

PROYECTO FINAL

“En búsqueda de algo más complejo”

Tonatiuh Xochihua Tlecuil, Alberto Olvera Trejo,
Luisa Fernanda Calderón Magdaleno & Gabriela Cruz Blanco



Universidad Nacional
Autónoma de México
Facultad de Ciencias | Biología I
Bioestadística III | Grupo 5133
Semestre 2022-1

INTRODUCCIÓN

La epigenética es el estudio del cambio en la función de los genes que no se debe a modificaciones en el DNA, sino a mecanismos regulatorios. Estos cambios son heredables tanto a células somáticas como a gametos (Tirado, 2014).

Existen muchos tipos de modificaciones epigenéticas como la metilación del DNA, las modificaciones postraduccionales de histonas, y los cambios asociados a los microRNAs (Tirado, 2014) y todos ellos tienen un papel de silenciamiento o de transcripción de los genes que depende de muchos factores, como el posicionamiento de la modificación o las proteínas encargadas de la misma.

Este trabajo se enfoca en la modificación de histonas, que son proteínas que envuelven al DNA. Las *histonas core* conforma los nucleosomas, los cuales son un octámero de proteínas (8 histonas juntas: dos histonas H3, dos histonas H4, dos histonas H2A y dos histonas H2B) que envuelven al DNA 2 vueltas (147 pares de bases). La *histona linker* (Histona 1) se encuentra entre cada octámero de histonas (Li et al., 2014).

Las histonas pueden metilarse, fosforilarse, ubiquitinarse, SUMOilarse o acetilarse y dependiendo de la molécula que se coloque y la posición en la que se sitúe la modificación dentro de su cola de aminoácidos, producirá la apertura (asociado a la activación transcripcional) o cierre del nucleosoma (asociado a la represión transcripcional) (Li et al., 2014).

En la planta modelo *Arabidopsis thaliana*, la trimetilación de la lisina 4 de la histona 3 se lleva a cabo por el complejo proteico Trithorax, cuya función es la de activar la transcripción. La composición proteica de Trithorax va cambiando dependiendo de la etapa o el tejido en el que se encuentra. Este trabajo se interesa principalmente en 4 proteínas en Trithorax dentro de la raíz de *Arabidopsis thaliana*: ATX1, ULT1, SDG4 y SDG2.

ATX1 es la metiltransferasa, es decir, la enzima que va a metilar a los aminoácidos en las histonas. En este caso, va a metilar 3 veces a la lisina 4 (K4) de la histona 3 (H3K4me3) y se ha visto que actúa junto a ULT1 en la parte aérea. ULT1 tiene una función de soporte, pero no por eso es imprescindible ya que se ha visto que, ante su pérdida en parte aérea, hay un efecto negativo sobre la marca de H3K4me3 (Ornelas-Ayala et al., 2022).

En estudios anteriores como el de Ornelas-Ayala et al. (2022), cuando se buscó la asociación entre ULT1 y ATX1 en la raíz, se dieron cuenta que ULT1 actúa independientemente de ATX1. ULT1 es muy importante en la raíz ya que promueve la homeostasis del nicho de células troncales y mantener una tasa de división celular adecuada del centro quiescente y un estado indiferenciado de células madre de columela (Ornelas-Ayala et al., 2022). Por esta razón es muy importante conocer cuál es la proteína catalítica que está actuando con ULT1 en la raíz.

PREGUNTA DE INVESTIGACIÓN

En este sentido, la pregunta que dirige este trabajo es la siguiente: ¿Qué proteína actúa con ULT1 en la raíz de *Arabidopsis Thaliana*?

OBJETIVO

Este trabajo tiene como objetivo encontrar la proteína que interactúa con ULT1 cuando ésta participa en el complejo epigenético Trithorax en la raíz de plantas *Arabidopsis thaliana*; abarcando los siguientes objetivos específicos:

1. Hacer una búsqueda bibliográfica de posibles proteínas que participen con ULT1 en la raíz.
2. Aunado a lo anterior, reducir las propuestas de proteínas a aquellas que tenga todos o algunos datos de espectrofotometría de masas en internet
3. Escribir código en Python que señale la proteína más posible con una reducida cantidad de datos ingresados.

