de factores de crecimiento epidérmico que, al unirse a su ligando, induce dimerización, autofosforilación en tirosina y señalización para la proliferación celular (Gan *et al*., 2013). | Choi y colaboradores demostraron que los inhibidores de EGFR atenúan la patología beta-amiloide y mejoran la función cognitiva en modelos de EA, reduciendo la neuroinflamación y el estrés oxidativo (Choi *et al*., 2023). | En la DM2, la activación transitoria del receptor de crecimiento epidérmico es esencial para la recuperación renal tras una lesión aguda, pero su activación persistente contribuye a enfermedades renales crónicas, incluida la enfermedad renal diabética, sugiriendo que dirigir al EGFR podría tener potencial terapéutico (Sheng *et al*., 2020). |
| IRS1 | Es una proteína que actúa como intermediaria en la transducción de señales desde el receptor de insulina a diversas señales celulares (H. Q. Wang *et al*., 2022). | Inhibir la ISR1 es arriesgado porque su disfunción se asocia con la EA. La resistencia a la insulina en el cerebro, provocada por inhibir la ISR1, puede promover la acumulación de beta-amiloide (Arnold *et al*., 2018). | La ISR1 es fundamental en la señalización del receptor de insulina, y su disfunción puede causar resistencia a la insulina (Gabbouj *et al*., 2019). |
| SRC | Es una proteína de la familia de las tirosino quinasas no receptoras, vinculada a procesos celulares como crecimiento, diferenciación y supervivencia (Uniprot, 2024). | La proteína SRC está implicada en la fosforilación anormal de proteínas tau, formando ovillos neurofibrilares en la EA. Esta hiperfosforilación es clave en la patogénesis de la EA, y SRC podría contribuir a este proceso (Arnold *et al*., 2018). | SRC está implicado en la señalización de insulina. Su activación anormal puede interferir con esta señalización, contribuyendo a la resistencia a la insulina, un factor clave en la patogénesis de la DM2 (Greeny *et al*., 2024). |
| MAPK1 | Las proteínas MAPK son quinasas serina/treonina esenciales en la transducción de señales intracelulares, involucradas en crecimiento, diferenciación, respuesta al estrés, apoptosis y funciones metabólicas (Cargnello y Roux, 2011). | La señalización aberrante de ERK2 puede causar la hiperfosforilación de proteínas tau, formando ovillos neurofibrilares, una característica de la EA (Miao *et al*., 2019). | Un estudio demostró que la desregulación de la señalización de MAPK contribuye a la disfunción endotelial progresiva en la DM2 (Huang *et al*., 2012). |
| PIK3R1 | Es esencial para regular la señalización celular a través de la vía PI3K/AKT, controlando el crecimiento, supervivencia celular, metabolismo y respuesta inmunológica. Las mutaciones en PIK3R1 están implicadas en enfermedades como el cáncer y la diabetes (NIH gene, 2024). | La disfunción de esta vía puede contribuir a la acumulación de beta-amiloide y la fosforilación de tau, características patológicas clave del Alzheimer, afectando la función cognitiva y la memoria (Cho, 2024). | PIK3R1 es crucial para la señalización de insulina en la vía PI3K/AKT, regulando la captación de glucosa en las células. Mutaciones o disfunción de PIK3R1 pueden causar resistencia a la insulina, una característica principal de la DM2 Boukhalfa *et al*., 2023). |
| HSP90AA1 | Es un gen que codifica una isoforma de Hsp90, una chaperona molecular esencial en el plegamiento, estabilidad y función de muchas proteínas. Hsp90 es crucial en situaciones de estrés celular, como temperaturas elevadas y desnaturalización de proteínas (NIH gene, 2024b). | Estabiliza y pliega proteínas como tau y beta-amiloide, esenciales en la patología del Alzheimer, e impide la agregación patológica de tau, manteniendo la homeostasis celular y la función neuronal (Liu *et al*., 2024). | Regula la señalización de la insulina, siendo necesaria para la funcionalidad de la proteína quinasa en la captación de glucosa. Su disfunción puede contribuir a la resistencia a la insulina, característica principal de la DM2 (Ding *et al*., 2022). |
| ESR1 | Codifica el receptor de estrógeno alfa, una proteína nuclear que regula la expresión de genes inducibles por estrógenos, implicados en crecimiento, metabolismo, desarrollo sexual, gestación y funciones reproductivas, y se expresa en tejidos no reproductivos (NIH gene, 2024a). | Un estudio demostró que polimorfismos en el gen ESR1, como PvuII y XbaI, aumentan el riesgo de Alzheimer. Estos polimorfismos pueden reducir la neuroprotección del receptor de estrógeno alfa, causando neuroinflamación y otros mecanismos patológicos (Ma *et al*., 2009). | Un estudio encontró que estos polimorfismos están significativamente asociados con la resistencia a la insulina y la diabetes tipo 2 en mujeres palestinas, sugiriendo que ESR1 influye en la susceptibilidad y patogénesis de la DM2 (Gallagher *et al*., 2007). |
| UBC | Este gen codifica la proteína ubiquitina C, que regula procesos celulares mediante la modificación postraduccional de otras proteínas. La ubiquitinación está asociada con la degradación de proteínas, reparación del ADN, ciclo celular, modificación de la quinasa y endocitosis (NIH gene, 2024d). | Aunque no hay estudios directos sobre UBC y EA, UBC es central en la ubiquitinación, reparación del ADN y ciclo celular. Inhibir UBC podría tener consecuencias catastróficas (Wang y Li, 2021). | No existen estudios que relacionen de manera directa a la UBC con la DM2. |

# 3. JUSTIFICACIÓN

La enfermedad de Alzheimer es la forma más común de demencia, representando la mayoría de los casos. Las proyecciones indican que el número de personas afectadas por este trastorno se duplicará cada 20 años. En México, la enfermedad de Alzheimer también representa un problema grave de salud pública. Aunque existen varios medicamentos aprobados por la FDA para tratar esta enfermedad, estos se enfocan principalmente en mitigar los efectos en etapas específicas y presentan efectos secundarios como náuseas, vómitos, diarrea, reacciones alérgicas, pérdida de apetito, dolor de cabeza, confusión, mareos y caídas. Debido a estos inconvenientes, se han llevado a cabo numerosas investigaciones para identificar potenciales inhibidores dirigidos a diferentes objetivos asociados con la enfermedad de Alzheimer. Sin embargo, la tasa de éxito en el desarrollo de nuevos fármacos para esta enfermedad es casi nula. Como resultado, han surgido múltiples estudios para desarrollar nuevos tratamientos. Diversas investigaciones han demostrado que, debido a la naturaleza multifactorial de la enfermedad de Alzheimer y su asociación con varias causas, se deben diseñar fármacos multidiana. Recientemente, se ha identificado la diabetes mellitus como una de estas causas. En la última década, los investigadores han probado diversos medicamentos para la diabetes mellitus en pacientes con Alzheimer, encontrando que pueden mejorar la cognición y la memoria. No obstante, actualmente no existe ningún estudio sobre el diseño de fármacos para la asociación entre Alzheimer y diabetes mellitus. Por lo tanto, esta investigación propone utilizar modelos de aprendizaje automático para identificar potenciales inhibidores multidiana enfocados a proteínas asociadas a Alzheimer y diabetes mellitus.

# 4. HIPÓTESIS

Al menos uno de los modelos QSAR combinados con técnicas de Machine Learning tendrá una precisión de al menos 0.85 para cada una de las proteínas diana, permitiendo así la identificación de potenciales inhibidores que actúen sobre al menos dos blancos moleculares relevantes para el Alzheimer y la Diabetes Mellitus tipo 2.

