

BIOI 3500, Advanced Bioinformatics Programming Assignment 11



Assignment 11 :: 50 points :: **Due: 11:59 p.m. Friday, 30th April 2021**

Objectives

This assignment will give you the opportunity to practice parsing BLAST output with Biopython. More specifically, your program will extract information from the Protein Data Bank.

Description

The file `multiQueryOutput.xml` (available on Canvas under **Files/Assignments/Assignment11**) contains the results of a multiquery BLAST search on five Uniprot sequences against the Protein Data Bank. Using Biopython, write a procedural Python (version 3) program (`netID_Assign11.py` where `netID` is your UNO NetID) that parses the provided BLAST output.

For each BLAST record in the output file, your program should display:

1. The Uniprot ID of the query. This can be extracted from the “query” attribute of each record (you might have to split the string using the ‘|’ character as delimiter).
2. The PDB ID of the hit. This can be extracted from “hit_def” attribute of the alignment objects (the PDB ID is the first four letters of the string).
3. The score, percentage identity, and coverage of the best HSP. If there are multiple HSPs for a given alignment, your program should only report these values for the best HSP, i.e., the HSP with the **highest score** (use the “score” attribute for each hsp).

The **percentage identity** of an HSP is defined as the number of identities over the length of the HSP alignment. The coverage of an HSP is defined as the length of the aligned part of the query sequence over the length of the query sequence. (*Hint: the hsp object contains a query_start and query_end attribute you can use to compute the coverage*).

The format should look like this (fields are tab separated):

```
P01315      6B3Q   416.0 0.855 1.000
P01315      2KQP   347.0 0.826 0.778
P01315      6ZI8   336.0 0.833 1.000
P01315      6PXV   331.0 0.798 0.778
...
```

Notice that the numbers in the sample output are the correct ones for the five Uniprot-PDB pairs reported above.

Tip: develop your program by trying things out in the Python REPL first.