



118- [PF] - Lab - Preparación para analizar la insulina

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.

1. Acceda a **NCBI** en <https://ncbi.nlm.nih.gov>.
2. 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).

NCBI Resources How To

Protein Protein human insulin

Create alert Advanced

1. 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](#)

1. 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 lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed
     61 lqvqqvelgg gpgagslqpl aegslqkrg iveqcctsic slyqlenycn
//
```

1. 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*.
2. Pegue la secuencia de la insulina en **preproinsulin-seq.txt**:

```
ORIGIN
      1 malwmrllpl lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed
     61 lqvqqvelgg gpgagslqpl aegslqkrg iveqcctsic slyqlenycn
//
```

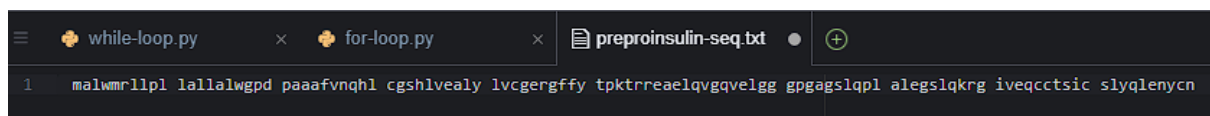
Adicional: Limpieza de preproinsulin-seq.txt mediante programación

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

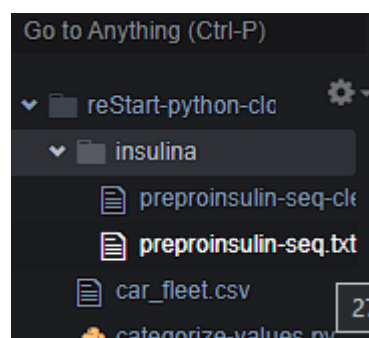
La insulina se obtiene de la preproinsulina mediante una serie de procedimientos de corte y pegado. La preproinsulina contiene una secuencia de señal 24aa y una molécula de proinsulina 86aa. Los aminoácidos 25–54 y los aminoácidos 90–110 son la molécula de la insulina procesada. Utilice Python, Bash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

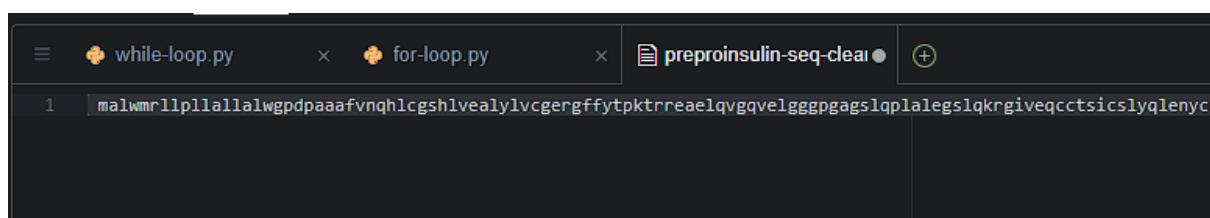


