



# Lab 118: Preparación para analizar la insulina con Python

link: <https://awsrestart.instructure.com/courses/1632/modules/items/886830>

## Información general sobre el laboratorio

En tecnología de la información, Python funciona bien como el lenguaje de programación preferido para la manipulación de cadenas, secuencias y números. Python es el lenguaje favorito en aplicaciones científicas, como la física, la química y la biología.

En algunos de los laboratorios de los módulos de Python, realizará manipulaciones y cálculos de secuencias sencillos con insulina, que es una hormona conocida del cuerpo humano responsable de regular los niveles de azúcar.

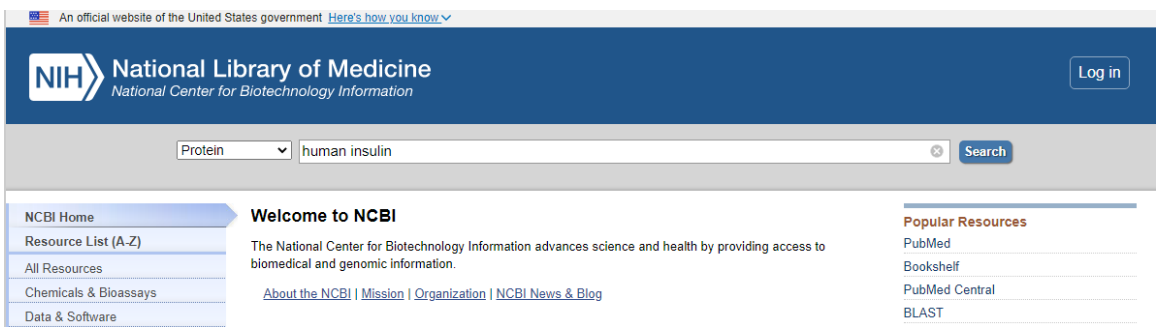
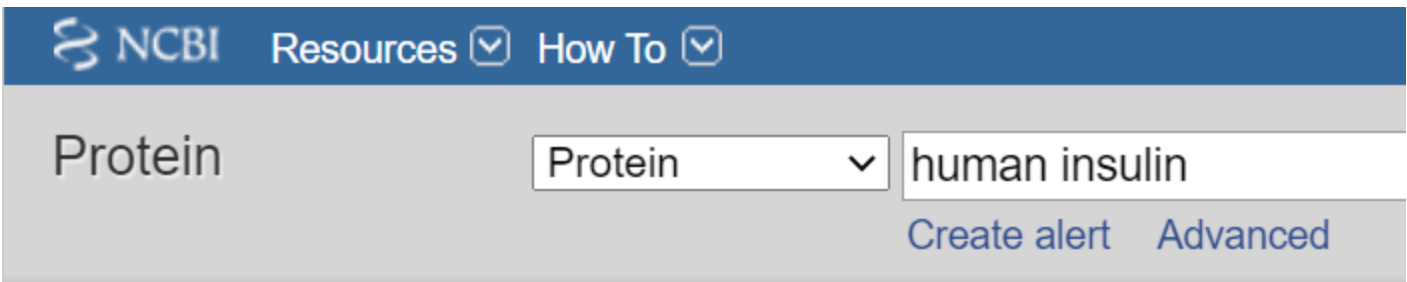
En este laboratorio, deberá realizar lo siguiente:

- recuperar la secuencia de proteínas de la insulina humana a partir de la preproinsulina humana

## Ejercicio 1: Recuperación de la secuencia de proteínas de la preproinsulina humana

El Centro Nacional para la Información Biotecnológica (NCBI) cuenta con información sobre muchas secuencias biológicas.

- Acceda a **NCBI** en <https://ncbi.nlm.nih.gov>.
- Junto a la barra de búsqueda, elija el menú desplegable y seleccione **Protein** (Proteína). A continuación, en la barra de búsqueda, escriba **human insulin** y elija **Search** (Buscar).



