

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 situó 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 quiascente 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.

En el Laboratorio de Genética Molecular, Evolución y Desarrollo de Plantas tenemos la hipótesis de que la ULT1 trabaja con alguna de las siguientes metiltransferasas: SDG2 o SDG4. Esto debido a que, cuando se hace knockdown de cualquiera de estas proteínas en la plantas, presentan fenotipos muy similares al knockdown de ULT1.

PREGUNTA DE INVESTIGACIÓN

En este sentido, la pregunta que dirige este trabajo es la siguiente: ¿qué proteína actúa con ULT1 en la raíz de 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 en con ULT1 y ULT2
2. Aunado a lo anterior, proponer una proteína que tenga sus datos de espectrofotometría de masas en internet
3. Realizar un análisis bioinformático de espectrofotometría de masas
4. Describir un algoritmo asociado al análisis anterior

MÉTODO

El primer paso fue hacer una búsqueda de las posibles proteínas que participen con ULT1 pues de esa manera se pudo acotar el conjunto de las proteínas candidatas y así únicamente se limitó el análisis a las proteínas candidatas más fuertes.

Posteriormente, se implementó un código en Python, el cual tomó como entrada la masa de las proteínas, se asoció un diccionario de las posibles proteínas candidatas para hacer la consulta a la API de EBI y regresa como resultado la información de la proteína que más coincidencia tiene.

Además, se agregó una pequeña interfaz interactiva de manera que el programa le pide los datos al usuario, es decir, el usuario no tiene que tocar el código para poder correrlo, únicamente tiene que teclear la información disponible y obtendrá los resultados.

```
def interfaz():
    min_mass = int(input('Escribe la masa mínima del rango \n'))
    max_mass = int(input('Escribe la masa máxima del rango \n'))
    domain = input('***Opcionalmente escribe el nombre del dominio de la proteína. Da Enter si quieres omitir \n')
    score_threshold = int(input('***Opcionalmente especifica el umbral de puntuación. Da Enter si quieres omitir \n'))

    proteins = search_protein_by_mass(min_mass, max_mass, domain, score_threshold)

    #Si se encontraron proteínas entonces imprimimos su ID y nombre
    if proteins:
        print("Proteínas encontradas:")
        for protein in proteins:
            print("ID: ", protein[0])
            print("Nombre: ", protein[1])
            print()
    else:
        print("No se encontraron proteínas que cumplan los criterios especificados.")
```

Figura 1. Muestra del código.

RESULTADOS

Después de hacer una búsqueda bibliográfica se encontró que las dos candidatas más probables a ser la proteína buscada son SDG2 y SDG4, ésta última presentando una mayor probabilidad a ser la proteína de interés.

Se hizo una búsqueda exhaustiva por internet con el fin de poder encontrar los resultados de la espectrofotometría pero el resultado no fue satisfactorio. Debido a ello es que únicamente se contempló la posibilidad de que la proteína fuera SDG4 o bien SDG2 y se ajustó el código a la información disponible. Usando la información anterior, se alimentó el código ya mencionado y se obtuvo como resultado que la proteína buscada es SDG4

