# Interrogating the PDB

**Authors**: Keith Johnson, Bradley University, IL and Shuchismita Dutta, Rutgers University, NJ

Inspired by “PDB: The Protein Data Bank – A Critical Guide” (by Atwood T.K., 2018; <https://doi.org/10.7490/f1000research.1116056.1>)

**Lesson Overview**:

This lesson introduces students to the search and browse functions of the RCSB Protein Data Bank (RCSB PDB) and a few other bioinformatics resources. Familiarity with these options will help search for relevant information and structures in the PDB and other bioinformatics resources to answer research questions. Finding specific structures in the PDB and relevant information about them can be done in different manners. Here are some ways to find and explore PDB structures.

**ASBMB Learning Objectives**

(<https://www.asbmb.org/education/core-concept-teaching-strategies/foundational-concepts/skills>)

2. Accessing, comprehending and communicating science

* Students should be able to **identify, locate and use the primary literature** *(Introductory)*.
* Students should be able to **use databases and bioinformatics tools** *(Intermediate)*.
* Students, when provided with appropriate background information, should be able to **identify consistencies and inconsistencies** *(Advanced)*.
* Students should be able to use **visual and verbal tools to explain concepts and data** *(Intermediate)*.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

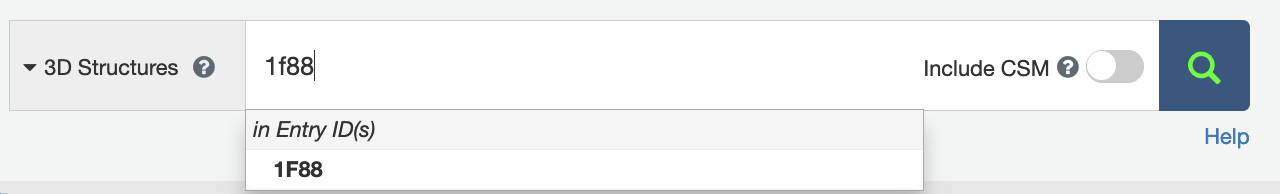
Each 3D-structure in the PDB is assigned a unique **Identifier (or PDB ID)**. The goal for finding PDB structures for a specific protein is identifying the PDB ID(s) for the relevant structure(s). You may find these PDB IDs listed in the scientific literature, in other bioinformatics resources (e.g., UniProt). In this exercise you will use the Search and Browse options available from the [RCSB PDB home page](https://www.rcsb.org/) to identify PDB IDs and gather information from these structures.

The molecule you will search for is rhodopsin. To begin, read the Molecule of the Month feature on rhodopsin (<https://pdb101.rcsb.org/motm/147>) to gain some insight into the structure and function of the rhodopsin protein.

In this worksheet, you will search in different manners for the protein rhodopsin in eight different ways.

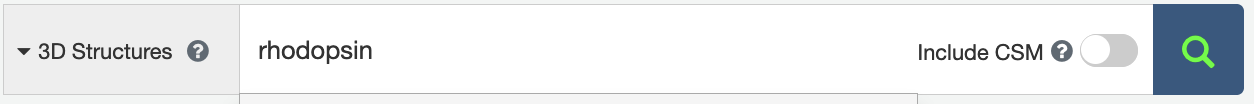
* Each way to search for the protein Rhodopsin is listed here with option numbers (A-F).
* Use these approaches to search for the protein and complete Table 1 shown below.
* Then answer the questions about the search results based on the data in the table.

1. ***PDB ID Search***: In the Molecule of the Month reading on rhodopsin, the structure that was shown is PDB entry 1f88. Using the generic ‘Search’ box at the top of the [RCSB PDB home page](https://www.rcsb.org/), type in the PDB ID 1f88 and launch the search by clicking on the ‘Go’ (magnifying glass icon).



