Version 1.2

**\*\* ONLY TEXT AND IMAGES APPEARING INSIDE THE RED BOXES WILL BE GRADED\*\***

1. This assignment explores IDRs [intrinsically disordered regions (of proteins)] and IDPs (intrinsically disordered proteins). You will use a web portal that finds IDR and IDPs from in input sequences. You will then use PyMOL to visualize your results.
2. Read the Introduction (Section 1) of this literature review: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4095912/>. Emphasis for our purposes is on the historical vs. current view of the relationship between sequence, structure, and function as discussed in Sections 1.2 and 1.4. What is the difference between an IDR and an IDP? Can IDRs and IDPs have biological function? Provide an example from Table 1 of the review to support your answer.

|  |
| --- |
| INSERT HERE |

Write your first and last name at the top of this file, and save it

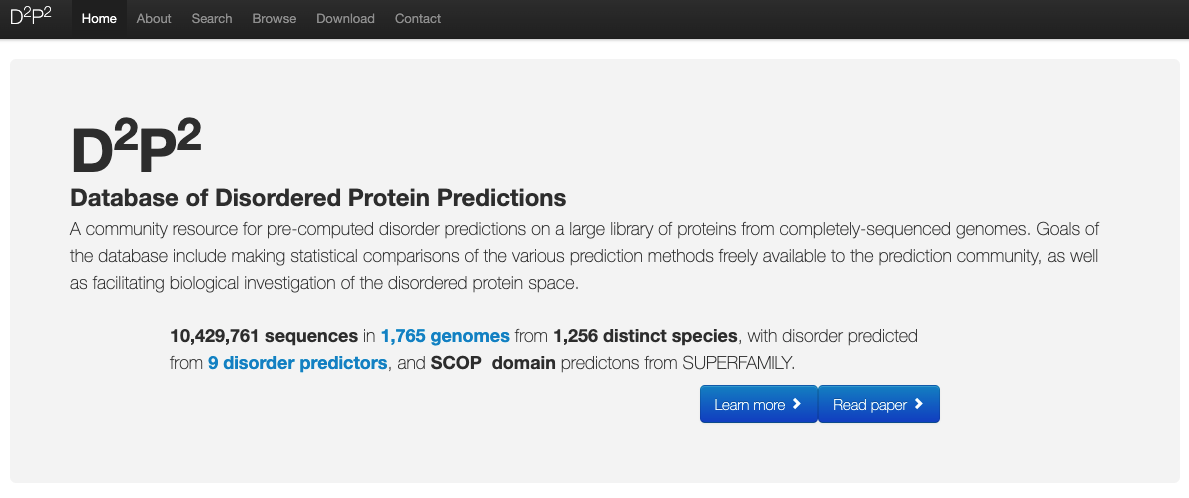
1. Follow the instructions below to use the Intrinsic Disorder Pymol Script (on the course webpage) and the Database of Disordered Protein Prediction, [D2P2](http://d2p2.pro/), to observe and predict disordered regions and structured motifs in a protein.

Open the Intrinsic Disorder PyMOL Script (the .pml) and turn off everything except the object called fus. Export the object sequence from the .pml file as a .fasta file with the following commands.

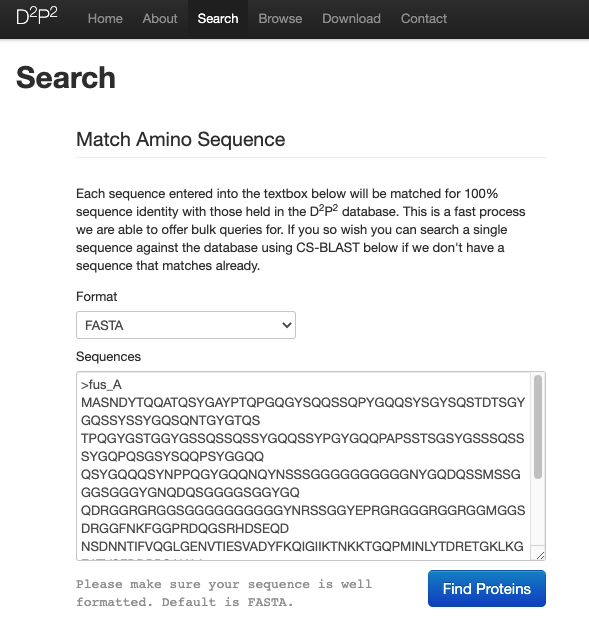
**PyMOL**>save fus.fasta, fus (to save a fasta file called fus from the fus object)

