PDB Analysis Interface Mini-project

The goal of this project is to write a program that provides users with different options to allow them to analyse PDB files. The program should always display the options available and only exit if the user enters in the option 'q' (for 'quit').

Note: Your program only needs to work with single chain PDB files, but a bonus mark will be given if you can get it to work with multiple chain PDBs.

The program should provide the following options.

R – Read in PDB files

If this option is selected, your program must ask the user to enter in the name of a PDB file on the system. The program will then read in this file to be used later on.

If the file has already been read in, an appropriate message should be displayed stating this.

Similarly, if the file does not exist, a message should be displayed explaining this (i.e. your program should **not** crash if this happens).

Your program should keep track of all PDB files uploaded this way (e.g. using a dictionary)

S – Search option

The program should give a list of PDB files already read in. The user should select one of these to work on. Then your program should report the number of 'glycosylation' sites in the protein, assuming glycosylation sites have the following motif consisting of **four** amino acids:

- The first amino acid is either 'G', 'Y' or 'L'.
- This is followed by an 'A'

- This is followed by a ‘P’, ‘F’ or ‘W’
- The final residue will then be a ‘T’, ‘L’, ‘V’, ‘G’, ‘M’ or ‘I’

Your program should print out the PDB sequence with all glycosylation motifs shown in lower case and the rest of the sequence in upper case.

W – Write out option

If this option is selected, the program should give a list of PDB files already read in. The user should select one of these to work on. Then the following sub-option(s) should be shown.

1. **Write out coordinate file**

If this option is selected, your program should write out only the coordinates section of the PDB file to another file. Your program should ask the user to provide a name for this file. Your program should provide the following sub-options for the user to select.

   a. **All atom**

This should provide all coordinate atoms.

   b. **Backbone atoms**

If this option is selected, your program should write out coordinates for all backbone atoms (CA, C, N, O). Example of a PDB that only has backbone atoms: 1MGX.pdb

   c. **Alpha carbon atoms only**

If this option is selected, your program should write out coordinates for all alpha-carbon atoms (CA). Example of a PDB that only has alpha-carbon atoms: 1QCR.pdb

2. **Write out sequence (Fasta format)**

If this option is selected, the program should give a list of PDB files already read in. The user should select one of these to work on. Then the following sub-options should be provided to write out the sequence of the protein in Fasta format.

1. **SEQRES Sequence**

This should write out the sequence of the PDB file according to the SEQRES section of the PDB file.

2. **Coordinate Sequence**

This should write out the sequence of the PDB file, based on the coordinates section.

3. **Alignment Sequence**

This should write out a sequence equivalent to the SEQRES section with all the missing residues converted to 'X' characters.

I – Information option

If this option is selected, the program should give a list of PDB files already read in. The user should select one of these to work on. Your program must print out the requested information (i.e. **not** write it to file), using the following options.

1. **Display coordinate sequence (as explained above)**
2. **Display SEQRES sequence (as explained above)**
3. **Display Alignment sequence (as explained above)**
4. **Display all non-water ligands in the protein (if any)**

A – Alignment option

If this option is selected, users must be asked to select two PDB files that they want to align. The program must then extract the sequences from these PDB files and align them. The alignment should then be written to file – the user should be asked to provide a name for this alignment file.

H – Help option

Lists all options and gives a brief description of each.

Q – quit

Terminates program