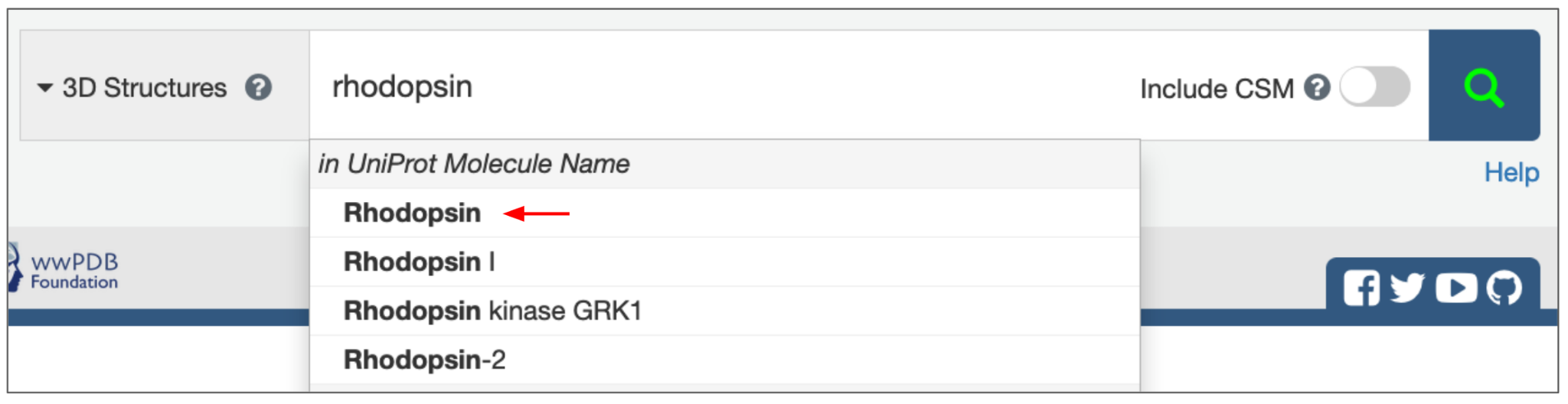
Use the results to complete information in Row A of Table 1.

1. ***Text search***: Using the generic ‘Search’ box at the top of the [RCSB PDB home page](https://www.rcsb.org/), type the keyword ‘rhodopsin’ and click on ‘Go’ (magnifying glass).



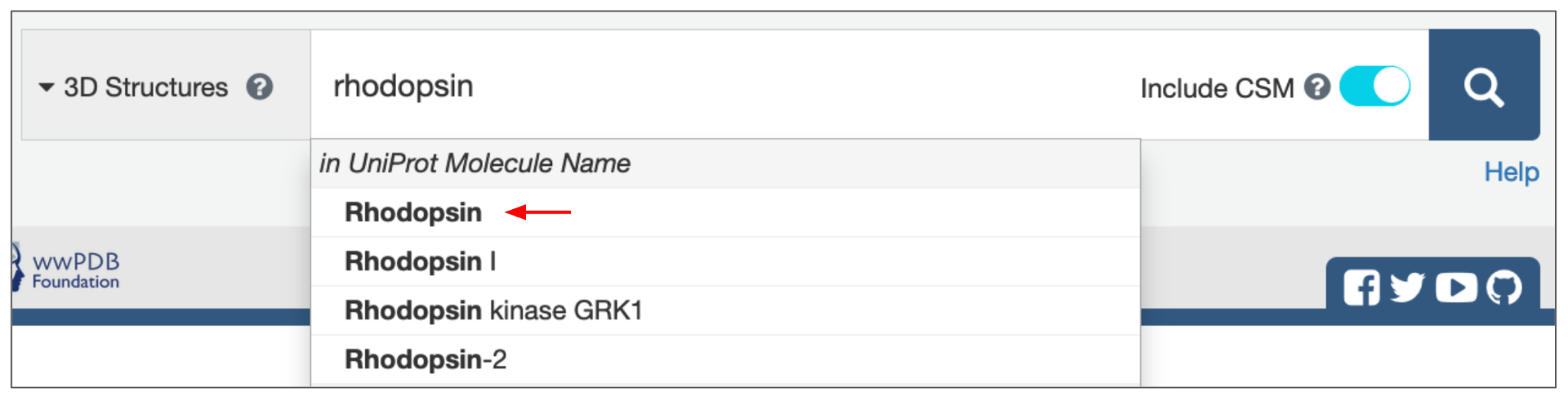
Use the results to complete information in Row B of Table 1.

1. ***Text Search (with autocomplete suggestions)***: Return to the ‘Search’ box and type ‘rhodopsin’ again. Before you click on the ‘Go’ button, pull down the ‘Rhodopsin’ under ‘in Uniprot Molecule Name’. *This provides some more specificity to the search.*



Use the results to complete information in Row C of Table 1.

