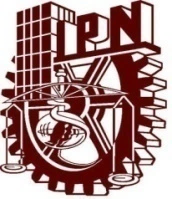
**INSTITUTO POLITÉCNICO NACIONAL**

**CENTRO DE BIOTECNOLOGÍA GENÓMICA**



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

**ASOCIACIÓN ALZHEIMER-DIABETES MELLITUS”**

ANTEPROYECTO DE INVESTIGACIÓN

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

PRESENTA

**ELIUD ULISES AGUILAR DURÁN**

CD. REYNOSA, TAMAULIPAS, MÉXICO JUNIO, 2024

**ÍNDICE**

**Sección Página**

asdasdad

# LISTA DE FIGURAS

**Figura Página**

# LISTA DE SÍMBOLOS Y/O NOMENCLATURA

|  |  |
| --- | --- |
| EA | Enfermedad de Alzheimer |
| APP |  |
| IMSS |  |
| (Aβ) |  |
|  |  |
|  |  |
|  |  |

# INTRODUCCIÓN

1 hoja máximo

# 2. ANTECEDENTES

## 2.1. La demencia

La demencia es un término empleado para diversas enfermedades que son crónicas y progresivas y que resultan en deterioros cognitivos e interfieren en la capacidad de realizar actividades de la vida diaria (OPS, 2023). La demencia suele manifestarse en edades avanzadas, siendo poco común en personas menores de 60 años la mayoría de las veces (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 anciana a nivel mundial. La enfermedad de Alzheimer (EA) es la forma más prevalente de demencia, abarcando entre el 60% y el 70% de los casos. (WHO, 2023). En todo el mundo, más de 55 millones de personas viven con demencia, generando un coste anual de $1 billón de dólares en 2018. En la Región de las Américas, más de 10 millones de personas conviven con esta condición. Las proyecciones indican que el número de personas afectadas por este trastorno se duplicará cada 20 años (OPS, 2023).

## 2.2. Alzheimer

La enfermedad de Alzheimer fue descrita por primera vez en el año 1906 en Alemannia por Alois Alzheimer (Stelzmann et al., 1995). Esta es la demencia neurodegenerativa más prevalente a nivel mundial, se caracteriza por el deterioro progresivo de las funciones mentales. Esta enfermedad afecta las células del cerebro (neuronas), provocando su degeneración y muerte. Las personas afectadas experimentan un deterioro gradual en habilidades cognitivas como la memoria, la orientación, el lenguaje, el aprendizaje, entre otros (IMSS, 2015). En México, se estima que 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. Esta condición afecta principalmente a personas mayores de 65 años (SSA México, 2021).

Las características principales de la enfermedad de Alzheimer (EA) son la formación de placas amiloides y la acumulación de ovillos neurofibrilares. Las placas amiloides están compuestas principalmente por la proteína beta-amiloide (Aβ), que se genera a partir del procesamiento secuencial de la proteína precursora de la beta-amiloide (APP, por sus siglas en inglés "Amyloid Precursor Protein"), (Sun et al., 2012). Por otro lado, los ovillos neurofibrilares son formados por la proteína tau, la cual 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 & Stoothoff, 2004). Existen diversas teorías acerca de la patogénesis del EA. Entre ellas, se encuentran la teoría del péptido amiloide, la hipótesis colinérgica el papel de la proteína Tau y la implicación del estrés oxidativo y el calcio (Sanabria-Castro et al., 2017).

Respecto a la hipótesis de la proteína amiloide, la enfermedad se considera como una serie de anomalías en el procesamiento de la proteína precursora amiloide, donde el desequilibrio entre la producción y eliminación de la beta-amiloide es responsable de la formación anormal de las placas amiloides observadas (Cummings et al., 2007). El péptido beta-amiloide es altamente resistente a la degradación proteolítica y consta de entre 37 y 43 aminoácidos. Las isoformas más comunes son Aβ1-40 y Aβ1-42 (Deane et al., 2009), siendo la forma Aβ1-41 la más hidrofóbica y la que tiene la mayor toxicidad (Mohandas et al., 2009).