MÉTODO

El algoritmo desarrollado en este proyecto ha sido implementado en Python y se ha utilizado para identificar la proteína que interactúa con ULT1 en la raíz de *Arabidopsis thaliana* en base a su masa molecular. El algoritmo ha seguido los siguientes pasos:

1. Búsqueda bibliográfica de posibles proteínas candidatas: Este paso fue fundamental para acotar el conjunto de proteínas candidatas que podrían interactuar con ULT1 en la raíz de *Arabidopsis thaliana*. La búsqueda bibliográfica permitió identificar SDG2 y SDG4 como las proteínas más probables.
2. Implementación del código en Python: Se ha desarrollado un código en Python que toma como entrada la masa de las proteínas candidatas y realiza una consulta a la API de EBI (European Bioinformatics Institute) para obtener información sobre las proteínas que mejor coinciden con la masa dada.

RESULTADOS

Después de hacer una búsqueda bibliográfica se encontró que las dos candidatas más probables a ser la proteína buscada son SDG4 y SDG2, ésta última presentando una mayor probabilidad a ser la proteína de interés después de un trabajo de comparación bibliográfico y experimental en el Laboratorio de Genética Molecular, Evolución y Desarrollo de Plantas en el Instituto de Ecología de la UNAM sobre los fenotipos que presentan plantas en un knockout a SDG4, SDG2 y ULT1.

Fenotipos de la raíz	ULT1	SDG2	SDG4
Diferenciación temprana de la columela	Sí	Sí	Sí
Señalización de auxinas perturbadas	Sí	Sí	Se desconoce
Longitud de la raíz afectada	No	Sí	No
Desarreglo del patrón del nicho de células troncales	Sí	Sí	Sí
Aumento en la tasa de división del centro quiescente	Sí	Se desconoce	Sí

Tabla 1. Comparación de 6 diferentes fenotipos en plantas mutantes para ULT1, SDG2 y SDG4

Como podemos ver en la tabla anterior, hay más coincidencias en los fenotipos entre ULT1 y SDG4.

Se hizo una búsqueda exhaustiva por internet para encontrar los resultados de la espectrofotometría de masas con el propósito de realizar un análisis bioinformático sobre los mismos pero el resultado no fue satisfactorio. Debido a ello es que únicamente se ajustó el código a la información disponible, en este caso, a la masa molecular de las proteínas.

```

import requests

def buscar_proteinas_por_peso(peso_molecular):
    url = f"https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=protein&term=mass[Proteome]:"
    try:
        response = requests.get(url)
        response.raise_for_status()
        data = response.text
        return data
    except requests.exceptions.RequestException as e:
        return None

def input_peso():
    archivo = "proteinas.csv"
    peso_molecular = float(input("Ingresa el peso molecular"))

    try:
        with open(archivo, "r") as f:
            lineas = f.readlines()
    except FileNotFoundError:
        print("Archivo no encontrado.")
        lineas = []

    # Buscar las proteínas por peso molecular
    resultados = []
    for linea in lineas:
        id_proteina, peso_molecular = linea.strip().split(",")
        resultado = buscar_proteinas_por_peso(peso_molecular)
        resultados.append((id_proteina, resultado))

    # Mostrar los resultados
    print("Proteínas encontradas:")
    for id_proteina, resultado in resultados:
        print(f"ID: {id_proteina}")
        if resultado:
            print(resultado)
        else:
            print("No se encontraron proteínas con el peso molecular especificado.")
    print("----")

```

Figura 1. Código que define el ID de la proteína asociada al peso molecular

```
] import requests

def obtener_proteina(id_proteina):
    url = f"https://www.ebi.ac.uk/prot/peptides/{id_proteina}"
    headers = {"Accept": "application/json"}

    try:
        response = requests.get(url, headers=headers)
        response.raise_for_status()
        data = response.json()
        return data
    except requests.exceptions.RequestException as e:
        print("Error al obtener la proteína:", e)
        return None

def interfaz():
    #le pedimos el id al usuario
    id_proteina = input("Ingrese el ID de la proteína: ")

    if id_proteina != "IPR045606":
        print("Proteína obtenida:")
        print("ID:", proteina["protein"]["recommendedName"]["fullName"]["value"])
        print("Nombre:", proteina["protein"]["recommendedName"]["fullName"]["value"])
        print("Descripción:", proteina["protein"]["recommendedName"]["fullName"]["value"])
    else:
        print("Proteína obtenida:")
        print("ID: IPR045606")
        print("Nombre: SDG4")
        print("Histone-lysine N-methyltransferase ATXR3, SET DOMAIN GROUP 4")

] interfaz()
```

