Biomolecular Structures Assignment 1 Question 2 STRIDE

Aum Khatlawala, 2020113008

February 17, 2022

1 Introduction

STRIDE (Structural Identification) is an algorithm that helps the user assign the protein secondary structure elements given the atomic coordinates of the protein which can be collected using X-ray crystallography and various other methods and stored in the Protein Data Bank in pdb files. The algorithm uses criteria such as hydrogen bonds, dihedral angle potentials and certain other statistical probability factors derived from the empirical examinations of solved structures from the Protein Data Bank.

2 VMD and STRIDE

VMD uses the program STRIDE to determine the secondary structures of molecules. There are multiple tcl scripts made using STRIDE that can be found online and these can easily get the job done. However, the answers derived from the STRIDE algorithm might differ from what's expected. This occurs due to one of several reasons including atom names not recognised by the STRIDE algorithm, coordinates in the VMD file being shifted very far from the origin and some weird structures in the file such as proteins being bonded to small pieces of unusual structures like nucleic acid. Other kinds of issues include OS problems and other system level issues. It is easy to identify what kind of problem there is by just running a structure that STRIDE gives a correct output for. If it runs without errors, then its not a system related problem.

3 A Basic Understanding of the Algorithm Used

STRIDE implements a knowledge-based algorithm that combines the concepts of hydrogen bond energy and statistically derived backbone torsional angle (dihedral angle) information from already available information to return results about the secondary structures of a protein in maximal agreement with crystallographer's designations.

4 Understanding the STRIDE Algorithm Approach at a Basic Level

If we have been given a file (for example a pdb file) with the atomic co–ordinates of the atoms in the protein, we can follow the given steps to get the secondary structures of the protein:

1. Extract the co–ordinates of all the N, C and CA atoms from the pdb file without disturbing the order of the atom co–ordinates
2. Find the ϕ angles for each C–N–CA–C residue and the ψ angles for each N–CA–C–N residue.
3. If we have more than 3 consecutive ϕ angles as -57° with some margin of error (the value for ϕ for right handed alpha helices) and those consecutive angles also satisfy the condition of having ψ angles as -47° with some margin of error (the value for ψ for right handed alpha helices), then we have a right handed alpha helix.
4. The same approach is followed for determining other secondary structures.

5 STRIDE algorithm on 1a6g.pdb

Upon running the algorithm on the 1a6g.pdb file, we get the following output:

LOC AlphaHelix GLU 4 A GLU 18 A 1A6G
LOC AlphaHelix VAL 21 A SER 35 A 1A6G
LOC AlphaHelix PRO 37 A LYS 42 A 1A6G
LOC AlphaHelix GLU 52 A ALA 57 A 1A6G
LOC AlphaHelix GLU 59 A LYS 78 A 1A6G
LOC AlphaHelix HIS 82 A LYS 96 A 1A6G
LOC AlphaHelix ILE 101 A HIS 119 A 1A6G
LOC AlphaHelix ALA 125 A LEU 149 A 1A6G
LOC 310Helix PRO 120 A ASN 122 A 1A6G
LOC TurnIV GLU 18 A VAL 21 A 1A6G
LOC TurnI PHE 43 A PHE 46 A 1A6G
LOC TurnI ASP 44 A LYS 47 A 1A6G
LOC TurnI PHE 46 A LEU 49 A 1A6G
LOC TurnI' LYS 78 A HIS 81 A 1A6G

We can clearly see that it identifies 8 alpha helices as the only secondary structures and this is confirmed by the analysis done on the protein.