**PyMOL>**pwd (to show the directory in which the .fasta file was saved)

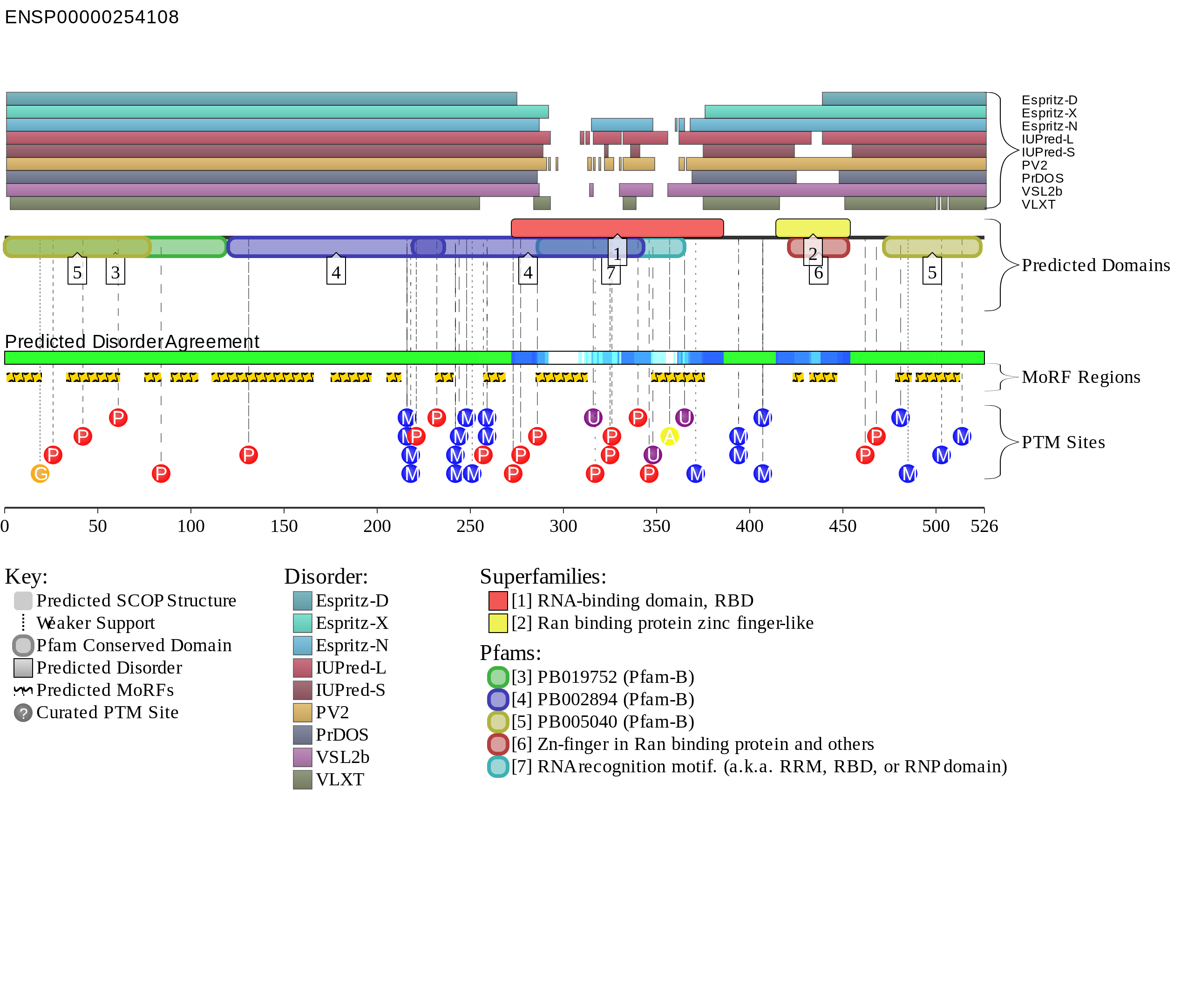
Go to the Database of Disordered Protein Prediction, [D2P2](http://d2p2.pro/), website and click ‘Search’.



Open the fus.fasta file in a text editor. Copy and paste the fus.fasta text (including the >fus\_A header) into the [D2P2](http://d2p2.pro/) search box. Click ‘Find Proteins’.



The search result should look like this (graph below). Notice that your amino acid numbers are the x-axis of the graph. The top part of the graph contains multiple predictions of intrinsically disordered regions. The middle part contains predictions of structured domains from two different databases ([SCOP](https://scop.mrc-lmb.cam.ac.uk/) and [Pfam](http://pfam.xfam.org/)). In the penultimate section, in lime green, are the predicted disorder agreement regions. Our focus is on the disorder agreement regions and the SCOP database superfamilies (yellow highlighted text).



Predicted Disorder Agreement (in lime green)

<Amino acid residue no.