# 5. OBJETIVOS

**5.1. Objetivo general**

Identificar inhibidores con capacidad para modular múltiples blancos relacionados con la asociación entre la enfermedad de Alzheimer y la diabetes mellitus 2.

**5.2. Objetivos específicos**

5.2.1. Recopilar y analizar datos experimentales de inhibición de moléculas dirigidas a las proteínas STAT3, EGFR, SRC, MAPK1 y ESR1.

5.2.2. Desarrollar modelos QSAR con aprendizaje automático para predecir la actividad inhibitoria de moléculas dirigidas a las proteínas STAT3, EGFR, SRC, MAPK1 y ESR1.

5.2.3. Identificar potenciales inhibidores aplicando los modelos QSAR desarrollados a una base de datos de compuestos naturales.

5.2.4. Realizar simulaciones de acoplamiento molecular para los inhibidores predichos por los modelos QSAR.

# 6. MATERIALES Y MÉTODOS

En la presente investigación se construirán modelos QSAR utilizando técnicas de Machine Learning para cada una de las proteínas dianas asociadas al Alzheimer y la Diabetes Mellitus tipo 2. Estos modelos se emplearán posteriormente para realizar un cribado virtual utilizando moléculas de bases de datos de compuestos naturales, con el objetivo de identificar inhibidores multidiana. Las moléculas obtenidas a través del cribado virtual con los modelos QSAR serán sometidas a un análisis de acoplamiento molecular para realizar un segundo cribado virtual y validar la actividad de estas moléculas. La Figura 4 muestra un esquema general de la metodología que se seguirá en esta investigación.

Gráfico, Diagrama, Gráfico de embudo

Descripción generada automáticamente

**Figura 4.** Esquema general de la metodología.

# 6.1. Recopilación y análisis de datos experimentales de inhibición

# 6.1.1. Obtención de los datos experimentales de los inhibidores

Los datos que se emplearán para entrenar los modelos QSAR con Machine Learning se descargarán de la base de datos bioinformática libre ChEMBL (Davies *et al*., 2015). De esta base de datos, se obtendrán datos experimentales de inhibición de moléculas dirigidas a las proteínas STAT3, EGFR, SRC, MAPK1 y ESR1. Para esto, se utilizará el buscador de ChEMBL, para recuperar todos los datos experimentales y la información de actividad inhibitoria correspondiente para cada proteína diana. Se realizará la descarga de aquellos datos que estén en el formato estándar IC50 y se almacenarán en un archivo CSV.

# 6.1.2. Tratamiento de datos

Para el análisis y procesamiento de los datos de las moléculas recolectadas para todas las proteínas, se utilizará el lenguaje de programación de alto nivel Python 3.11.0 (Van y Fred, 2009) junto con la biblioteca de análisis de datos pandas (The pandas development team, 2020). Este proceso se debe a que las bases de datos a menudo contienen datos faltantes o duplicados y deben ser tratados previo al entrenamiento de los modelos de Machine Learning. El entorno de desarrollo será Visual Studio Code, versión 1.89.1 (Microsoft, 2011). Cada uno de los cinco archivos CSV descargados previamente se cargará en un DataFrame respectivo, uno para cada proteína diana, utilizando la función read\_csv de pandas. Luego, cada DataFrame se filtrará para conservar únicamente las columnas de interés: Molecule ChEMBL ID, Molecule Name, Molecular Weight, SMILES, Standard Type, Standard Relation, Standard Value, y Standard Units. Posteriormente, se analizarán los valores faltantes en cada columna utilizando la función “isna” de pandas, que se utiliza para se utiliza para detectar valores faltantes en un DataFrame. A continuación, se limpiarán los valores faltantes en las columnas "Standard Value" y "SMILES", ya que son esenciales para el entrenamiento de los modelos de aprendizaje automático y el cálculo de descriptores moleculares. También se eliminarán las moléculas duplicadas basándose en la columna "Molecule Name" utilizando la función “drop\_duplicates” de pandas en cada DataFrame. Finalmente, los DataFrames resultantes de la limpieza de los datos se almacenarán en archivos CSV. La Figura 5 esquematiza las etapas generales del tratamiento de los datos para cada una de las proteínas diana.

Diagrama

Descripción generada automáticamente

**Figura 5.** Esquema general del tratamiento de datos y selección de columnas de interés.

# 6.1.3. Establecimiento del umbral de inhibición

Los nuevos archivos guardados en formato CSV se cargarán como DataFrames de pandas. En cada uno de estos DataFrames, se añadirá una columna denominada 'actividad' con valores binarios, 0 o 1. El valor 1 indicará que una molécula es activa respecto a la proteína correspondiente del DataFrame, mientras que el valor 0 indicará inactividad. El umbral para definir la actividad se establecerá en 1 μM; las moléculas con un IC50 inferior a 1 μM serán consideradas activas, mientras que aquellas con un IC50 superior a 1 μM serán consideradas inactivas. Este umbral es ampliamente utilizado en la creación de modelos QSAR, ya que los inhibidores por debajo de este valor son considerados potentes, asegurando así una alta probabilidad de eficacia en aplicaciones terapéuticas y de investigación (Krishna *et al*., 2022; Lee *et al*., 2019; Mukerjee *et al*., 2022). Como análisis exploratorio inicial, se calcularán la media, mediana y moda para los grupos de moléculas activas e inactivas en cada DataFrame, utilizando la función groupby de pandas y las respectivas funciones de agregación. Posteriormente, se crearán diagramas de caja (boxplots) empleando la función boxplot de la librería Seaborn en Python (Waskom, 2021), ampliamente utilizada para visualizaciones gráficas en Python. Finalmente, los DataFrames actualizados se guardarán nuevamente en archivos CSV para su uso en etapas posteriores. Como ejemplo de visualización, la Figura 6 muestra cómo se verían los DataFrames con las columnas filtradas y la columna de actividad añadida. En este caso, se observa que, dado que la molécula tiene un valor de IC50 de 892, su actividad se clasifica como 0.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Molecule ChEMBL ID | Molecule Name | Molecular Weight | Smiles | Standard Type | Standard Relation | Standard Value | Standard Units | Actividad |
| CHEMBL378581 | NaN | 309.77 | O=S1(=O)c2ccc(O)cc2CCN1c1ccccc1Cl | IC50 | '=' | 892 | nM | 0 |

**Figura 6.** Ejemplo de los DataFrames con las columnas filtradas y la columna de actividad añadida.

# 6.1.4. Cálculo de los descriptores moleculares

Un descriptor molecular es una cantidad numérica que describe una propiedad de una molécula de manera cuantitativa y se utilizan ampliamente en modelos QSAR. Para calcular estos descriptores, se cargarán los cinco archivos CSV con los umbrales de actividad establecidos como nuevos DataFrames de pandas. Utilizando el módulo RDKit, una colección de herramientas para la quimioinformática y el diseño de fármacos, se calcularán descriptores moleculares para todas las moléculas. RDKit ofrece una amplia gama de más de 200 descriptores moleculares de varios tipos. Previo al cálculo de descriptores, se creará una nueva columna en cada uno de los cinco DataFrames para almacenar objetos tipo mol generados a partir de los SMILES de cada molécula, utilizando la función “MolFromSmiles” del módulo RDKit, versión 2023.9.5 (RDKit team, 2024). Estos objetos tipo mol son representaciones internas de moléculas químicas que permiten manipular y analizar estructuras moleculares dentro de la biblioteca RDKit. Posteriormente, se utilizará la función “all\_descriptors” para calcular específicamente descriptores 1D y 2D, abarcando un total de 110 descriptores bidimensionales ofrecidos por RDKit. Estos descriptores incluyen huellas moleculares, carga y polaridad, topología, anillos y ciclos, y estado electrónico. Se añadirá una nueva columna a cada DataFrame con el nombre del descriptor respectivo y los valores calculados para cada molécula. Finalmente, los DataFrames con los descriptores calculados se guardarán como archivos CSV.