1. ***Text Search (with autocomplete suggestions) + Include CSM*** *(Computed Structure Models)*: Return to the ‘Search’ box and first turn on the “Include CSM” slider switch and then type ‘rhodopsin’ again. Before you click on the ‘Go’ button, pull down the ‘Rhodopsin’ under ‘in Uniprot Molecule Name’. *This provides some more specificity to the search.*

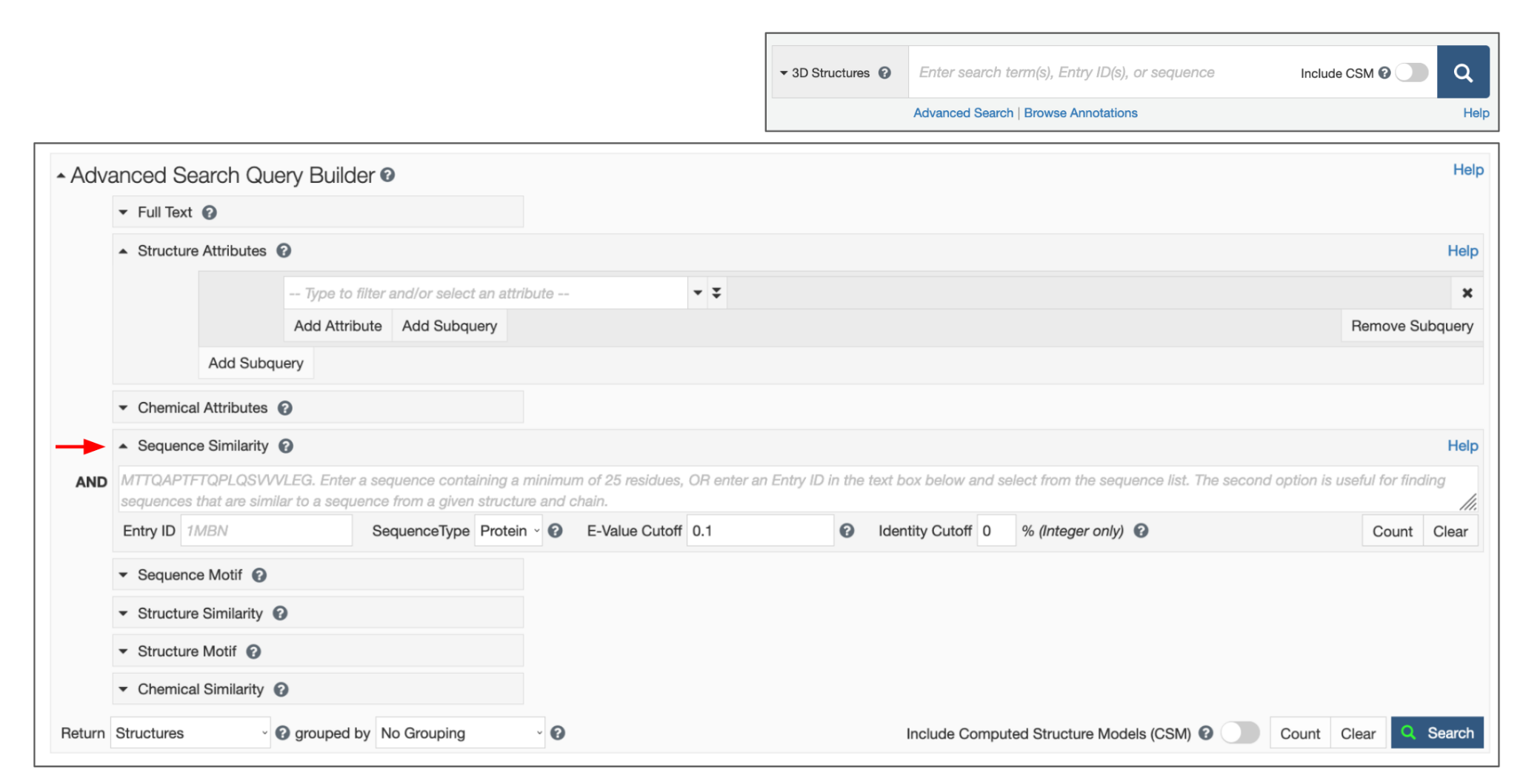


Use the results to complete information in Row D of Table 1.