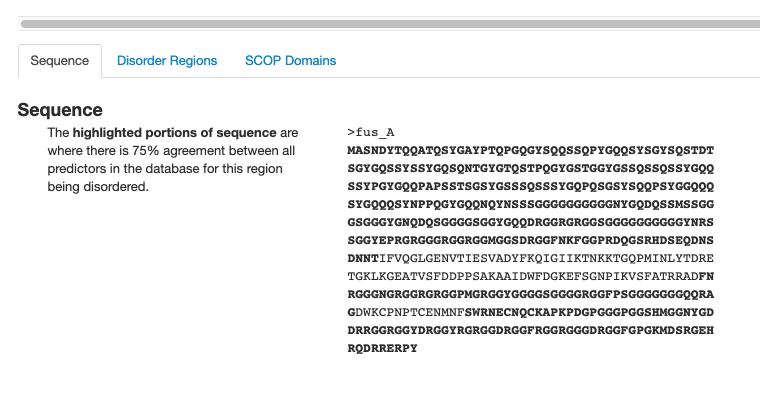
<SCOP superfamily

<Pfam family

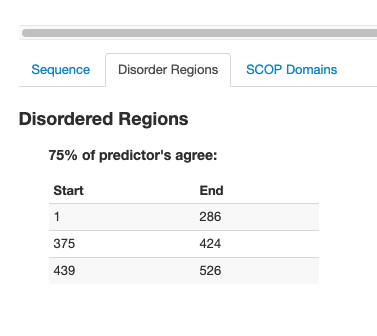
Regions of disorder as predicted by each of these different algorithms

^structured

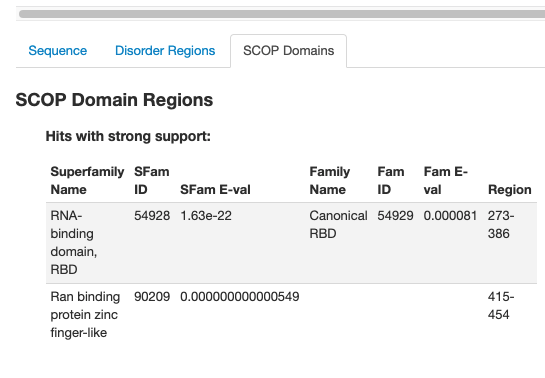
Scroll down the page, below the graph. The ‘Sequence’ tab gives the sequence of the predicted disorder agreement in bold fond.



The ‘Disordered Regions’ tab tabulates the starting and ending amino acid residue numbers of the Predicted Disorder Agreement.



The ‘SCOP Domains’ tab tabulates the starting and ending amino acid residue numbers of predicted structured domains.



1. Return to the Intrinsic Disorder PyMOL Script (the .pml). Create two new objects, one called ‘fus\_IDR’ and one called ‘fus\_structured’ with the amino acid regions D2P2 identified as IDRs or as structured domains. Recall that you need to use the ‘or’ operator to put multiple regions into a single object

**PyMOL**>create structured, resi 273-386 or resi 415-454

Notice that for fus, the IDR and structured regions predicted by D2P2 partially overlap.

1. Turn off all objects except fus\_IDR and fus\_structured, color the IDR object blue and the structured object red, and save a .png. Insert the .png here.

|  |
| --- |
| INSERT HERE |

1. Choose one other protein object in the Intrinsic Disorder PyMOL Script and analyze it for IDRs and structured regions following the same process. The protein chosen is:

|  |
| --- |
| INSERT HERE |

1. Which SCOP superfamily structures were identified?

|  |
| --- |
| INSERT HERE |

1. Analyze your assigned protein sequence with [D2P2](http://d2p2.pro/). Note: this will not work for all protein.

First, set up Pymol by copying and pasting the following (all at once) into the command line

**PyMOL>**

#-----first line in copy------------------------------------------------------------------------

## Header: General Commands ##

# delete all objects and reset pymol

reinitialize

# set the background color to white

bg\_color white

# make the background transparent for ray trace

set ray\_opaque\_background, 0

# set the ray trace mode

# normal color

set ray\_trace\_mode, 0

# normal color + black outline

#set ray\_trace\_mode, 1

# black outline only

#set ray\_trace\_mode, 2

# turn off shadows during ray trace

set ray\_shadows, 0

# set the mouse mode for laptop.

config\_mouse one\_button

# get rid of double bonds and skinny bonds to H

set stick\_h\_scale, 1

set valence, 0

# high quality surfaces

set surface\_quality, 3

## End of Header: General Commands ##

#-------last line in copy---------------------------------

Fetch your protein.

**PyMOL>**fetch YYYY #YYYY is your PDB Code#

If you have multiple polypeptide chains in your protein, create a new object containing only one of the chains to analyze with [D2P2](http://d2p2.pro/).

**PyMOL>**create chain A, chain A  #creates an object called chain A#

Then, export the chain A amino acid sequence to a file

**PyMOL**>save chainA.fasta, chain A #save a fasta file with the sequence of chain A#

**PyMOL>**pwd #show the directory where saved#

**PyMOL>**ls #lists the files in that directory#

Analyze your protein with [D2P2](http://d2p2.pro/). Turn off all PyMOL objects except [your protein]\_IDR and [your protein] \_structured. Color the IDR object blue and the structured object red. Write out a png file

**PyMOL>**save image\_name.png

Insert the image\_name.png here. If your D2P2 search yielded ‘Sorry, no results found’, insert a screenshot of that result instead.

|  |
| --- |
| INSERT HERE |

1. Save the document as Assignment\_17\_lastname.docx.