El péptido amiloide se origina a través del procesamiento de la proteína precursora de amiloide (APP) en la membrana plasmática, como se ilustra en la figura (Sanabria-Castro et al., 2017). En el proceso normal, la APP es descompuesta principalmente por enzimas con actividad α-secretasa, pertenecientes a las familias de desintegrina y metaloproteasa (ADAM) (De Strooper, 2010). La escisión de la APP por ADAM resulta en la formación y liberación de APPsα, que es soluble bajo ciertas condiciones (Tanzi & Bertram, 2005). En personas con EA, la primera ruptura genera una porción amino terminal más corta (APPsβ) (Velliquette et al., 2005). Esta división es realizada por BACE1, una proteasa transmembrana ubicua con actividad β-secretasa. Después, la γ-secretasa, que es un complejo con 4 subunidades: presenilinas, nicastrina, PEN-2 y APH-1 (Zhang et al., 2014) produce un corte liberando el péptido B-amiloide (Sanabria-Castro et al., 2017). Todo este proceso puede ser observado en la Figura 1.

Diagrama, Esquemático

Descripción generada automáticamente

Figura 1. (Sanabria-Castro et al., 2017)

La hipótesis colinérgica sugiere que la disfunción de las neuronas que contienen acetilcolina contribuye al deterioro cognitivo observado en personas de edad avanzada y en EA (Terry & 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 hallazgos principales que sustentan esta hipótesis radican en que los antagonistas de estos receptores, como la escopolamina, promueven la producción de péptido beta-amiloide y disminuyen la actividad de la α-secretasa (Liskowsky & 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 servido de base para varios tratamientos y, además, hay 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. Sin embargo, esta relación no establece por sí sola una causalidad definitiva de la enfermedad (Pena et al., 2006).

Respecto a la proteína Tau, esta es el componente principal de los filamentos helicoidales (PHF) emparejados que forman los ovillos neurofibrilares (NFT) en el cerebro del EA, y que la tau en los PHF y NFT está anormalmente fosforilada (Grundke-Iqbal et al., 1986).

3. Hablar de qué puede provocar Alzheimer

4. Hablar de la diabetes.

5. Hablar de qué provoca la diabetes.

6. Hablar de datos de la diabetes.

7. Hablar de los productos del Alzheimer

8. Hablar de la relación Alzheimer – diabetes.

9. Proteínas involucradas del artículo.

10. Buscar artículos que respalden a las proteínas del artículo.

11. Hablar sobre el diseño de fármacos

12g. Buscar trabajos sobre diseño de fármacos sobre la asociación AD-DM2.

# 3. JUSTIFICACIÓN

# 4. HIPÓTESIS

Mediante modelos de inteligencia artificial es posible predecir nuevos inhibidores con características similares a inhibidores confirmados experimentalmente para el tratamiento de la asociación Alzheimer-Diabetes Mellitus.

# 5. OBJETIVOS

**5.1. Objetivo general**

Identificar inhibidores multi-blanco con capacidad para modular múltiples blancos relacionados con la asociación entre la enfermedad de Alzheimer y la diabetes mellitus

**5.2. Objetivos específicos**

5.2.1. Recopilar datos experimentales de IC50 sobre inhibidores conocidos para los blancos seleccionados de una base de datos.

5.2.2. Realizar una comparación entre los datos obtenidos para cada blanco y determinar los inhibidores comunes entre los blancos seleccionados.

5.2.3. Calcular descriptores moleculares para cada inhibidor identificado.

5.2.4. Clasificar cada inhibidor como activo o inactivo para cada blanco en función de su valor de IC50 y otros criterios pertinentes.

5.2.5. Desarrollar varios modelos de Machine Learning clasificatorios utilizando los descriptores moleculares de los inhibidores y sus efectos de inhibición para cada blanco y Seleccionar el mejor modelo de Machine Learning basado en su rendimiento.

5.2.6. Aplicar el modelo de Machine Learning seleccionado a una base de datos externa de inhibidores e identificar los mejores inhibidores según los resultados del modelo de Machine Learning.

5.2.7. Realizar un análisis de acoplamiento molecular para las moléculas seleccionadas con los distintos blancos y determinar su puntaje de unión.