1. ***Advanced Search (by Sequence)***: Search for Rhodopsin structures using the sheep rhodopsin protein sequence.

Return to the home page (<https://www.rcsb.org/>) and follow the directions below:

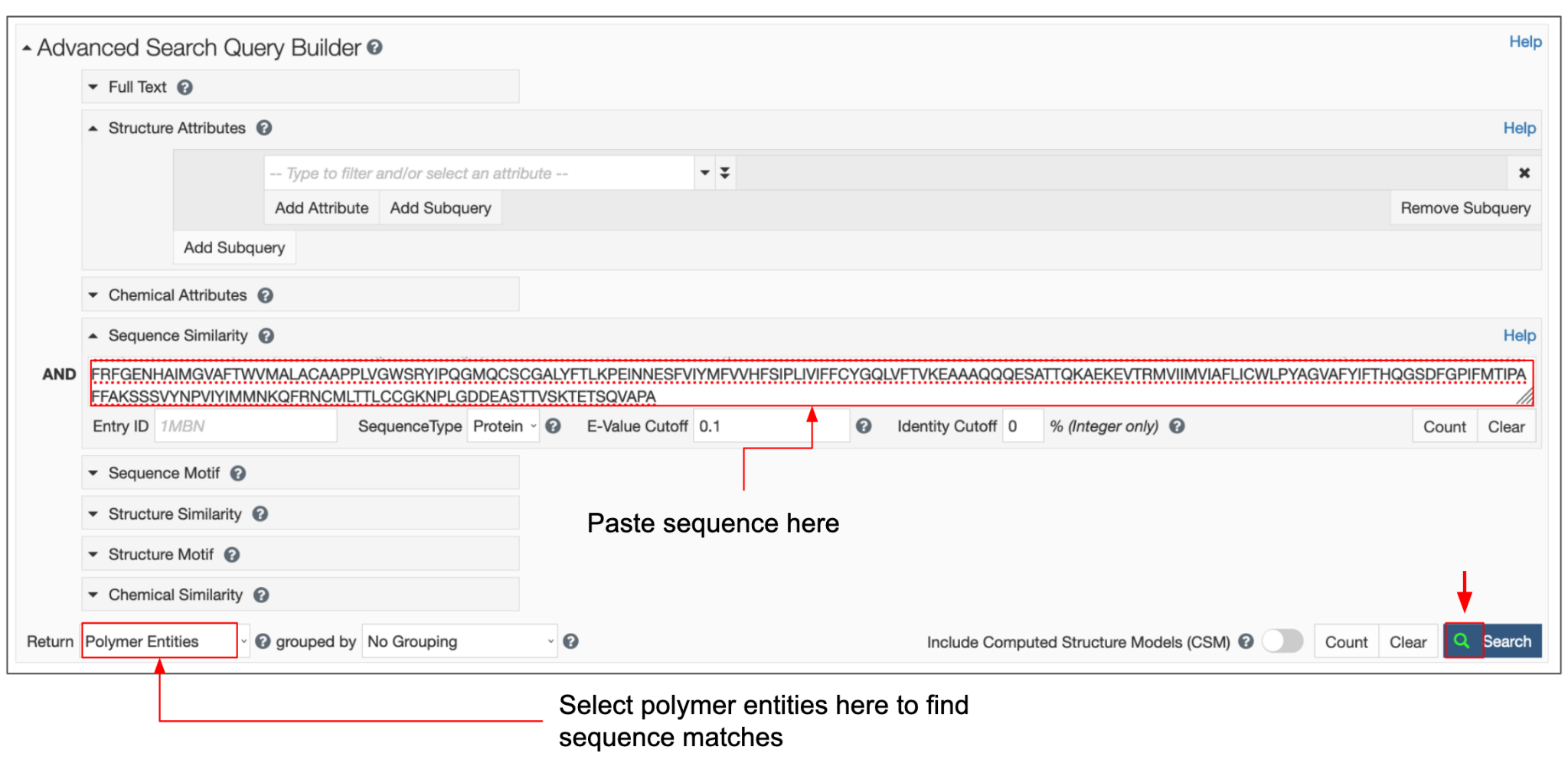
* 1. Click on the ‘Advanced Search’ link under the search box on the top of the page.
  2. In the Advanced Query Builder options presented, click to open the ‘Sequences Similarity’ option.



