# Preparación para el análisis de insulina con Python

## Información general sobre el laboratorio

En la tecnología de la información, Python destaca como el lenguaje de programación elegido 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 los laboratorios de los módulos de Python, realizará sencillas manipulaciones y cálculos de secuencias con insulina humana, que es una hormona conocida del cuerpo humano responsable de regular los niveles de azúcar.

En este laboratorio, realizará lo siguiente:

* Recuperar la secuencia de proteínas de la insulina humana a partir de la preproinsulina. Iniciará sesión con su ID de estudiante (y privilegios mediante sudo)

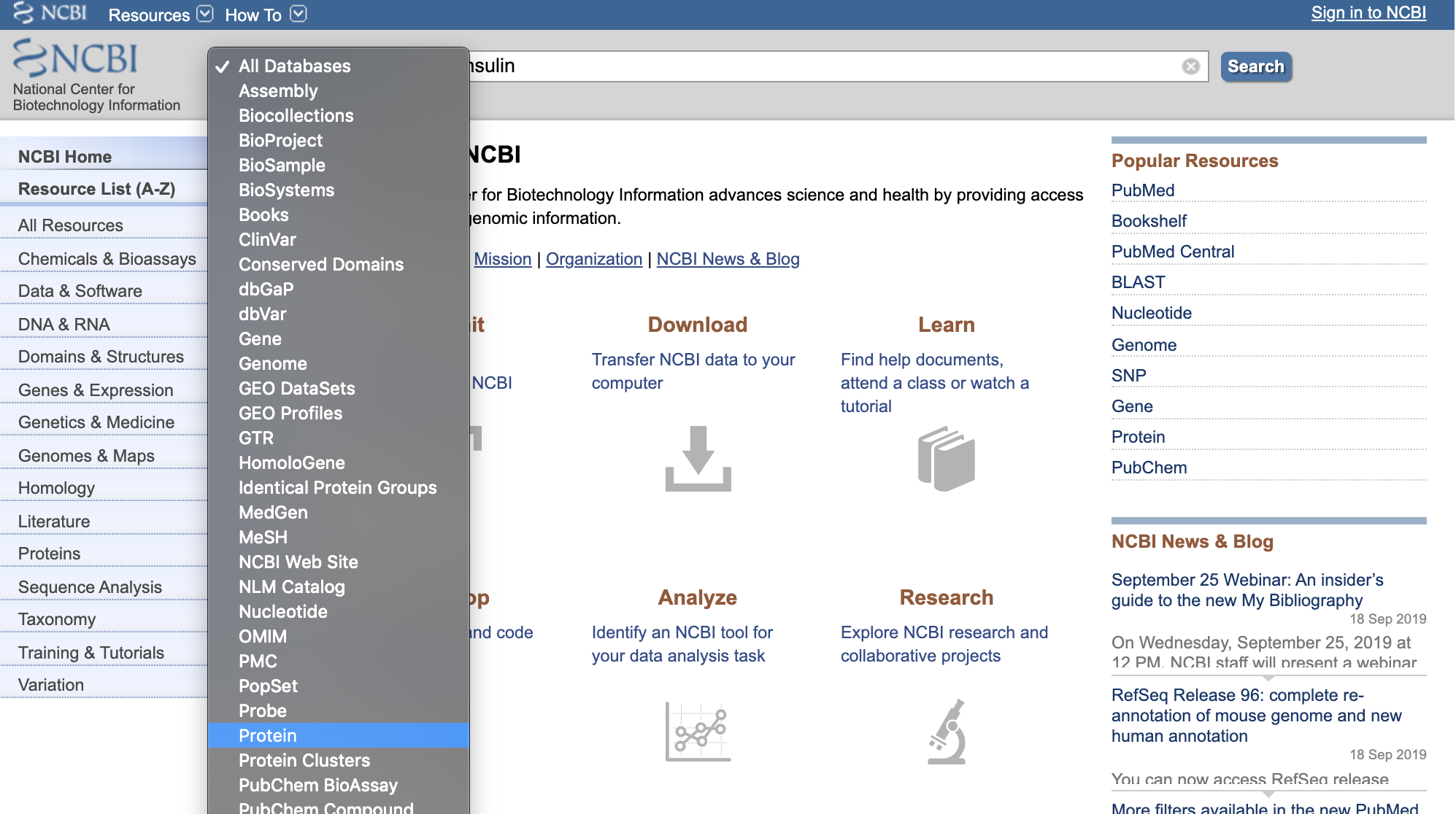
## Tiempo estimado para finalizar

30 minutos

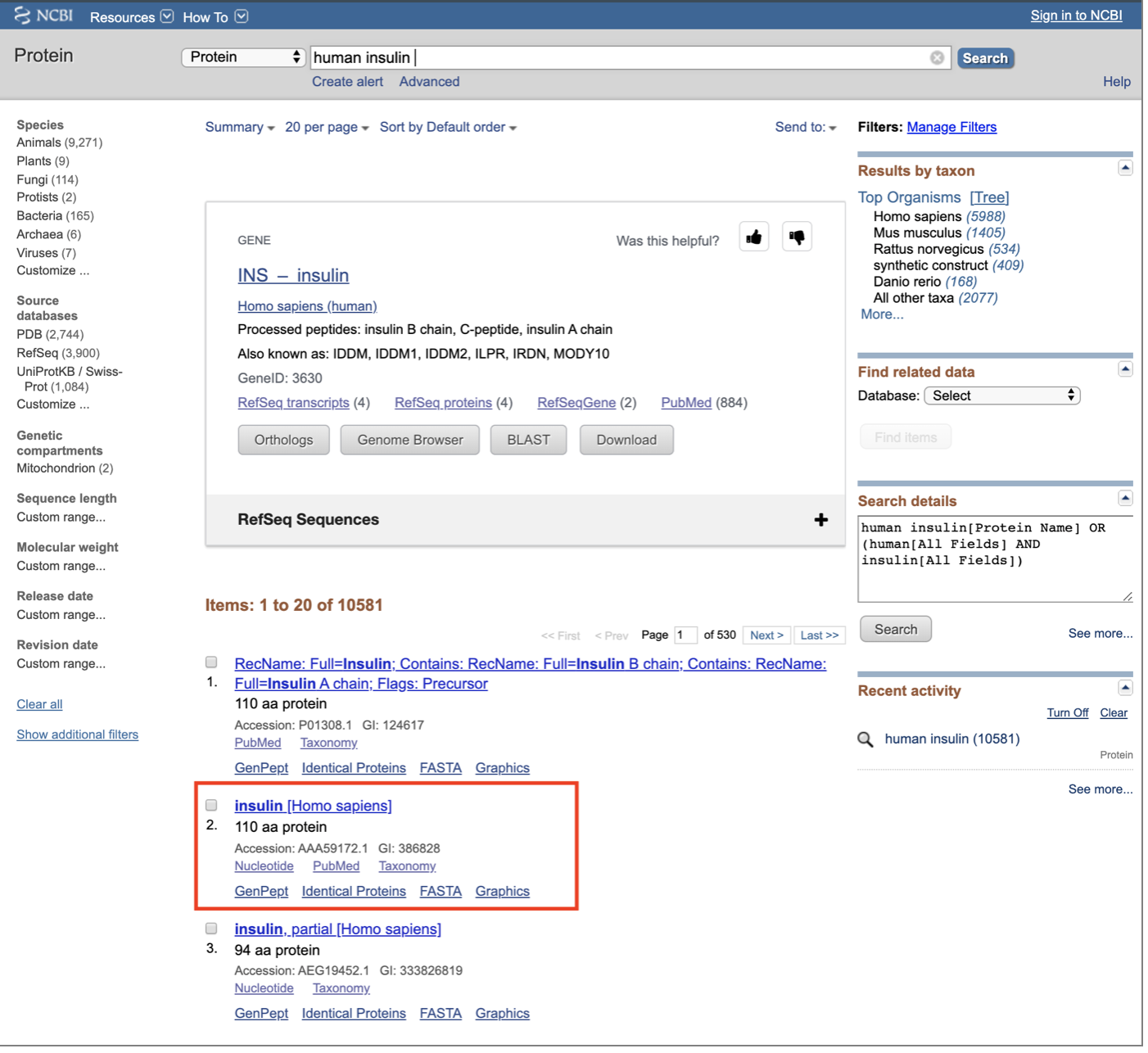
## Ejercicio 1: Recuperación de la secuencia de proteínas de la preproinsulina

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

1. Acceda al NCBI en [ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/).
2. Haga clic en el menú desplegable junto a la barra de búsqueda y seleccione Protein (Proteína). A continuación, escriba “human insulin” (insulina humana) en la barra de búsqueda y haga clic en Search(Buscar).



1. Haga clic en el segundo resultado de la búsqueda: insulin (insulina)[ Homo sapiens].



Haga clic en la carpeta work (trabajo) del árbol de archivos a la izquierda y, a continuación, haga clic en el archivo preproinsulin\_seq.txt.

Copie la secuencia de la insulina en preproinsulin\_seq.txt:  
ORIGIN

1 malwmrllpl lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed

61 lqvgqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn

1. //