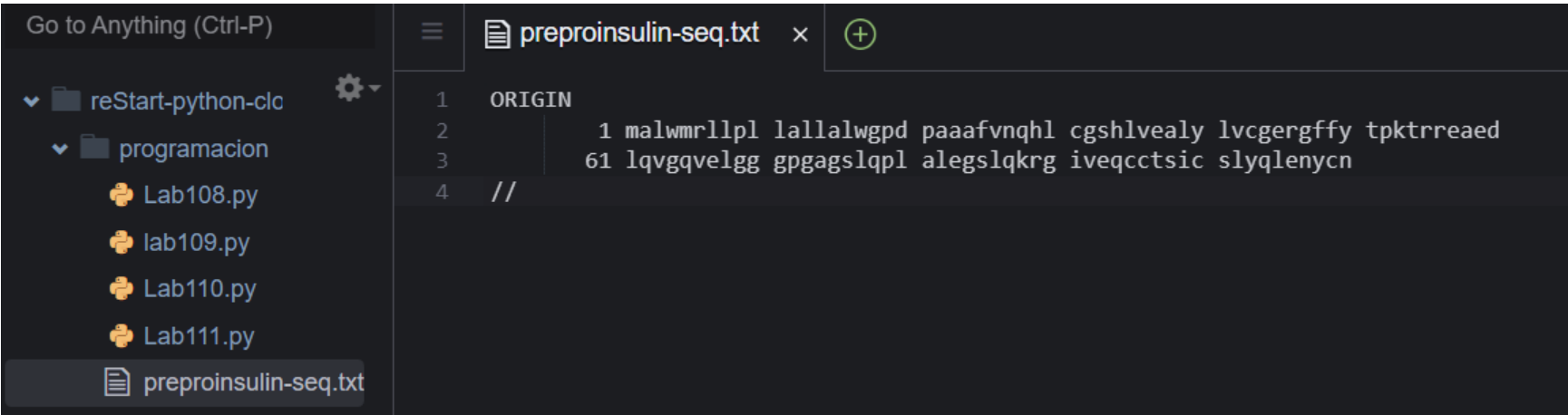
- Elija el siguiente resultado de búsqueda: **insulin [Homo sapiens]** (insulina Homo sapiens).

☐ [insulin \[Homo sapiens\]](#)  
3. 110 aa protein  
Accession: AAA59172.1 GI: 386828  
[Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

4. En la parte inferior del registro de búsqueda, copie la secuencia de insulina, que comienza con la palabra `ORIGIN` y termina con `//`.

```
ORIGIN
1 malwmrllpl lallalwgpdpaaafvnqhl cgshlvealy lvcgergffy tpktrreaed
61 lqvqqvelgg gpgagslqpl alegslqkrgiveqcctsic slyqlenycn
//
```

5. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *preproinsulin-seq.txt*.
6. Pegue la secuencia de la insulina en **preproinsulin-seq.txt**:



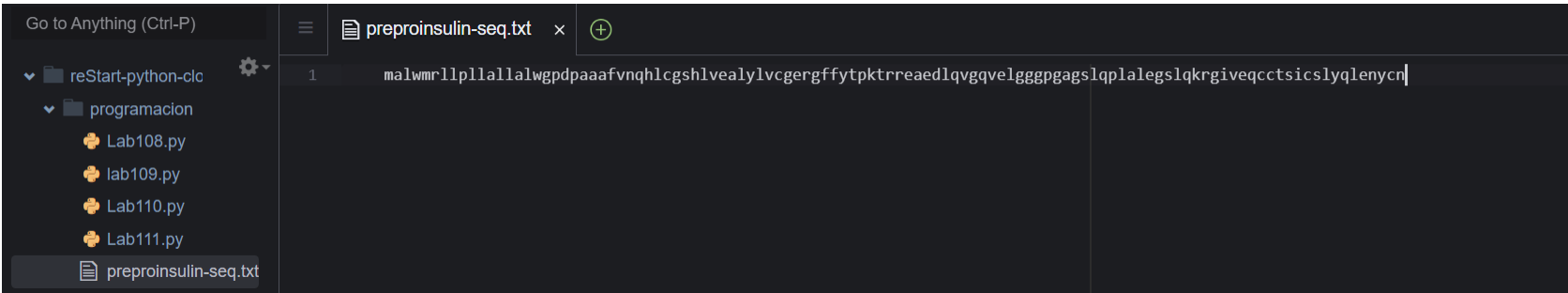
Adicional: Limpieza de preproinsulin-seq.txt mediante programación