In a separate window, open the UniProt (the protein sequence database) website (<https://www.uniprot.org/>).

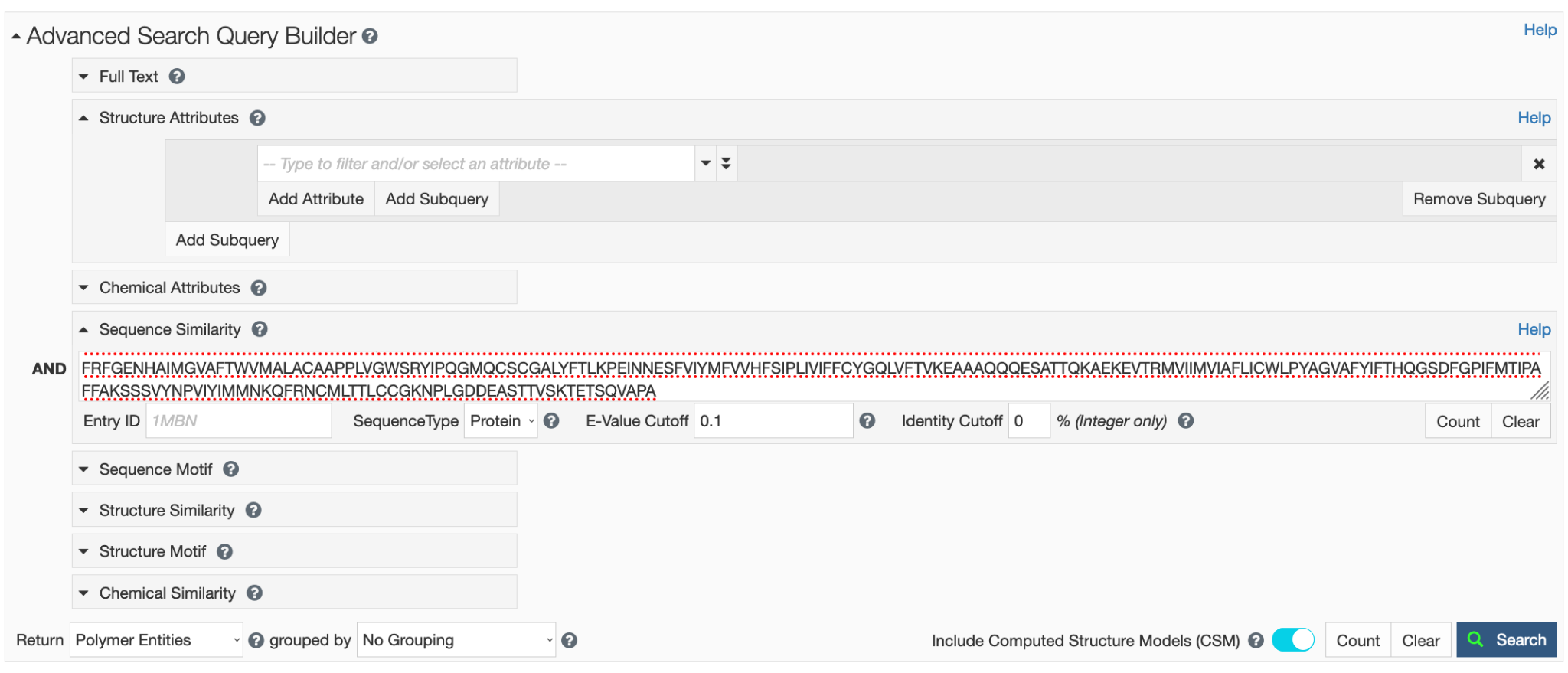
* 1. Search for Rhodopsin by typing the protein name in the search box on this website. 
  2. Scroll through the results returned and select the sheep protein from the list - entry ‘OPSD\_SHEEP’ ([P02700](https://www.uniprot.org/uniprotkb/P02700)). Open the page for this protein sequence, click on Download link to access the FASTA sequence (<https://rest.uniprot.org/uniprotkb/P02700.fasta>) (Hint: use the FASTA (canonical) sequence). Alternatively you can scroll down to the Sequence section of the page and click on the Copy Sequence link to copy the sequence to the clipboard. Go back to the Advanced Search Window.
  3. Paste in sequence the ‘Sequence Similarity’ window to run the sequence based search in the PDB.