### Adicional: limpieza mediante programación de insulin\_seq.txt

La limpieza de archivos de datos iniciales es una tarea común en programación. Existen varias formas de realizar la limpieza mediante programación de insulin\_seq.txt utilizando bash, Python u otro lenguaje de programación de su elección. Pruebe a utilizar expresiones regulares para eliminar del archivo el “ORIGIN”, sus números, “//”, espacios y saltos de línea o retornos de carro. También puede verificar que el archivo tenga 110 caracteres por código.

## 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 edición manual para recuperar solo los aminoácidos en la secuencia que compone la insulina.

1. Elimine manualmente o mediante programación “ORIGIN”, “1”, “61”, “//”, así como los espacios y los retornos de carro.
2. Copie los resultados en el archivo preproinsulin\_seq\_clean.txt.
3. Verifique que el archivo tenga 110 caracteres en minúscula, que representan los aminoácidos en la secuencia de la preproinsulina.
4. Guarde los aminoácidos 1-24 como lsinsulin\_seq\_clean.txt. Verifique que el archivo tenga 24 caracteres.
5. Guarde los aminoácidos 25-54 como binsulin\_seq\_clean.txt. Verifique que el archivo tenga 30 caracteres.
6. Guarde los aminoácidos 55-89 como cinsulin\_seq\_clean.txt. Verifique que el archivo tenga 35 caracteres.
7. Guarde los aminoácidos 90-110 como ainsulin\_seq\_clean.txt. Verifique que el archivo tenga 21 caracteres.

### Cuándo automatizar y cuándo trabajar manualmente: 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 de codificación adicional que se necesita para limpiar mediante programación insulin\_seq.txt y convertirlo en insulin\_seq\_clean.txt, pero si tuviera que descargar miles o millones de archivos y repetir la acción, sin duda valdría la pena la automatización.

Preparar los archivos manualmente le hará apreciar la automatización que Python puede proporcionar.

## Fin

Ha completado correctamente el laboratorio.