5.2.8. Realizar un análisis de dinámica molecular para las moléculas seleccionadas.

# 6. MATERIALES Y MÉTODOS

## 6.1. Selección de las proteínas diana

Se seleccionarán solo las proteínas de la lista propuesta (STAT3, EGFR, IRS1, MAPK1, SRC, HSP90AA1, PIK3R1, UBC, MAPK3 y ESR1) que cuenten con al menos 300 inhibidores en la base de datos ChEMBL (Davies et al., 2015). Aquellas que tengan menos de este valor límite serán descartadas, ya que los modelos de Machine Learning requieren una cantidad adecuada de datos de entrenamiento. Además, dado que se realizará una intersección para analizar la coincidencia de inhibidores descargados para todas las proteínas, es crucial contar con una cantidad suficiente.

## 6.2. Obtención de datos.

Para cada proteína seleccionada, se descargarán moléculas junto con su actividad de inhibición correspondiente empleando el buscador de la base de datos ChEMBL (Davies et al., 2015). En el buscador correspondiente, se introducirá el nombre de cada proteína seleccionada y se descargarán los inhibidores en el formato de actividad estándar IC50, en un archivo CSV.

**6.3. Tratamiento de datos**

Se empleará Python 3.11.0 (Van & Fred, 2009) junto con la biblioteca pandas (The pandas development team, 2020) para llevar a cabo el análisis y procesamiento de los datos recolectados para cada proteína. Los datos serán cargados utilizando la función pd.read\_csv() para cada proteína. Posteriormente, se creará un nuevo DataFrame para cada proteína con el fin de extraer las columnas relevantes del DataFrame original, que incluyen: Molecule ChEMBL ID, Molecule Name, Molecular Weight, Smiles, Standard Type, Standard Relation, Standard Value, y Standard Units. A continuación, se realizará un análisis para identificar las columnas con la mayor cantidad de valores faltantes. Se procederá a limpiar los valores faltantes en las columnas "Standard Value" y "SMILES", dado que son esenciales para el entrenamiento de los modelos de Machine Learning. Finalmente, se llevará a cabo una limpieza de los valores duplicados basada en la columna "Molecule ChEMBL ID". Esto es importante, ya que dos moléculas idénticas podrían tener distintos identificadores ChEMBL ID, lo que podría afectar el rendimiento del modelo. Los DataFrames modificados finales se guardarán en formato CSV.

**6.4. Intersección de los datos**

Una vez realizado el tratamiento para los datos obtenidos de cada proteína, se realizará la intersección de los datos obtenidos de las proteínas empleando la función merge de la librería Pandas. Esta función permite realizar la intersección entre dos DataFrames distintos basándose en una sola columna. La intersección se realizará basándose en la columna “Molecule ChEMBL ID”. El DataFrame resultante tendrá las filas cuyo ChEMBL ID coincida en ambos dataframes. Además, se tendrán las columnas tanto del primer DataFrame como del segundo DataFrame. Se eliminarán las columnas extra generadas y se quedarán únicamente las columnas de Standard Value que corresponde al valor de IC50 de la molécula en cuestión con la proteína en cuestión. Por lo tanto,

1. **Intersección de los inhibidores presentes en todas las proteínas seleccionadas.**
2. **Calcular descriptores moleculares para las moléculas.**
3. **Selección de mejores características.**
4. **Construcción de modelos que sean capaces de predecir inhibidores.**
5. **Selección del mejor modelo.**
6. **Aplicar a una base de datos externa a los datos de entrenamiento.**
7. **Docking molecular para selección de los mejores candidatos.**
8. **Dinámica molecular.**

# 7. BIBLIOGRAFÍA

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

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 Strooper, B. (2010). Proteases and proteolysis in Alzheimer disease: a multifactorial view on the disease process. *Physiological Reviews*, *90*(2), 465–494. https://doi.org/10.1152/PHYSREV.00023.2009

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

Velliquette, R. A., O’Connor, T., & Vassar, R. (2005). Energy inhibition elevates beta-secretase levels and activity and is potentially amyloidogenic in APP transgenic mice: possible early events in Alzheimer’s disease pathogenesis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *25*(47), 10874–10883. https://doi.org/10.1523/JNEUROSCI.2350-05.2005

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

Zhang, X., Li, Y., Xu, H., & Zhang, Y. W. (2014). The γ-secretase complex: from structure to function. *Frontiers in Cellular Neuroscience*, *8*(DEC). https://doi.org/10.3389/FNCEL.2014.00427