Note: Since the query is a sequence, the search results will return sequences in the PDB archive that match. Select Polymer entities in the Return type and then click on the magnifying glass on right to perform the search (see image below).



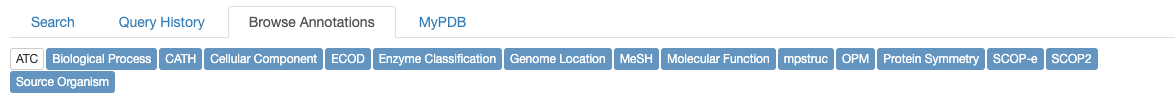
Use the results to complete information in Row E of Table 1.

1. ***Advanced Search (by Sequence) + Include CSMs***: Scroll up to the Advances Search Query Builder section of the page and turn on the “Include CSMs” slider next to the Search button. Now click on the search button and run the search again.

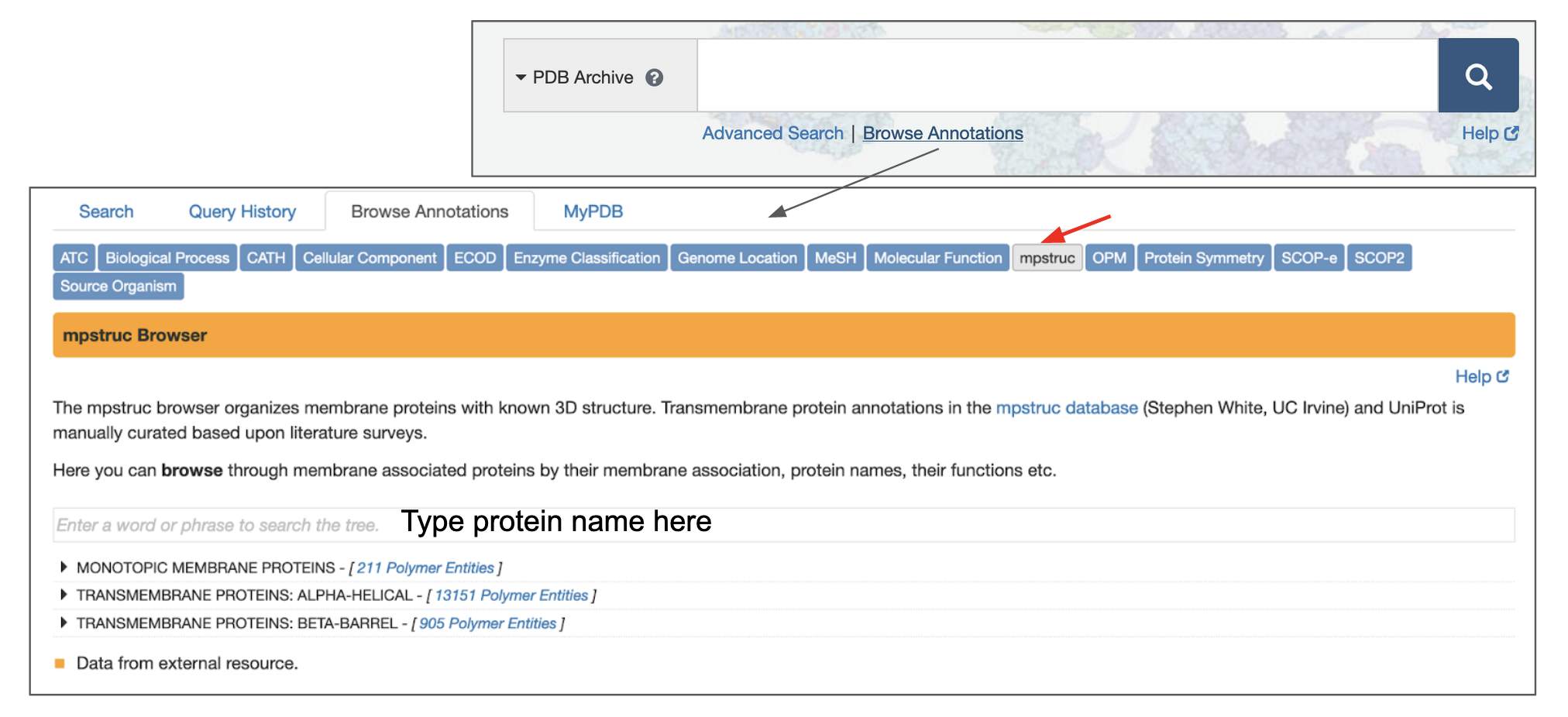


Use the results to complete information in Row F of Table 1.