```
1 malwmrllpl lalllalwgpdpaaafvnqhl cgshlvealy lvcgergffy tpktrreaelqvgqvelgg gpgagslqpl alegsllqkrq iveqcctsic slyqlenycn
```

1. 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*.



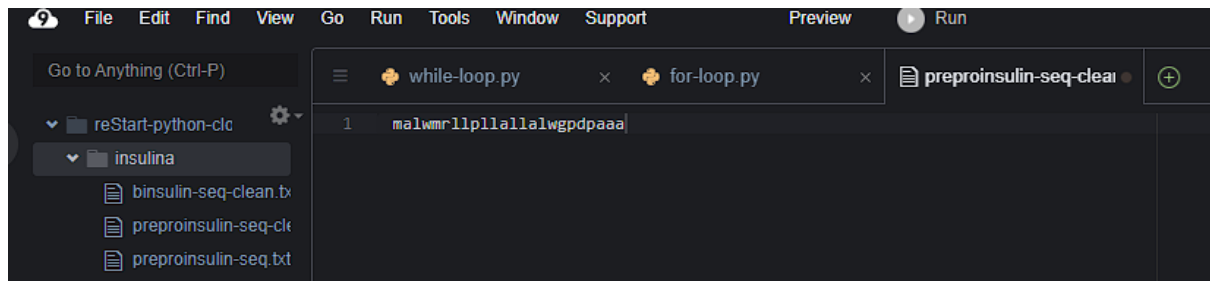
1. Copie los resultados en el archivo **preproinsulin-seq-clean.txt**.
2. Confirme que el archivo tenga 110 caracteres en minúscula, que representan los aminoácidos en la secuencia de la preproinsulina.



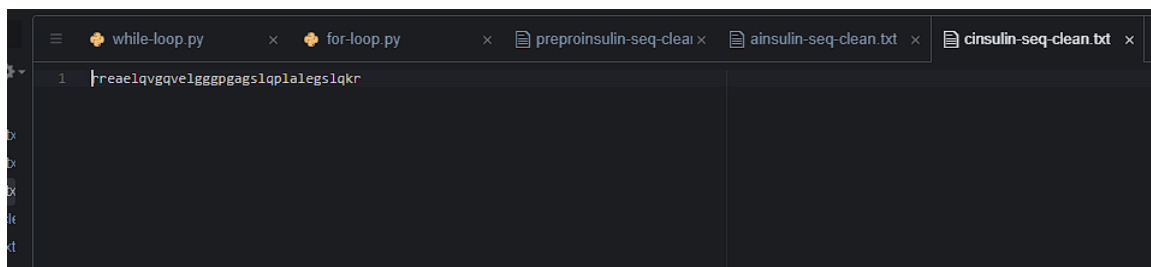
```
1 malwmrllpllalllalwgpdpaaafvnqhlcgshlvealylvcgergffypktrreaelqvgqvelggpgagslqplalegslqkrqiveqcctsicsslyqlenycn
```

1. 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*.

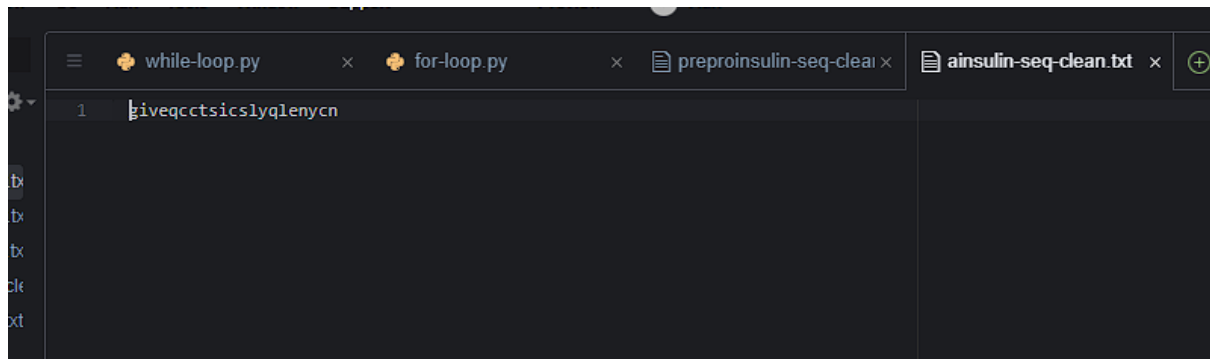
2. Guarde los aminoácidos 1–24 en **lsinsulin-seq-clean.txt**. Verifique que el archivo tenga 24 caracteres.
3. En el IDE de AWS Cloud9, en el panel de navegación, elija **File > New File** (Archivo > Archivo nuevo) y guarde el archivo como *binsulin-seq-clean.txt*.



1. Guarde los aminoácidos 25–54 en **binsulin-seq-clean.txt**. Verifique que el archivo tenga 30 caracteres.
2. 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*.
3. Guarde los aminoácidos 55–89 en **cinsulin-seq-clean.txt**. Verifique que el archivo tenga 35 caracteres.



4. 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*.
5. 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.

Finalizar laboratorio

¡Felicitaciones! Ha llegado al final del laboratorio.

1. Elige **End Lab** (Finalizar laboratorio) en la parte superior de esta página y, a continuación, selecciona Yes (Sí) para confirmar que deseas finalizar el laboratorio.

Un panel muestra el mensaje *DELETE has been initiated... You may close this message box now* (Se ha iniciado la ELIMINACIÓN... Ya puedes cerrar este mensaje).

1. Aparece brevemente el mensaje *Ended AWS Lab Successfully* (El laboratorio de AWS finalizó correctamente), que indica que el laboratorio ha finalizado.