La limpieza de archivos de datos iniciales es una tarea común en programación. Puede realizar la limpieza de preproinsulin-seq.txt mediante programación de varias formas, por ejemplo, con el uso de Bash, Python u otro lenguaje de programación de su elección. Pruebe utilizar expresiones regulares para eliminar mediante programación el archivo de *ORIGIN*, sus números, las dos barras diagonales (*//*), los espacios y los saltos de línea o los carros de retorno. También puede confirmar mediante programación que el archivo tiene 110 caracteres.

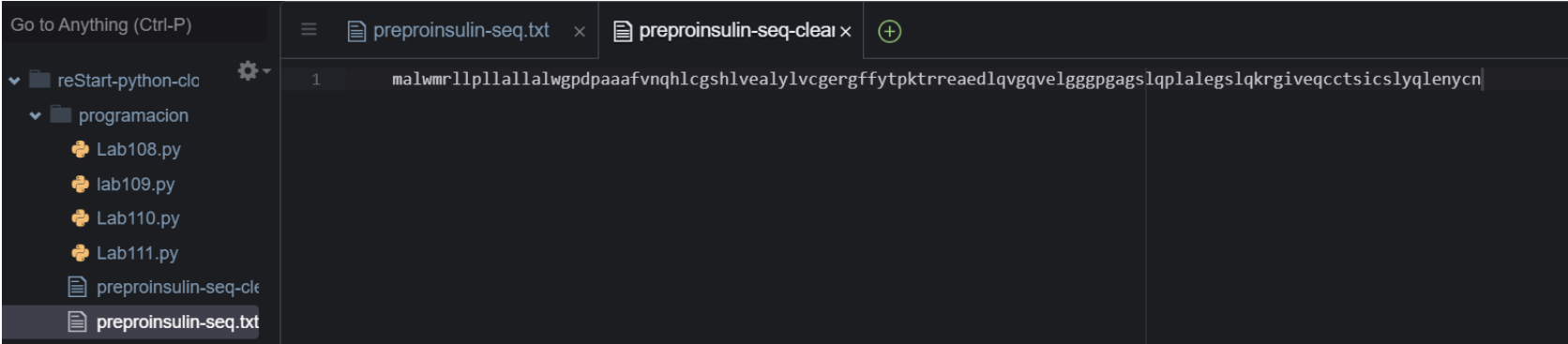
Ejercicio 2: Obtención de la secuencia de proteínas de la insulina humana

La insulina se obtiene de la preproinsulina mediante una serie de procedimientos de corte y pegado. La preproinsulina contiene una secuencia de señal 24aa y una molécula de proinsulina 86aa. Los aminoácidos 25–54 y los aminoácidos 90–110 son la molécula de la insulina procesada. Utilice Python, Bash o manipulación manual para recuperar solo los aminoácidos en la secuencia que componen la insulina.

