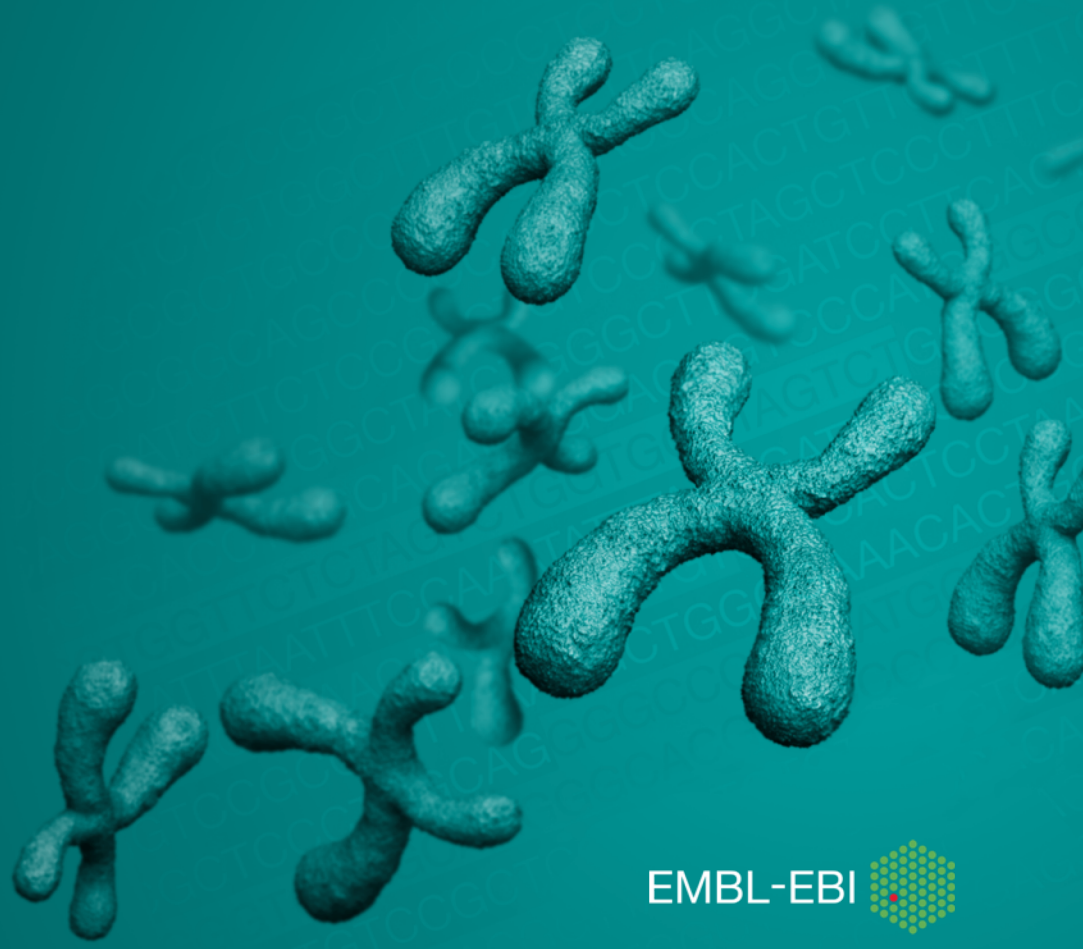
Proteins API Exercises



EMBL-EBI

Aim

1. Hands-on experience of navigating around the Swagger UI
2. An introduction to making simple requests
3. Familiarisation with the UniProt data model(s)

What you need

1. Jupyter notebook up and running
2. And a cup of coffee

All exercises are based upon the UniProt Entry – P21802 (FGFR2_HUMAN)

Part 1: Navigating and using the Proteins API Swagger UI

Introduction

Open a browser or browser tab and navigate to:

https://www.ebi.ac.uk/proteins/api/doc/

Using: Proteins – Proteins inc. isoforms

Question 1 – How many end points are available via proteins inc. isoforms?

Question 2 – Which endpoint is best to get a UniProt entry with the PDB identifier 1DJS? ‘Try It’

Question 3 – Can you use an endpoint to determine how many isoforms P21802 has? ‘Try It’

Question 4 – Is there an endpoint to end a specific isoform? ‘Try It’

Question 5 – Is there an endpoint that could allow you to retrieve every entry in UniProt (not recommended). Quickly think how you might do it?

Part 2: Adding filters and using basic script

Introduction

We are going to move to using Proteins – Features for this section. We will start by still using the Swagger UI but take the example Python script and use it in Jupyter to quickly modify the query to get specific results.

Note: In UniProt, Feature (FT) is used as a general term for an annotation on a specific part of the protein that contributes to that protein's function.

Using: Proteins – Features

Question 1 – Which is the best end point to get all the Features for P21802? 'Try It'

Question 2 – Who many different types of Features does P21802 have?

- Save the basic python script with format type JSON.

Question 3 – Is there an endpoint that allows you to **only** get Glycosylation sites?

- This question is tricky, you might have to ask a friend on what glycosylation does and you may have to think about what the potential filter could be.
- Modify you basic Python script to carry out the query and 'run it'.

Question 4 – How many Glycosylation sites are there on P21802?