# 6.1.5. Filtros de baja varianza y de alta correlación

Se cargarán los archivos CSV, que ahora incluyen los descriptores moleculares bidimensionales calculados con RDKit, en nuevos DataFrames de pandas. Aunque se han calculado los descriptores moleculares, muchos de ellos no ofrecen información útil para el modelo final, requieren capacidad de cómputo significativa y pueden afectar negativamente la precisión del modelo. Por ello, es esencial emplear solo los descriptores útiles para el entrenamiento de los modelos de Machine Learning. Uno de los métodos para seleccionar descriptores útiles consiste en aplicar filtros de baja varianza y alta correlación. Estos filtros mejoran el rendimiento del modelo al reducir la dimensionalidad de los datos y eliminar descriptores redundantes o no informativos. Los descriptores con baja varianza son aquellos que muestran poca variabilidad entre las muestras, por lo que no contribuyen significativamente a la diferenciación entre clases. Por otro lado, los descriptores altamente correlacionados proporcionan información similar o duplicada al modelo, lo que puede llevar a problemas de multicolinealidad (Dhamodharan y Mohan, 2022; Lee *et al*., 2019). Se creará un correlograma de los descriptores moleculares utilizando la función “corr” de la librería pandas. Para la visualización gráfica del correlograma, se empleará la función heatmap de seaborn. Posteriormente, se descartarán aquellos descriptores moleculares cuya correlación entre sí sea superior a 0.7 utilizando máscaras de la librería pandas sobre el correlograma. Asimismo, se calculará la varianza para cada descriptor y se eliminarán aquellos con una varianza de 0 o cercana a 0. Para cada DataFrame, se eliminarán las columnas de los descriptores que cumplan con estos criterios, con el fin de evitar el sobreajuste y asegurar la optimización de los modelos de aprendizaje automático.

# 6.2. Entrenamiento de los modelos de aprendizaje automático

# 6.2.1. División del conjunto de datos

Los datos correspondientes a los descriptores seleccionados se cargarán como archivos DataFrame para cada una de las proteínas diana. Para cada archivo, se creará una variable llamada “X” que almacenará la información de los descriptores de todas las moléculas correspondientes a cada proteína diana. Además, se generará una variable “y” que contendrá únicamente la columna de actividad de las moléculas. Posteriormente, tanto la variable “X” como “y” se dividirán en dos subconjuntos de datos denominados “conjunto de entrenamiento” y “conjunto de prueba”, utilizando la función “train\_test\_split” de la librería scikit-learn 1.3.2 para Python (Pedregosa *et al*., 2011a). La división se realizará en una proporción estándar del 80-20, donde el 80% de los datos de “X” se asignarán al conjunto de entrenamiento y el 20% restante al conjunto de prueba. Este mismo criterio se aplicará a la variable “y”. Las variables resultantes de estas divisiones se guardarán en formato CSV para su posterior uso en la validación interna y externa, así como para el entrenamiento de los modelos de aprendizaje automático. La Figura 6 muestra de manera ilustrativa cómo se llevará a cabo la división de los datos.

Diagrama

Descripción generada automáticamente

**Figura 7.** Esquema ilustrativo de la creación de los conjuntos de entrenamiento y prueba.

# 6.2.2. Validación interna

En la presente investigación se emplearán los modelos de Machine Learning XGBoost y Random Forest. El Random Forest es un algoritmo basado en múltiples árboles de decisión. Un árbol de decisión es una estructura jerárquica similar a un diagrama de flujo, donde se toman decisiones basadas en reglas derivadas de las variables de los datos. Cada árbol comienza con un nodo raíz, donde se realiza la primera división de los datos según la variable que mejor los separa. A partir de ahí, se forman nodos internos que continúan dividiendo los datos según diferentes variables, hasta llegar a los nodos hoja, que representan las decisiones finales o predicciones del modelo. En la Figura 8 se ilustra un árbol de decisión para la aprobación de créditos bancarios a clientes. La variable del nodo raíz es el ingreso anual del cliente, con la condición de que, si este ingreso supera los 50 mil, se pasa al nodo interno de historial crediticio. Si el cliente tiene un buen historial crediticio, se aprueba el crédito. Este proceso continúa con el resto de las decisiones hasta llegar a una conclusión final (Blockeel *et al*., 2023).

Diagrama

Descripción generada automáticamente

**Figura 8.** Árbol de decisión para la aprobación de créditos bancarios a clientes.

El modelo Random Forest se basa en la creación de múltiples árboles de decisión, entrenados con muestras aleatorias del conjunto de datos original. La decisión final se obtiene promediando los resultados de todos los árboles (Breiman, 2001). En XGBoost, los árboles de decisión se crean de manera secuencial, mejorando los errores de los anteriores mediante funciones de pérdida (Chen y Guestrin, 2016a). Se emplearán Random Forest de la librería scikit-learn (Pedregosa *et al*., 2011b) y XGBoost de la librería XGBoost 2.0.3 (T. Chen y Guestrin, 2016). La validación interna es crucial para ajustar y evaluar el rendimiento de los modelos, determinando los mejores hiperparámetros. Los hiperparámetros son configuraciones ajustadas antes del entrenamiento de los modelos (Elgeldawi *et al*., 2021). Para ello, se cargarán los archivos CSV de las variables divididas en conjuntos de entrenamiento y prueba como DataFrames. Se entrenarán los modelos XGBoost y Random Forest con varios ajustes de hiperparámetros para cada DataFrame, buscando la mejor configuración. Para Random Forest, se ajustarán “n\_estimators”, “max\_depth” y “max\_features”. Las combinaciones probadas serán: “n\_estimators” = [50, 100, 150], “max\_depth” = [10, 20, 30], y “max\_features” = [“auto”, “sqrt”, “log2”]. Para XGBoost, se ajustarán “n\_estimators”, “learning\_rate” y “max\_depth” con las combinaciones: “n\_estimators” = [50, 100, 150], “max\_depth” = [10, 20, 30], y “learning\_rate” = [0.01, 0.05, 0.1]. Se utilizará la función “GridSearchCV” de scikit-learn con validación cruzada de 5 pliegues y precisión como métrica de evaluación para encontrar los mejores hiperparámetros. De esta forma, se entrenarán múltiples modelos con combinaciones únicas de hiperparámetros y se determinará cuál es la mejor combinación de hiperparámetros para cada tipo de modelo. Los mejores hiperparámetros de los modelos se guardarán en un DataFrame para su posterior uso en el entrenamiento final.

# 6.2.3. Entrenamiento de los modelos y validación externa

Se cargarán nuevamente los datos de los archivos CSV que contienen las variables divididas en conjuntos de entrenamiento y prueba en DataFrames. Se iniciarán los modelos utilizando las clases “XGBClassifier” y “RandomForestRegressor”, configurando cada uno con los mejores hiperparámetros obtenidos durante la validación interna para cada proteína. Luego, cada modelo se entrenará con su respectivo conjunto de datos de entrenamiento. Tras el entrenamiento, los modelos se utilizarán para predecir en los conjuntos de prueba, estimando métricas como precisión, exactitud, recall y F1 score con la función “classification\_report” de scikit-learn así como el cálculo de las curvas ROC. Para evitar el reentrenamiento, se empleará la librería joblib para guardar y cargar los modelos en archivos “.pkl”, permitiendo su uso posterior en la predicción de la actividad de moléculas en una base de datos de compuestos naturales.