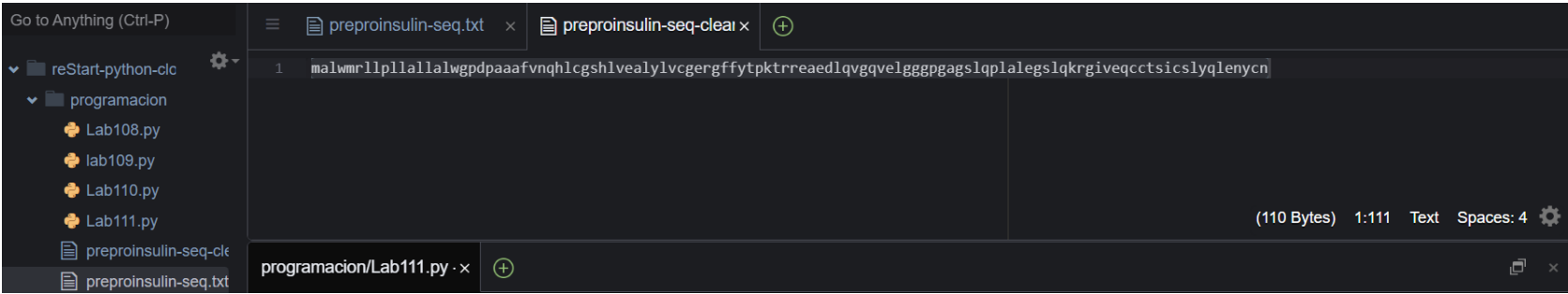
1. Elimine de forma manual o mediante programación *ORIGIN*, *1*, *61*, *//*, así como los espacios y los retornos de carro.



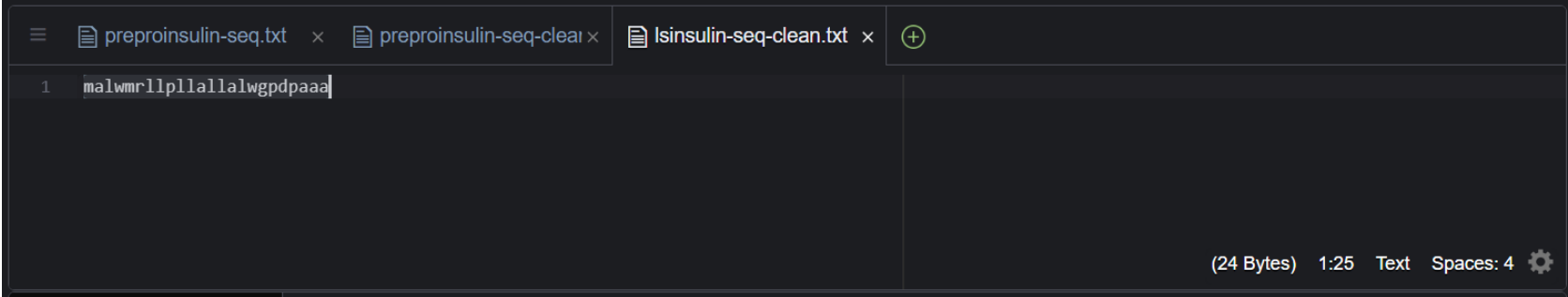
2. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *preproinsulin-seq-clean.txt*



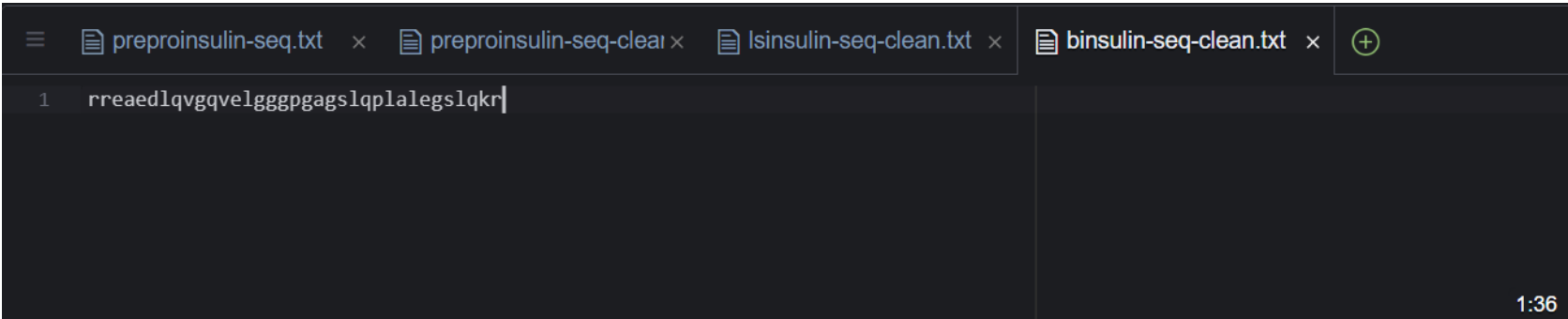
- 3. Copie los resultados en el archivo **preproinsulin-seq-clean.txt**.
- 4. Confirme que el archivo tenga 110 caracteres en minúscula, que representan los aminoácidos en la secuencia de la preproinsulina.



