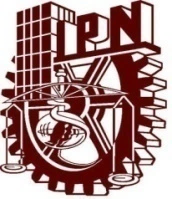
**INSTITUTO POLITÉCNICO NACIONAL**

**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 |
|  |  |
|  |  |
|  |  |

# 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

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).

La enfermedad de Alzheimer se distingue por la presencia conjunta de dos tipos de lesiones en el cerebro: las placas extracelulares de beta amiloide (Aβ), también conocidas como placas seniles (SP), y los ovillos neurofibrilares intracelulares (NFT), los cuales contribuyen a la atrofia cerebral (Farooqui, 2016).

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 potenciales inhibidores multi-blanco para el tratamiento de la asociación Alzheimer-Diabetes mellitus

**5.2. Objetivos específicos**

5.2.1. Descargar los valores de

5.2.2.

5.2.3.

5.2.4.

5.2.5.

# 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.**
9. **Análisis de toxicidad.**

# 7. BIBLIOGRAFÍA

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

Farooqui, A. A. (2016). Neurochemical Aspects of Neurological Disorders. *Trace Amines and Neurological Disorders: Potential Mechanisms and Risk Factors*, 237–256. https://doi.org/10.1016/B978-0-12-803603-7.00016-1

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

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

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

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

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

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

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