# 6.3. Determinación de potenciales inhibidores

# 6.3.1. Obtención de las moléculas a partir de las bases de datos

Se descargarán datos de moléculas químicas para el reposicionamiento de fármacos de las bases de datos ReframeDB (Janes *et al*., 2018) y COCONUT (Sorokina *et al*., 2021). Estas bases de datos se utilizarán debido a las ventajas que ofrecen, como la validación previa de los compuestos y la inclusión de productos naturales. Se descargará la totalidad de los compuestos de ReframeDB, junto con sus representaciones SMILES, resultando en un total de 12,790 compuestos en formato CSV. De manera similar, se descargarán los compuestos y sus SMILES de la base de datos COCONUT, obteniendo un total de 407,270 compuestos en formato SMI. Estos archivos se emplearán como fuente de datos para los modelos QSAR previamente entrenados

# 6.3.2. Tratamiento de los datos descargados

Los datos descargados de las bases de datos se cargarán como DataFrames utilizando la librería pandas de Python (The pandas development team, 2020). Para la base de datos de Reframe, se utilizará la función read\_csv de pandas para cargar los datos en un DataFrame. Se creará un nuevo DataFrame que incluya solo las columnas de interés, que son “smile” y “rfm\_id”. De manera similar, se cargará el archivo de la base de datos de COCONUT, que está en formato “smi”, leyéndolo con pandas como si fuera un CSV, especificando que la separación se realiza mediante espacios en blanco. Para ambos DataFrames, se empleará la función “dropna” de pandas con el objetivo de limpiar los datos faltantes y la función “drop\_duplicates” para eliminar datos repetidos. Los DataFrames resultantes después del tratamiento de datos se guardarán como archivos en formato CSV.

# 6.3.3. Cálculo de descriptores moleculares

Se cargarán los archivos CSV con los datos tratados de las bases de datos como DataFrames de pandas. Para ambos DataFrames, se añadirá una nueva columna que almacenará objetos de tipo mol generados a partir de los SMILES de cada molécula utilizando la función MolFromSmiles del módulo RDKit (RDKit team, 2024), siguiendo el mismo procedimiento usado para los datos de entrenamiento y prueba de los modelos de aprendizaje automático. Luego, se empleará la función “all\_descriptors” para calcular los descriptores seleccionados. Cada descriptor se añadirá como una nueva columna en los DataFrames. Finalmente, los DataFrames con los descriptores calculados se guardarán como archivos CSV.

# 6.3.4. Determinación de potenciales inhibidores multiblanco

Se cargarán los archivos CSV que contienen las moléculas de las bases de datos, así como los 10 modelos QSAR entrenados (5 de Random Forest y 5 de XGBoost), utilizando la función load de la librería joblib de Python. Se creará una variable llamada X que almacenará todos los datos de los descriptores moleculares de las moléculas descargadas. Esta variable X se introducirá en cada uno de los diez modelos QSAR para predecir la actividad de estas moléculas utilizando la función predict de scikit-learn (Pedregosa *et al*., 2011a). Los resultados de las predicciones de los 10 modelos se guardarán en un único DataFrame, donde cada columna llevará el nombre del respectivo modelo y la proteína diana. Posteriormente, se filtrarán los resultados usando pandas para identificar moléculas que inhiben al menos 2 proteínas diana y con una probabilidad superior a 0.9 de pertenecer a dicha clase. Estas moléculas se identificarán como potenciales inhibidores multidiana para la asociación Alzheimer-Diabetes mellitus. El DataFrame resultante se guardará como un archivo CSV para su posterior análisis mediante acoplamiento molecular.

# 6.4. Acoplamiento molecular

# 6.4.1. Preparación del receptor

Se descargarán las estructuras tridimensionales de las proteínas diana en formato PDB desde la base de datos Protein Data Bank (PDB) (Berman *et al*., 2000). Posteriormente, se utilizará el software ChimeraX 1.17.3 (Ding *et al*., 2022) para limpiar estas estructuras, eliminando moléculas de agua y ligandos no deseados. Finalmente, se empleará MGLTools 1.5.7 (Morris *et al*., 2009) para convertir los archivos PDB a formato PDBQT, adecuado para estudios de acoplamiento molecular.

# 6.4.2. Preparación del ligando

Se prepararán los potenciales inhibidores predichos utilizando modelos QSAR basados en Machine Learning. Primero, se empleará el software RDKit 2023.9.5 (RDKit team, 2024) para Python para generar las coordenadas tridimensionales de las moléculas y añadir átomos de hidrógeno polares. Luego, se utilizará el software Meeko (Giménez-Dejoz *et al*., 2022) para convertir los archivos de estos ligandos a formato PDBQT, adecuado para estudios de acoplamiento molecular.

# 6.4.3. Acoplamiento molecular

Se realizará el acoplamiento molecular de los ligandos con sus respectivas proteínas utilizando el software SMINA (Koes *et al*., 2013). Se obtendrán hasta tres puntajes de acoplamiento: Vina, Vinardo y dkoes por esta herramienta. Para comparar y promediar los resultados de estos tres métodos, se transformarán los puntajes a valores Z-score. Finalmente, se calculará el Z-consenso promediando los Z-scores de Vina, Vinardo y dkoes para cada inhibidor, proporcionando así una evaluación integral de los potenciales inhibidores. Estos resultados se compararán con los resultados obtenidos por los modelos de Machine Learning con el objetivo de realizar una validación y hasta un segundo cribado virtual con los resultados de los acoplamientos moleculares.

# 7. CRONOGRAMA DE ACTIVIDADES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **2023** | | **2024** | |
|  | **S1** | **S2** | **S3** | **S4** |
| **Obtención y preparación de los datos** |  |  |  |  |
| **Entrenamiento y validación de los modelos** |  |  |  |  |
| **Determinación de potenciales inhibidores empleando los modelos** |  |  |  |  |
| **Acoplamiento molecular** |  |  |  |  |
| **Revisión de literatura científica** |  |  |  |  |
| **Redacción de la tesis** |  |  |  |  |

# 8. BIBLIOGRAFÍA

Arnold, S. E., Arvanitakis, Z., Macauley-Rambach, S. L., Koenig, A. M., Wang, H. Y., Ahima, R. S., Craft, S., Gandy, S., Buettner, C., Stoeckel, L. E., Holtzman, D. M., y Nathan, D. M. (2018). Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. *Nature Reviews Neurology 2018 14:3*, *14*(3), 168–181. https://doi.org/10.1038/nrneurol.2017.185

Balczon, R., Lin, M. T., Voth, S., Nelson, A. R., Schupp, J. C., Wagener, B. M., Pittet, J. F., & Stevens, T. (2024). Lung Endothelium, Tau, and Amyloids in Health and Disease. *Physiological Reviews*, *104*(2), 533–587. https://doi.org/10.1152/PHYSREV.00006.2023/ASSET/IMAGES/LARGE/PHYSREV.00006.2023\_F016.JPEG

Bao, L. Q., Baecker, D., Mai Dung, D. T., Phuong Nhung, N., Thi Thuan, N., Nguyen, P. L., Phuong Dung, P. T., Huong, T. T. L., Rasulev, B., Casanola-Martin, G. M., Nam, N. H., & Pham-The, H. (2023). Development of Activity Rules and Chemical Fragment Design for In Silico Discovery of AChE and BACE1 Dual Inhibitors against Alzheimer’s Disease. *Molecules*, *28*(8), 3588. https://doi.org/10.3390/MOLECULES28083588/S1

Batool, M., Ahmad, B., & Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. *International Journal of Molecular Sciences 2019, Vol. 20, Page 2783*, *20*(11), 2783. https://doi.org/10.3390/IJMS20112783

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/NAR/28.1.235

Bird, T. D. (2008). Genetic Aspects of Alzheimer Disease. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *10*(4), 231. https://doi.org/10.1097/GIM.0B013E31816B64DC

