Introduction to Bioinfromatics

Name: Priyank Patel

UIN: 00985578

Email: ppatel@cs.odu.edu

Extraction of SLS Patterns from Protein Structures

Project Requirements:

In this project we have to create a program that can identify SLS structures from user defined PDB with the user desired length of the loop.

My approach:

I have created a program in "perl" language to parse the PDB file to get the required data to serve the task and process the data.

I have processed the files using the approach below,

1. Ask user for the input file name and size of the desired loop in a SLS.
2. Open the PDB file
3. Parse all the location information about Sheets and helixes from the file
4. Sort this information to make the searching easy.
5. Compute interval length between all the pairs of Sheets
6. Compare this computed length with the desired length of the loop.
7. If the computed length and the desired lengths are equal then check if the helix is present in that interval or not.
8. If helix is not present then extract the Atom coordinates information of that SLS and write it to the file, else go to the next structure.

The program if it finds any suitable SLS in the file will write a file with the following file name convention.

NameOfProtein_startPositionOfLoop_SizeOfLoop.pdb

Running the program:

To run the program, follow the steps below,

1. Put all PDB structure files in a folder with a1.pl file.
2. Run the perl file with the command "perl –w a1.pl".
3. The program will ask you to give the PDb file name.
4. The program will ask you for the length of the loop you want.
5. It will print all the sheet and loops present in the file in sorted order and creates a file for each SLS found.
6. If there are no SLS present it will give you a message saying that.

Tests:

I have tested my program with couple of files and below are the screen shoots of the run,

File: 1AOP.pdb

Size: 3

```
1AOP.pdb
Enter the length:
3
=========Sheet=========
81 85
89 91
111 115
119 126
155 158
215 221
234 242
245 254
272 279
347 351
354 361
365 367
389 393
397 404
431 434
473 476
490 497
500 506
516 523
=========Helix=========
93 106
132 140
166 183
229 233
281 297
306 315
316 328
375 386
407 418
426 430
450 465
485 489
524 542
547 556
written to the file 1AOP_242_3.pdb
written to the file 1AOP_351_3.pdb
written to the file 1AOP_497_3.pdb
```

File: 1NVS.pdb

Size: 2

```
atria:~/bnf/structures> perl -w al.pl
Enter the file name:
1NVS.pdb
Enter the length:
2
=========Sheet=========
8 18
21 28
34 41
70 76
79 85
90 91
126 127
136 138
144 146
153 154
156 157
160 161
302 304
=========Helix=========
51 60
93 96
103 124
132 134
170 174
175 175
176 181
185 204
215 223
230 234
235 246
255 260
written to the file 1NVS_154_2.pdb
```

File: 1NVS.pdb

Size: 6

```
Enter the file name:
1NVS.pdb
Enter the length:
6
=========Sheet=========
8 18
21 28
34 41
70 76
79 85
90 91
126 127
136 138
144 146
153 154
156 157
160 161
302 304
=========Helix=========
51 60
93 96
103 124
132 134
170 174
175 175
176 181
185 204
215 223
230 234
235 246
255 260
written to the file 1NVS_28_6.pdb
written to the file 1NVS_138_6.pdb
atria:~/bnf/structures> 
```