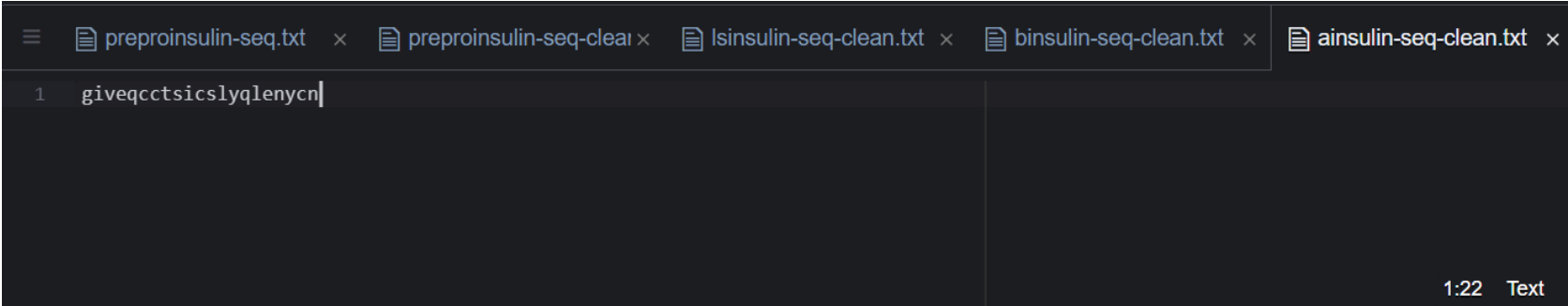
- 5. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *lsinsulin-seq-clean.txt*.
- 6. Guarde los aminoácidos 1–24 en **lsinsulin-seq-clean.txt**. Verifique que el archivo tenga 24 caracteres.



- 7. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *binsulin-seq-clean.txt*.
- 8. Guarde los aminoácidos 25–54 en **binsulin-seq-clean.txt**. Verifique que el archivo tenga 30 caracteres.
- 9. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *cinsulin-seq-clean.txt*.
- 10. Guarde los aminoácidos 55–89 en **cinsulin-seq-clean.txt**. Verifique que el archivo tenga 35 caracteres.



- 11. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *ainsulin-seq-clean.txt*.
- 12. Guarde los aminoácidos 90–110 en **ainsulin-seq-clean.txt**. Verifique que el archivo tenga 21 caracteres.



### Decisión sobre cuándo automatizar y cuándo trabajar de forma manual: un debate sobre el alcance y el tiempo

Automatizar el trabajo frente a trabajar manualmente es un dilema de los programadores. Demasiada automatización desperdicia tiempo en la codificación, mientras que muy poca automatización restringe el alcance de su programa. Intente equilibrar la automatización con el trabajo manual en un esfuerzo por crear un programa con el mayor alcance posible y con el menor tiempo dedicado a la codificación. En este caso, probablemente no valga la pena el tiempo

adicional de codificación para limpiar insulin-seq.txt a insulin-seq-clean.txt mediante programación. Sin embargo, si necesita descargar miles o millones de archivos y hacer la misma tarea, sería bueno explorar la automatización.

¡Felicitaciones! Ha preparado los datos para su procesamiento posterior. La preparación manual de estos archivos debería ayudarlo a apreciar la automatización que Python puede proporcionar.