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, *6*, 1124553. https://doi.org/10.3389/FRAI.2023.1124553/BIBTEX

Bosc, N., Atkinson, F., Felix, E., Gaulton, A., Hersey, A., & Leach, A. R. (2019). Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *Journal of Cheminformatics*, *11*(1), 1–16. https://doi.org/10.1186/S13321-018-0325-4/TABLES/6

Brands, A. M. A., Biessels, G. J., De Haan, E. H. F., Kappelle, L. J., & Kessels, R. P. C. (2005). The Effects of Type 1 Diabetes on Cognitive PerformanceA meta-analysis. *Diabetes Care*, *28*(3), 726–735. https://doi.org/10.2337/DIACARE.28.3.726

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Carracedo-Reboredo, P., Liñares-Blanco, J., Rodríguez-Fernández, N., Cedrón, F., Novoa, F. J., Carballal, A., Maojo, V., Pazos, A., & Fernandez-Lozano, C. (2021). A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal*, *19*, 4538–4558. https://doi.org/10.1016/J.CSBJ.2021.08.011

Castillo-Velázquez, R., Martínez-Morales, F., Castañeda-Delgado, J. E., García-Hernández, M. H., Herrera-Mayorga, V., Paredes-Sánchez, F. A., Rivera, G., Rivas-Santiago, B., & Lara-Ramírez, E. E. (2023). Bioinformatic prediction of the molecular links between Alzheimer’s disease and diabetes mellitus. *PeerJ*, *11*. https://doi.org/10.7717/PEERJ.14738

Cavas, L., Topcam, G., Gundogdu-Hizliates, C., & Ergun, Y. (2019). Neural Network Modeling of AChE Inhibition by New Carbazole-Bearing Oxazolones. *Interdisciplinary Sciences – Computational Life Sciences*, *11*(1), 95–107. https://doi.org/10.1007/S12539-017-0245-4/FIGURES/7

Chen, T., & Guestrin, C. (2016a). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T., & Guestrin, C. (2016b). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cho, S. B. (2024). Comorbidity Genes of Alzheimer’s Disease and Type 2 Diabetes Associated with Memory and Cognitive Function. *International Journal of Molecular Sciences*, *25*(4), 2211. https://doi.org/10.3390/IJMS25042211/S1

Choi, H. J., Jeong, Y. J., Kim, J., & Hoe, H. S. (2023). EGFR is a potential dual molecular target for cancer and Alzheimer’s disease. *Frontiers in Pharmacology*, *14*, 1238639. https://doi.org/10.3389/FPHAR.2023.1238639/BIBTEX

Chollet, F. (2018). *Deep Learning with Python* (Toni Arritola, Ed.). Manning Publications Co.

Clodfelder-Miller, B. J., Zmijewska, A. A., Johnson, G. V. W., & Jope, R. S. (2006). Tau Is Hyperphosphorylated at Multiple Sites in Mouse Brain In Vivo After Streptozotocin-Induced Insulin Deficiency. *Diabetes*, *55*(12), 3320–3325. https://doi.org/10.2337/DB06-0485

Cobley, J. N., Fiorello, M. L., & Bailey, D. M. (2018). 13 reasons why the brain is susceptible to oxidative stress. *Redox Biology*, *15*, 490–503. https://doi.org/10.1016/J.REDOX.2018.01.008

Concu, R., González-Durruthy, M., & Cordeiro, M. N. D. S. (2020). Developing a Multi-target Model to Predict the Activity of Monoamine Oxidase A and B Drugs. *Current Topics in Medicinal Chemistry*, *20*(18), 1593–1600. https://doi.org/10.2174/1568026620666200603121224

Coumar, M. S. (2021). *Molecular docking for computer-aided drug design : fundamentals, techniques, resources and applications*.

Cukierman, T., Gerstein, H. C., & Williamson, J. D. (2005). Cognitive decline and dementia in diabetes - Systematic overview of prospective observational studies. *Diabetologia*, *48*(12), 2460–2469. https://doi.org/10.1007/S00125-005-0023-4/FIGURES/3

Cummings, J. L., Doody, R., & Clark, C. (2007). Disease-modifying therapies for Alzheimer disease: challenges to early intervention. *Neurology*, *69*(16), 1622–1634. https://doi.org/10.1212/01.WNL.0000295996.54210.69

Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer’s disease drug-development pipeline: Few candidates, frequent failures. *Alzheimer’s Research and Therapy*, *6*(4), 1–7. https://doi.org/10.1186/ALZRT269/TABLES/3

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., & Overington, J. P. (2015). ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, *43*(Web Server issue), W612. https://doi.org/10.1093/NAR/GKV352

de Boer, R. J., van Lidth de Jeude, J. F., & Heijmans, J. (2024). ER stress and the unfolded protein response in gastrointestinal stem cells and carcinogenesis. *Cancer Letters*, *587*. https://doi.org/10.1016/J.CANLET.2024.216678

De Felice, F. G. (2013). Alzheimer’s disease and insulin resistance: translating basic science into clinical applications. *The Journal of Clinical Investigation*, *123*(2), 531–539. https://doi.org/10.1172/JCI64595

de la Monte, S. M. (2009). Insulin resistance and Alzheimer’s disease. *BMB Reports*, *42*(8), 475. https://doi.org/10.5483/BMBREP.2009.42.8.475

Deane, R., Bell, R., Sagare, A., & Zlokovic, B. (2009). Clearance of amyloid-beta peptide across the blood-brain barrier: implication for therapies in Alzheimer’s disease. *CNS & Neurological Disorders Drug Targets*, *8*(1), 16–30. https://doi.org/10.2174/187152709787601867

Devendra, D., & Eisenbarth, G. S. (2003). 17. Immunologic endocrine disorders. *Journal of Allergy and Clinical Immunology*, *111*(2 SUPPL. 2). https://doi.org/10.1067/mai.2003.81

Dhamodharan, G., & Mohan, C. G. (2022). Machine learning models for predicting the activity of AChE and BACE1 dual inhibitors for the treatment of Alzheimer’s disease. *Molecular Diversity*, *26*(3), 1501–1517. https://doi.org/10.1007/S11030-021-10282-8/TABLES/6

Ding, X., Meng, C., Dong, H., Zhang, S., Zhou, H., Tan, W., Huang, L., He, A., Li, J., Huang, J., Li, W., Zou, F., & Zou, M. (2022). Extracellular Hsp90α, which participates in vascular inflammation, is a novel serum predictor of atherosclerosis in type 2 diabetes. *BMJ Open Diabetes Research and Care*, *10*(1), e002579. https://doi.org/10.1136/BMJDRC-2021-002579

Docherty, F. M., & Russ, H. A. (2019). Cell–Cell Interactions Driving Differentiation of Adult Pancreatic Stem Cells. *Encyclopedia of Tissue Engineering and Regenerative Medicine: Volumes 1-3*, *1–3*, 367–374. https://doi.org/10.1016/B978-0-12-801238-3.65615-5

Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. *Informatics 2021, Vol. 8, Page 79*, *8*(4), 79. https://doi.org/10.3390/INFORMATICS8040079

Erf, G. F. (2014). Autoimmune Diseases of Poultry. *Avian Immunology: Second Edition*, 315–332. https://doi.org/10.1016/B978-0-12-396965-1.00018-2

FDA. (2019). *Drugs@FDA Glossary of Terms | FDA*. https://www.fda.gov/drugs/drug-approvals-and-databases/drugsfda-glossary-terms

Fisher, A. (2008). M1 muscarinic agonists target major hallmarks of Alzheimer’s disease--the pivotal role of brain M1 receptors. *Neuro-Degenerative Diseases*, *5*(3–4), 237–240. https://doi.org/10.1159/000113712