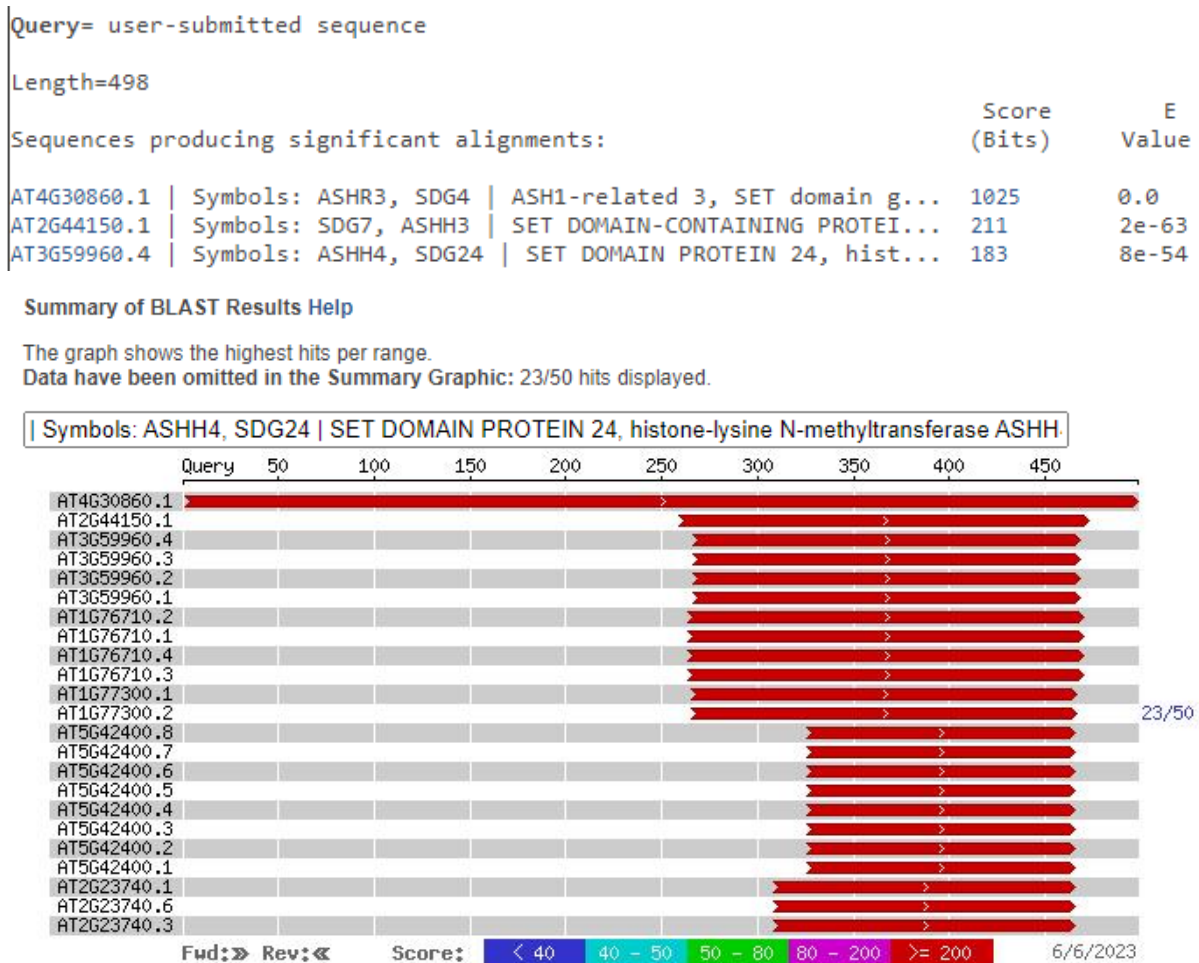


Figura 2. Resultados de BLAST. (Fuente: The Arabidopsis Information Resource (TAIR))

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. Esta funcionalidad puede ser útil en diversas aplicaciones relacionadas con la investigación y el análisis de proteínas.

Al aprovechar la API, hemos simplificado el proceso de búsqueda y obtención de información sobre proteínas. 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, y puede servir como punto de partida para futuros desarrollos en el campo de la investigación y el análisis proteómico.

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 y el análisis bioinformático, 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.

Es importante destacar que la identificación de la proteína SDG2 como la posible interactora de ULT1 se basa en la información bibliográfica disponible y en la búsqueda en bases de datos públicas. Sin embargo, la falta de resultados de la espectrofotometría de masas limita la confirmación experimental de esta interacción. Por lo tanto, es necesario considerar esta limitación y la posibilidad de realizar experimentos adicionales, como ensayos de co-inmunoprecipitación o técnicas de co-localización, para validar la interacción entre ULT1 y SDG2.

Es importante mencionar que ULT1 y SDG2 están involucradas en el complejo epigenético Trithorax, el cual juega un papel fundamental en la regulación de la expresión génica y el desarrollo de las plantas. La identificación de la interacción entre ULT1 y SDG2 en la raíz de *Arabidopsis thaliana* podría tener implicaciones significativas en la comprensión de los mecanismos moleculares y epigenéticos involucrados en el desarrollo de las raíces y en la regulación de la expresión génica.

Además, el desarrollo de un código en Python para la identificación de proteínas basado en la masa molecular ha demostrado ser una herramienta útil en la búsqueda y selección de candidatas. Este enfoque bioinformático puede ser aplicado en futuros estudios para

analizar grandes conjuntos de datos proteómicos y agilizar el proceso de identificación de proteínas de interés.

En cuanto a las limitaciones del estudio, es importante destacar la falta de resultados de la espectrofotometría de masas, lo cual restringió la cantidad de información disponible para respaldar la identificación de la proteína SDG2 como la interactora de ULT1. Además, la disponibilidad de información bibliográfica sobre las interacciones específicas de ULT1 en la raíz de *Arabidopsis thaliana* puede ser limitada, lo que dificulta la comparación y validación de los resultados obtenidos.

En resumen, aunque este proyecto ha logrado identificar a SDG2 como la proteína más probable que interactúa con ULT1 en la raíz de *Arabidopsis thaliana*, es necesario considerar las limitaciones mencionadas y realizar estudios experimentales adicionales para confirmar esta interacción. Los resultados obtenidos hasta ahora brindan una base sólida para futuras investigaciones en el campo de la genética molecular y el desarrollo de plantas, y el enfoque bioinformático utilizado puede ser aplicado en otras investigaciones relacionadas con la identificación de proteínas de interés.

Conclusión

A pesar de que no se pudo contar con los resultados de la espectrofotometría de masas deseada, el análisis hecho y el código creado nos permitieron llegar al objetivo del proyecto, el cual era identificar la proteína desconocida que actúa con ULT1, la cual sabemos que es

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>

