
Healx Data Content Technical Assessment

Please complete and return this technical test before attending your interview at Healx. This isn't something we would like you to spend a long time on, we recommend no more than around 3 hours. Please send us your code with instructions on how to execute it.

Your answer will form the basis of a technical discussion about how you implemented your solution, structured your code and the next steps you would take to make your solution ready for a production environment.

The task

We've provided you with a **healx-test-rare-disease-genes.csv** file listing 100 monogenic rare diseases with the gene symbol of the causative gene and the UniProt identifier of the associated Protein.

Please write **Python** code to read disease information from this file and fetch the UniProt record for each protein identifier using the UniProt REST API. Parse the data fetched from UniProt to identify which of these diseases have a **protein shorter than 600 amino acids** in length and are **annotated with the 'Apoptosis' UniProt keyword**. You may make use of any Python libraries you need to.

You can fetch the UniProt record for a protein using a RESTful web service call of the form:

<https://www.uniprot.org/uniprot/P37231.xml>

If you have any questions then please don't hesitate in reaching out to the team on: recruitment@healx.io who can forward your questions to the team and get back to you ASAP

Good luck!

Talent Team @ Healx 😊