1. **Browse Annotations (mpstruc)** (Membrane Protein): Rhodopsin is a membrane protein. A few membrane protein data resources have identified and organized [membrane protein structures](https://www.rcsb.org/docs/general-help/membrane-protein-resources) based on their structures locations with respect to membranes. You can explore these resources at [mpstruc](http://blanco.biomol.uci.edu/mpstruc/), [OPM](https://opm.phar.umich.edu/) etc. The membrane protein annotations have been mapped on PDB structures and may be browsed to find a protein of interest (e.g., Rhodopsin). On the rcsb.org website, browse/search the membrane protein browsers [mpstruc](https://www.rcsb.org/search/browse/membrane) and [OPM](https://www.rcsb.org/search/browse/opm) to find rhodopsin structures as follows.
   1. Click on the “Browse Annotations” link under the search box on the top of the page.



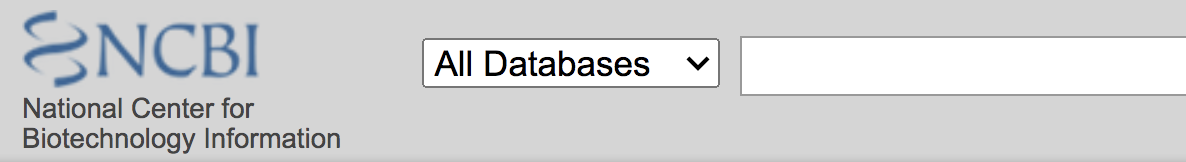
* 1. From the options presented, click on the tab for mpstruc and search for the protein of interest (by typing the name in the box presented on this page) and scroll through the options to select it (e.g., “Rhodopsin”) and see the results. Alternatively, browse through the classification to locate the protein of interest.



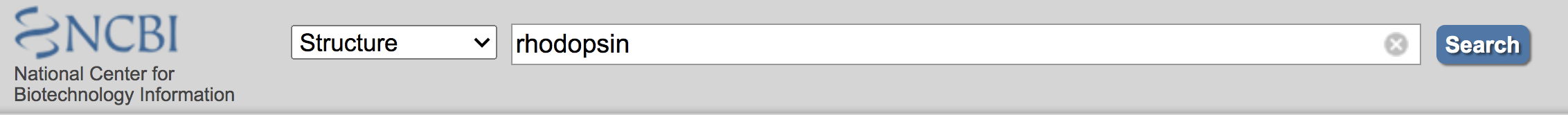
From the options presented, select the branch that is “Rhodopsin” without any prefixes, suffixes, or other words. Click on the structures linked to this branch of the browser tree.

Use the results to complete information in Row G of Table 1.

1. ***NCBI Structure Search***: Go to the NCBI main [web page](https://www.ncbi.nlm.nih.gov/).
   1. The page should look similar to



* 1. Select ‘Structure’ from the pulldown at ‘All Databases’ and enter ‘rhodopsin’ in the blank space.



* 1. Click on ‘Search’ on the right-hand side.

Use the results to complete information in Row F of Table 1.

Date on which the searches were run:

*Table 1: Search Results for PDB structures of Rhodopsin using different approaches.*

|  | Number of structures returned | PDB ID of top structure match in results list | Scientific name of top three most frequent source organisms in results |
| --- | --- | --- | --- |
| A | 1 | 1f88 | *Bos taurus* |
| B |  |  |  |
| C |  |  |  |
| D |  |  |  |
| E |  |  |  |
| F |  |  |  |
| G |  |  |  |
| H |  |  | **Do not fill this box** |

Answer the following questions about results of the search approaches tried above.

1. What are the main differences between the results returned in searches A-H?
2. Which search was the most selective?
3. Which search returned the most structures?
4. Which was the least focused search and why?
5. Searching the PDB archive with sheep rhodopsin sequence did not find any experimental structures. When CSMs were included, did you find any CSM for this sequence? List how many CSMs were found?

1. If you did not have a PDB ID to look up, which search option would you use? Why?