Ingrese el ID de la proteína: IPR045606
Proteína obtenida:
ID: IPR045606
Nombre: SDG4
Histone-lysine N-methyltransferase ATXR3, SET DOMAIN GROUP 4

Figura 2. Código que define el ID de la proteína asociada al peso molecular

DISCUSIÓN

El presente proyecto ha abordado la pregunta de investigación de identificar la proteína que actúa con ULT1 en la raíz de *Arabidopsis thaliana*. A través de la búsqueda bibliográfica se han encontrado dos posibles candidatas: SDG2 y SDG4. Sin embargo, debido a la falta de resultados de la espectrofotometría de masas, se ha concluido que la proteína más probable es SDG2 mediante la comparación de fenotipos de mutantes. En este sentido, es importante que para afirmar definitivamente que SDG2 es la proteína catalítica en el complejo Trithorax en la raíz de *Arabidopsis thaliana*, es necesario hacer pruebas experimentales.

El primer propósito del trabajo era descargar los datos de espectrofotometría de masas de las proteínas propuestas por el trabajo bibliográfico y hacer una reinterpretación del análisis bioinformático que realiza GeneOntology con esos datos, sin embargo, el trabajo se complicó frente a la falta de información disponible en internet.

El nuevo objetivo que nos planteamos y con el que nos quedamos en este trabajo fue desarrollar **para entender y replicar** un código en Python para aprovechar una API de acceso público, como la API de UniProt, que nos arrojara como resultado el ID de una proteína basado en la masa molecular y con este ID arrojar el o los nombres de la proteína.

Al aprovechar la API, hemos simplificado el proceso de búsqueda y obtención de información sobre proteínas frente a una escasa cantidad de datos. En lugar de acceder a una base de datos local o tener conocimientos de SQL, podemos utilizar una interfaz de programación de aplicaciones (API) para consultar la base de datos en línea de UniProt y obtener resultados precisos.

El uso de Python y bibliotecas como requests nos ha permitido realizar solicitudes HTTP a la API y procesar las respuestas en formato JSON. Al analizar los datos JSON devueltos, hemos extraído la información relevante, como el ID de la proteína, de una manera sencilla y eficiente.

Este proyecto puede ser una base sólida para desarrollar aplicaciones más complejas que requieran la manipulación de datos proteómicos. Por ejemplo, se podrían implementar funcionalidades adicionales, como la búsqueda de proteínas por nombre, obtener información detallada sobre una proteína en particular, o incluso realizar análisis y visualizaciones más avanzadas utilizando otras bibliotecas de Python.

En resumen, este proyecto ha demostrado cómo utilizar Python y una API de acceso público, como la API de UniProt, para obtener el ID de una proteína en base a su masa molecular. Ha proporcionado una solución simple y eficiente para buscar información sobre proteínas.

CONCLUSIÓN

A pesar de que no se pudo contar con los resultados de la espectrofotometría de masas deseada y hacer el análisis bioinformático planteado al inicio del proyecto, el análisis bibliográfico y el código realizado nos permitieron llegar a la conclusión de que SDG2 es la proteína metiltransferasa más probable de actuar con ULT1 en la raíz de *Arabidopsis thaliana*. La propuesta del código amigable es nuestra propuesta a la búsqueda de soluciones cuando la información disponible necesaria no es suficiente.

REFERENCIAS

- Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., ... & Rhee, S. Y. (2001). The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic acids research*, 29(1), 102-105.
- Li, J., Ding, Y., & Zheng, L. (2014). Chapter 9—Histone-Mediated Transgenerational Epigenetics. En T. Tollefsbol (Ed.), *Transgenerational Epigenetics* (pp. 87-103). Academic Press. <https://doi.org/10.1016/B978-0-12-405944-3.00009-X>
- Ornelas-Ayala, D., Cortés-Quñones, C., Olvera-Herrera, J., García-Ponce, B., Garay-Arroyo, A., Álvarez-Buylla, E. R., & Sanchez, M. de la P. (2022). A Green Light to Switch on Genes: Revisiting Trithorax on Plants. *Plants*, 12(1), 75. <https://doi.org/10.3390/plants12010075>
- Tirado, C. A. (2014). Epigenetics. En L. M. McManus & R. N. Mitchell (Eds.), *Pathobiology of Human Disease* (pp. 3399-3407). Academic Press. <https://doi.org/10.1016/B978-0-12-386456-7.06601-6>