Gan, H. K., Cvrljevic, A. N., & Johns, T. G. (2013). The epidermal growth factor receptor variant III (EGFRvIII): Where wild things are altered. *FEBS Journal*, *280*(21), 5350–5370. https://doi.org/10.1111/febs.12393

Geng, C., Wang, Z. Bin, & Tang, Y. (2024). Machine learning in Alzheimer’s disease drug discovery and target identification. *Ageing Research Reviews*, *93*, 102172. https://doi.org/10.1016/J.ARR.2023.102172

Gillespie, K. M. (2006). Type 1 diabetes: pathogenesis and prevention. *CMAJ*, *175*(2), 165–170. https://doi.org/10.1503/CMAJ.060244

Giménez-Dejoz, J., Trott, O., & Hall, D. R. (2022). Meeko: A library for preparing high-quality molecular docking input with machine learning and cheminformatics. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btab856

Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D., & Crowther, R. A. (1989). Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer’s disease. *Neuron*, *3*(4), 519–526. https://doi.org/10.1016/0896-6273(89)90210-9

Gong, C. X., Liu, F., & Iqbal, K. (2018). Multifactorial Hypothesis and Multi-Targets for Alzheimer’s Disease. *Journal of Alzheimer’s Disease : JAD*, *64*(s1), S107–S117. https://doi.org/10.3233/JAD-179921

Greeny, A., Nair, A., Sadanandan, P., Satarker, S., Famurewa, A. C., & Nampoothiri, M. (2024). Epigenetic Alterations in Alzheimer’s Disease: Impact on Insulin Signaling and Advanced Drug Delivery Systems. *Biology 2024, Vol. 13, Page 157*, *13*(3), 157. https://doi.org/10.3390/BIOLOGY13030157

Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(13), 4913. https://doi.org/10.1073/PNAS.83.13.4913

Gu, F., Zhu, M., Shi, J., Hu, Y., & Zhao, Z. (2008). Enhanced oxidative stress is an early event during development of Alzheimer-like pathologies in presenilin conditional knock-out mice. *Neuroscience Letters*, *440*(1), 44–48. https://doi.org/10.1016/J.NEULET.2008.05.050

Henriksen, E. J., Diamond-Stanic, M. K., & Marchionne, E. M. (2011). Oxidative stress and the etiology of insulin resistance and type 2 diabetes. *Free Radical Biology and Medicine*, *51*(5), 993–999. https://doi.org/10.1016/J.FREERADBIOMED.2010.12.005

Henson, J., Anyiam, O., & Vishnubala, D. (2023). Type 2 Diabetes. *Exercise Management for Referred Medical Conditions*, 223–252. https://doi.org/10.4324/9781315102399-12

Huang, A., Yang, Y. M., Yan, C., Kaley, G., Hintze, T. H., & Sun, D. (2012). Altered MAPK Signaling in Progressive Deterioration of Endothelial Function in Diabetic Mice. *Diabetes*, *61*(12), 3181. https://doi.org/10.2337/DB12-0559

IMSS. (2015). *Enfermedad de Alzheimer*. https://www.imss.gob.mx/salud-en-linea/enfermedad-alzheimer

Ishihara, T., Hong, M., Zhang, B., Nakagawa, Y., Lee, M. K., Trojanowski, J. Q., & Lee, V. M. Y. (1999). Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform. *Neuron*, *24*(3), 751–762. https://doi.org/10.1016/S0896-6273(00)81127-7

Ismail, L., Materwala, H., & Al Kaabi, J. (2021). Association of risk factors with type 2 diabetes: A systematic review. *Computational and Structural Biotechnology Journal*, *19*, 1759–1785. https://doi.org/10.1016/J.CSBJ.2021.03.003

Janes, J., Young, M. E., Chen, E., Rogers, N. H., Burgstaller-Muehlbacher, S., Hughes, L. D., Love, M. S., Hull, M. V., Kuhen, K. L., Woods, A. K., Joseph, S. B., Petrassi, H. M., McNamara, C. W., Tremblay, M. S., Su, A. I., Schultz, P. G., & Chatterjee, A. K. (2018). The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(42), 10750–10755. https://doi.org/10.1073/PNAS.1810137115/SUPPL\_FILE/PNAS.1810137115.SD01.XLSX

Johnson, G. V. W., & Stoothoff, W. H. (2004). Tau phosphorylation in neuronal cell function and dysfunction. *Journal of Cell Science*, *117*(24), 5721–5729. https://doi.org/10.1242/JCS.01558

Kandimalla, R., Thirumala, V., & Reddy, P. H. (2017). Is Alzheimer’s disease a Type 3 Diabetes? A critical appraisal. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1863*(5), 1078–1089. https://doi.org/10.1016/J.BBADIS.2016.08.018

Koes, D. R., Baumgartner, M. P., & Camacho, C. J. (2013). Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. *Journal of Chemical Information and Modeling*, *53*(8), 1893–1904. https://doi.org/10.1021/ci300604z

Kopf, D., & Frölich, L. (2009). Risk of Incident Alzheimer’s Disease in Diabetic Patients: A Systematic Review of Prospective Trials. *Journal of Alzheimer’s Disease*, *16*(4), 677–685. https://doi.org/10.3233/JAD-2009-1011

Krischer, J. P., Liu, X., Lernmark, Å., Hagopian, W. A., Rewers, M. J., She, J. X., Toppari, J., Ziegler, A. G., & Akolkar, B. (2022). Predictors of the Initiation of Islet Autoimmunity and Progression to Multiple Autoantibodies and Clinical Diabetes: The TEDDY Study. *Diabetes Care*, *45*(10), 2271. https://doi.org/10.2337/DC21-2612

Krishna, S., Borrel, A., Kleinstreuer, N., Huang, R., Zhao, J., & Xia, M. (2022). High-Throughput Chemical Screening and Structure-Based Models to Predict hERG Inhibition. *Biology*, *11*(2), 209. https://doi.org/10.3390/BIOLOGY11020209/S1

Kumar, V., Saha, A., & Roy, K. (2023). Multi-target QSAR modeling for the identification of novel inhibitors against Alzheimer’s disease. *Chemometrics and Intelligent Laboratory Systems*, *233*, 104734. https://doi.org/10.1016/J.CHEMOLAB.2022.104734

Kurochkin, I. V., Guarnera, E., & Berezovsky, I. N. (2018). Insulin-Degrading Enzyme in the Fight against Alzheimer’s Disease. *Trends in Pharmacological Sciences*, *39*(1), 49–58. https://doi.org/10.1016/J.TIPS.2017.10.008

Lavecchia, A., & Cerchia, C. (2016). In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discovery Today*, *21*(2), 288–298. https://doi.org/10.1016/J.DRUDIS.2015.12.007

Lee, J., Kumar, S., Lee, S. Y., Park, S. J., & Kim, M. H. (2019). Development of Predictive Models for Identifying Potential S100A9 Inhibitors Based on Machine Learning Methods. *Frontiers in Chemistry*, *7*, 453573. https://doi.org/10.3389/FCHEM.2019.00779/BIBTEX

Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D. M., Oshima, J., Pettingell, W. H., Yu, C. E., Jondro, P. D., Schmidt, S. D., Wang, K., Crowley, A. C., Fu, Y. H., Guenette, S. Y., Galas, D., Nemens, E., Wijsman, E. M., Bird, T. D., Schellenberg, G. D., & Tanzi, R. E. (1995). Candidate gene for the chromosome 1 familial Alzheimer’s disease locus. *Science (New York, N.Y.)*, *269*(5226), 973–977. https://doi.org/10.1126/SCIENCE.7638622

