

Desarrollo y programación de herramientas informáticas para el análisis y manipulación de archivos *PDBs* para proveer análisis rápidos de biología estructural

ERIK HAZEL VILCHIS MACIAS¹

¹Unidad de Bioinformática del IFC

*macias_neuro@hotmail.com

Compiled May 12, 2023

El proyecto se enfoca en el desarrollo y programación en Python de herramientas informáticas para analizar y manipular archivos *PDBs* de manera rápida y eficiente. Estas herramientas permitirán un análisis detallado de la biología estructural, lo que resultará de gran utilidad en para acortar los tiempos en investigaciones como el diseño y desarrollo de fármacos, investigaciones sobre la estructura de un grupo determinado de proteínas, etc.

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. IMPORTANCIA

La base de datos de proteínas (*PDB*) es un recurso crítico y altamente valorado por la comunidad científica, ya que proporciona información detallada sobre estructuras de proteínas y complejos proteicos. La necesidad de herramientas avanzadas para el procesamiento y visualización de esta información es evidente, y es por ello que el presente proyecto se enfoca en desarrollar una herramienta que pueda agilizar el proceso de análisis, fomentar la colaboración, aumentar la reproducibilidad y mejorar la eficiencia de la investigación.

La herramienta en cuestión será capaz de procesar grandes cantidades de datos de forma rápida y precisa, lo que permitirá a los investigadores ahorrar tiempo y recursos valiosos. Además, su interfaz de usuario intuitiva y fácil de usar, facilitará la colaboración entre los miembros del equipo, lo que resultará en una investigación más eficaz y detallada.

El desarrollo de esta herramienta no solo contribuirá al avance de la ciencia, sino que también tendrá un impacto significativo en la solución de problemas importantes en la sociedad, tales como la lucha contra enfermedades graves y el diseño de fármacos más efectivos. En definitiva, este proyecto representa un paso importante en la mejora de la eficiencia, colaboración y confiabilidad de la investigación científica, lo que tendrá un impacto positivo en la sociedad en general.

2. OBJETIVO

Realizar scripts ejecutables en Python que permitan extraer y procesar información de archivos *PDB* actualizados

3. OVERVIEW DEL PROCESO

En primer lugar, se debe acceder a la API de la base de datos y filtrar los archivos *PDB* para seleccionar solo los que se desean descargar. Luego, se deben descargar los archivos seleccionados y almacenarlos en una ubicación específica en el sistema de archivos. Además, se debe crear un registro de los archivos descargados para evitar la descarga de archivos duplicados en el futuro. Se debe verificar periódicamente si hay nuevos archivos disponibles en la base de datos y descargarlos si es necesario.

Para extraer información sobre las cadenas de proteínas, se deben leer los archivos *PDB* utilizando la biblioteca *biopython* o scripts ad hoc y analizar la estructura de cada archivo para extraer información sobre las cadenas de proteínas. Luego, se deben escribir los resultados obtenidos en un archivo de salida que incluya información sobre el nombre de cada archivo *PDB*, el número de cadenas, el tamaño de cada cadena y el tipo de cadena.

Por último, se recomienda utilizar *Git* y *Github* para realizar un seguimiento de las modificaciones del código y crear un video tutorial sobre el funcionamiento del script utilizando *Adobe After Effects* y *Wondershare Filmora*.

4. QUICK OVERVIEW

El *PDB* (*Protein Data Bank*) es una base de datos pública y centralizada que contiene estructuras tridimensionales de proteínas y otras moléculas biológicas, utilizada en el diseño de drogas y la investigación en campos como la química computacional y la ingeniería de proteínas. Python y *Git* son herramientas que permiten manipular y extraer información de los archivos *PDB* de manera eficiente y organizada. El análisis de los archivos *PDB* permite estudiar la estructura tridimensional de proteínas, identificar posibles sitios de unión para drogas o compuestos terapéuticos, y es una técnica fundamental en la investigación de la biología molecular y la biotecnología. El análisis del número y tamaño de las cadenas en los archivos *PDB* ayuda a comprender la organización, estructura, función y actividad biológica de la proteína, y se puede utilizar para identificar proteínas con múltiples subunidades o dominios específicos. Las herramientas que permiten obtener estos datos básicos de manera sencilla son relevantes.