Liskowsky, W., & Schliebs, R. (2006). Muscarinic acetylcholine receptor inhibition in transgenic Alzheimer-like Tg2576 mice by scopolamine favours the amyloidogenic route of processing of amyloid precursor protein. *International Journal of Developmental Neuroscience : The Official Journal of the International Society for Developmental Neuroscience*, *24*(2–3), 149–156. https://doi.org/10.1016/J.IJDEVNEU.2005.11.010

Lovestone, S., Boada, M., Dubois, B., Hüll, M., Rinne, J. O., Huppertz, H. J., Calero, M., Andrés, M. V., Gómez-Carrillo, B., León, T., & Del Ser, T. (2015). A Phase II Trial of Tideglusib in Alzheimer’s Disease. *Journal of Alzheimer’s Disease*, *45*(1), 75–88. https://doi.org/10.3233/JAD-141959

Mahley, R. W., & Rall, S. C. (2000). Apolipoprotein E: Far more than a lipid transport protein. *Annual Review of Genomics and Human Genetics*, *1*(2000), 507–537. https://doi.org/10.1146/ANNUREV.GENOM.1.1.507/CITE/REFWORKS

Medline. (2020). *Diabetes: MedlinePlus en español*. https://medlineplus.gov/spanish/diabetes.html

Medline. (2024). *Demencia: MedlinePlus enciclopedia médica*. https://medlineplus.gov/spanish/ency/article/000739.htm

Michailidis, M., Tata, D. A., Moraitou, D., Kavvadas, D., Karachrysafi, S., Papamitsou, T., Vareltzis, P., & Papaliagkas, V. (2022). Antidiabetic Drugs in the Treatment of Alzheimer’s Disease. *International Journal of Molecular Sciences*, *23*(9). https://doi.org/10.3390/IJMS23094641

Microsoft. (2011). *Visual Studio Code* (1.89.1). Microsoft.

Mohandas, E., Rajmohan, V., & Raghunath, B. (2009). Neurobiology of Alzheimer’s disease. *Indian Journal of Psychiatry*, *51*(1), 55. https://doi.org/10.4103/0019-5545.44908

Morris, G. M., Ruth, H., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/JCC.21256

Mukerjee, N., Al-Khafaji, K., Maitra, S., Suhail Wadi, J., Sachdeva, P., Ghosh, A., Buchade, R. S., Chaudhari, S. Y., Jadhav, S. B., Das, P., Hasan, M. M., Rahman, M. H., Albadrani, G. M., Altyar, A. E., Kamel, M., Algahtani, M., Shinan, K., Theyab, A., Abdel-Daim, M. M., … Sharma, R. (2022). Recognizing novel drugs against Keap1 in Alzheimer’s disease using machine learning grounded computational studies. *Frontiers in Molecular Neuroscience*, *15*, 1036552. https://doi.org/10.3389/FNMOL.2022.1036552/BIBTEX

Mullins, R. J., Diehl, T. C., Chia, C. W., & Kapogiannis, D. (2017). Insulin resistance as a link between amyloid-beta and tau pathologies in Alzheimer’s disease. *Frontiers in Aging Neuroscience*, *9*(MAY), 251216. https://doi.org/10.3389/FNAGI.2017.00118/BIBTEX

NIH. (2017). *¿Qué causa la enfermedad de Alzheimer? | National Institute on Aging*. https://www.nia.nih.gov/espanol/enfermedad-alzheimer/causa-enfermedad-alzheimer

NIH. (2023). *¿Cómo se trata la enfermedad de Alzheimer? | National Institute on Aging*. https://www.nia.nih.gov/espanol/cuidado-medico-enfermedad-alzheimer/como-se-trata-enfermedad-alzheimer

NIH gene. (2024). *PIK3R1 phosphoinositide-3-kinase regulatory subunit 1 [Homo sapiens (human)] - Gene - NCBI*. https://www.ncbi.nlm.nih.gov/gene/5295

OPS. (2023a). *Demencia - OPS/OMS | Organización Panamericana de la Salud*. https://www.paho.org/es/temas/demencia

OPS. (2023b). *Diabetes - OPS/OMS | Organización Panamericana de la Salud*. https://www.paho.org/es/temas/diabetes

Parvathy Dharshini, S. A., Sneha, N. P., Yesudhas, D., Kulandaisamy, A., Rangaswamy, U., Shanmugam, A., Taguchi, Y.-H., & Gromiha, M. M. (2022). Exploring Plausible Therapeutic Targets for Alzheimer’s Disease using Multi-omics Approach, Machine Learning and Docking. *Current Topics in Medicinal Chemistry*, *22*(22), 1868–1879. https://doi.org/10.2174/1568026622666220902110115

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, *26*(1), 80. https://doi.org/10.1016/J.DRUDIS.2020.10.010

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011a). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011b). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pena, F., Gutierrez-Lerma, A., Quiroz-Baez, R., & Arias, C. (2006). The role of beta-amyloid protein in synaptic function: implications for Alzheimer’s disease therapy. *Current Neuropharmacology*, *4*(2), 149–163. https://doi.org/10.2174/157015906776359531

Pillai, N., Dasgupta, A., Sudsakorn, S., Fretland, J., & Mavroudis, P. D. (2022). Machine Learning guided early drug discovery of small molecules. *Drug Discovery Today*, *27*(8), 2209–2215. https://doi.org/10.1016/J.DRUDIS.2022.03.017

Raghunathan, S., & Priyakumar, U. D. (2022). Molecular representations for machine learning applications in chemistry. *International Journal of Quantum Chemistry*, *122*(7), e26870. https://doi.org/10.1002/QUA.26870

RDKit team. (2024). *RDKit: Open-source cheminformatics* (2023.9.5).

Rogaev, E. I., Sherrington, R., Rogaeva, E. A., Levesque, G., Ikeda, M., Liang, Y., Chi, H., Lin, C., Holman, K., Tsuda, T., Mar, L., Sorbi, S., Nacmias, B., Piacentini, S., Amaducci, L., Chumakov, I., Cohen, D., Lannfelt, L., Fraser, P. E., … George-Hyslop, P. H. S. (1995). Familial Alzheimer’s disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer’s disease type 3 gene. *Nature 1995 376:6543*, *376*(6543), 775–778. https://doi.org/10.1038/376775a0

Rosenberg, P. B. (2005). Clinical aspects of inflammation in Alzheimer’s disease. *International Review of Psychiatry*, *17*(6), 503–514. https://doi.org/10.1080/02646830500382037

Roy, K., Kar, S., Das, R. N., Roy, K., Kar, S., & Das, R. N. (2015). Chapter 12 – Future Avenues. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. http://www.sciencedirect.com:5070/book/9780128015056/understanding-the-basics-of-qsar-for-applications-in-pharmaceutical-sciences-and-risk-assessment

Salazar, D. E., & Gormley, G. (2017). Modern Drug Discovery and Development. *Clinical and Translational Science: Principles of Human Research: Second Edition*, 719–743. https://doi.org/10.1016/B978-0-12-802101-9.00041-7

Sanabria-Castro, A., Alvarado-Echeverría, I., & Monge-Bonilla, C. (2017). Molecular Pathogenesis of Alzheimer’s Disease: An Update. *Annals of Neurosciences*, *24*(1), 46–54. https://doi.org/10.1159/000464422

Schneider, P., Walters, W. P., Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow, R. A., Fisher, J., Jansen, J. M., Duca, J. S., Rush, T. S., Zentgraf, M., Hill, J. E., Krutoholow, E., Kohler, M., Blaney, J., Funatsu, K., Luebkemann, C., & Schneider, G. (2019). Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery 2019 19:5*, *19*(5), 353–364. https://doi.org/10.1038/s41573-019-0050-3

Schneider, Patrick., & Xhafa, Fatos. (2022). *Anomaly detection and complex event processing over IoT data streams with application to eHealth and patient data monitoring*. Academic Press. http://www.sciencedirect.com:5070/book/9780128238189/anomaly-detection-and-complex-event-processing-over-iot-data-streams

Sheng, L., Bayliss, G., & Zhuang, S. (2020). Epidermal Growth Factor Receptor: A Potential Therapeutic Target for Diabetic Kidney Disease. *Frontiers in Pharmacology*, *11*. https://doi.org/10.3389/FPHAR.2020.598910

Sies, H. (2019). Oxidative Stress: Eustress and Distress in Redox Homeostasis. *Stress: Physiology, Biochemistry, and Pathology Handbook of Stress Series, Volume 3*, 153–163. https://doi.org/10.1016/B978-0-12-813146-6.00013-8

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, *13*(1), 1–13. https://doi.org/10.1186/S13321-020-00478-9/FIGURES/4

Speight, J., & Pouwer, F. (2023). Type 1 Diabetes. *Cambridge Handbook of Psychology, Health and Medicine: Third Edition*, 477–480. https://doi.org/10.29309/tpmj/2017.24.12.614

SSA México. (2021). *Enfermedad de Alzheimer, demencia más común que afecta a personas adultas mayores | Secretaría de Salud | Gobierno | gob.mx*. https://www.gob.mx/salud/es/articulos/enfermedad-de-alzheimer-demencia-mas-comun-que-afecta-a-personas-adultas-mayores?idiom=es

Starks, E. J., Patrick O’grady, J., Hoscheidt, S. M., Racine, A. M., Carlsson, C. M., Zetterberg, H., Blennow, K., Okonkwo, O. C., Puglielli, L., Asthana, S., Dowling, N. M., Gleason, C. E., Anderson, R. M., Davenport-Sis, N. J., Derungs, L. A. M., Sager, M. A., Johnson, S. C., & Bendlin, B. B. (2015). Insulin Resistance is Associated with Higher Cerebrospinal Fluid Tau Levels in Asymptomatic APOE ɛ4 Carriers. *Journal of Alzheimer’s Disease*, *46*(2), 525–533. https://doi.org/10.3233/JAD-150072

Stelzmann, R. A., Norman Schnitzlein, H., & Reed Murtagh, F. (1995). An English translation of Alzheimer’s 1907 paper, “Uber eine eigenartige Erkankung der Hirnrinde.” *Clinical Anatomy (New York, N.Y.)*, *8*(6), 429–431. https://doi.org/10.1002/CA.980080612

Strachan, M. W. J., Deary, I. J., Ewing, F. M. E., & Frier, B. M. (1997). Is Type II Diabetes Associated With an Increased Risk of Cognitive Dysfunction?: A critical review of published studies. *Diabetes Care*, *20*(3), 438–445. https://doi.org/10.2337/DIACARE.20.3.438

Sumi, N. J., Ctortecka, C., Hu, Q., Bryant, A. T., Fang, B., Remsing Rix, L. L., Ayaz, M., Kinose, F., Welsh, E. A., Eschrich, S. A., Lawrence, H. R., Koomen, J. M., Haura, E. B., & Rix, U. (2019). Divergent Polypharmacology-Driven Cellular Activity of Structurally Similar Multi-Kinase Inhibitors through Cumulative Effects on Individual Targets. *Cell Chemical Biology*, *26*(9), 1240-1252.e11. https://doi.org/10.1016/j.chembiol.2019.06.003

Sun, X., Bromley-Brits, K., & Song, W. (2012). Regulation of β-site APP-cleaving enzyme 1 gene expression and its role in Alzheimer’s Disease. *Journal of Neurochemistry*, *120*(SUPPL. 1), 62–70. https://doi.org/10.1111/J.1471-4159.2011.07515.X

Tanzi, R. E., & Bertram, L. (2005). Twenty years of the Alzheimer’s disease amyloid hypothesis: a genetic perspective. *Cell*, *120*(4), 545–555. https://doi.org/10.1016/J.CELL.2005.02.008

Terry, A. V., & Buccafusco, J. J. (2003). The cholinergic hypothesis of age and Alzheimer’s disease-related cognitive deficits: recent challenges and their implications for novel drug development. *The Journal of Pharmacology and Experimental Therapeutics*, *306*(3), 821–827. https://doi.org/10.1124/JPET.102.041616

The pandas development team. (2020). *pandas-dev/pandas: Pandas* (2.1.4). Zenodo.

Troutwine, B. R., Hamid, L., Lysaker, C. R., Strope, T. A., & Wilkins, H. M. (2022). Apolipoprotein E and Alzheimer’s disease. *Acta Pharmaceutica Sinica B*, *12*(2), 496–510. https://doi.org/10.1016/J.APSB.2021.10.002

Tumminia, A., Vinciguerra, F., Parisi, M., & Frittitta, L. (2018). Type 2 Diabetes Mellitus and Alzheimer’s Disease: Role of Insulin Signalling and Therapeutic Implications. *International Journal of Molecular Sciences 2018, Vol. 19, Page 3306*, *19*(11), 3306. https://doi.org/10.3390/IJMS19113306

Uddin, M. S., Al Mamun, A., Kabir, M. T., Ashraf, G. M., Bin-Jumah, M. N., & Abdel-Daim, M. M. (2020). Multi-Target Drug Candidates for Multifactorial Alzheimer’s Disease: AChE and NMDAR as Molecular Targets. *Molecular Neurobiology 2020 58:1*, *58*(1), 281–303. https://doi.org/10.1007/S12035-020-02116-9

Ugwu, D. I., & Conradie, J. (2023). Anticancer properties of complexes derived from bidentate ligands. *Journal of Inorganic Biochemistry*, *246*. https://doi.org/10.1016/j.jinorgbio.2023.112268

van Gerven, M., & Bohte, S. (2017). Editorial: Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience*, *11*, 333898. https://doi.org/10.3389/FNCOM.2017.00114/BIBTEX

Van, R., & Fred, L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vatansever, S., Schlessinger, A., Wacker, D., Kaniskan, H. Ü., Jin, J., Zhou, M. M., & Zhang, B. (2021). Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Medicinal Research Reviews*, *41*(3), 1427–1473. https://doi.org/10.1002/MED.21764

Voicu, A., Duteanu, N., Voicu, M., Vlad, D., & Dumitrascu, V. (2020). The rcdk and cluster R packages applied to drug candidate selection. *Journal of Cheminformatics*, *12*(1), 1–8. https://doi.org/10.1186/S13321-019-0405-0/FIGURES/7

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/JOSS.03021

Weggen, S., & Beher, D. (2012). Molecular consequences of amyloid precursor protein and presenilin mutations causing autosomal-dominant Alzheimer’s disease. *Alzheimer’s Research and Therapy*, *4*(2), 1–14. https://doi.org/10.1186/ALZRT107/TABLES/3

WHO. (2023). *Demencia*. https://www.who.int/es/news-room/fact-sheets/detail/dementia

Yang, Y., & Song, W. (2013). Molecular links between Alzheimer’s disease and diabetes mellitus. *Neuroscience*, *250*, 140–150. https://doi.org/10.1016/J.NEUROSCIENCE.2013.07.009

Ye, Q., Hsieh, C. Y., Yang, Z., Kang, Y., Chen, J., Cao, D., He, S., & Hou, T. (2021). A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. *Nature Communications 2021 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-27137-3

Zhang, H., Cao, Y., Ma, L., Wei, Y., & Li, H. (2021). Possible Mechanisms of Tau Spread and Toxicity in Alzheimer’s Disease. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/FCELL.2021.707268

Zhao, W. Q., & Alkon, D. L. (2001). Role of insulin and insulin receptor in learning and memory. *Molecular and Cellular Endocrinology*, *177*(1–2), 125–134. https://doi.org/10.1016/S0303-7207(01)00455-5