GSoC 2019 Project Proposal

Manan Goel
manangoel1999@gmail.com

BioShell

A New Activity for Sugar

My Introduction

- I am Manan Goel, a sophomore at the International Institute of Information Technology - Hyderabad pursuing a B.Tech. in Computer Science followed by an M.S. in Computational Natural Sciences.
- My first language is Hindi but I am also fluent in English.
- My email is manangoel1999@gmail.com and this is the link(manangoel99) to my GitHub profile.
- I am placed in Hyderabad, Telangana, India (+5:30 GMT).

Previous Projects and Open Source Contributions

- I have to admit that I haven't contributed a lot to large scale open source projects like sugar. I have made a few pull requests which are listed below -
  - https://github.com/sugarlabs/www-sugarlabs/pull/292
  - https://github.com/sugarlabs/www-sugarlabs/pull/291
- I have worked quite a bit on a lot of solo projects which I made using different frameworks in different programming languages. The following is a list of my personal projects that I've worked on over the past two years -
  - The following are my versions of existing games that I made using OpenGL and WebGL
    - https://github.com/manangoel99/SubwaySurfer
    - https://github.com/manangoel99/FlightSimulator
    - https://github.com/manangoel99/JetpackJoyride

- This is a web-based billing system I worked on for a catering start-up company in my hometown
  - https://github.com/manangoel99/BestBhojv2
- A quizzing web app with its front end created using ReactJS and back end written in GoLang.
  - https://github.com/manangoel99/QuizPortal
- A web app clone of the Android app SplitWise made in Flask and BootStrap was used for the Front End.
  - https://github.com/manangoel99/LetsSplit
- As a part of an internship, I did last year I worked on creating a proxy rotator for a web scraper created using BeautifulSoup and Selenium and also optimize the amount of data collected along with the time taken to do it.

Motivation

- **My motivation for taking part in the Google Summer of Code:**
  - My major motivation for being a part of Google Summer of Code starts with a little story. During my last two years of school, I was one of the very few students who pursued computer science and at the time there was a persistent thought in my mind that writing code for these basic algorithms has to mean something. Writing code can't just stop at sorting an array of numbers. When I got into college I found open source to be what I was looking for, something which brought people together to work as a community to build something that affects so many.

    As a first year, I was never confident enough about my skills and was hence, too shy leading to the fact that I couldn't effectively communicate with the people of different open source communities.

    I believe that GSoC is an amazing platform to get involved with open source communities since it provides a single place through which a person can get in touch with the world of open source.

- ## Why I chose Sugar Labs:

    - I came to know about Sugar Labs through GSoC organisations while I was searching for different projects I could start working on. The first thing that made me want to work with Sugar Labs was that almost all the projects were built using Python, a language I believe I'm extremely proficient in.

      I grew up in a small town in interior Haryana where the idea of getting educated from a computer-based platform is extremely new, so I felt that Sugar Labs brings something to children which I did not have for a long time and I would love to contribute to that.

      I found the Sugar Labs community to be amazing with prompt replies from mentors who are extremely helpful. Moreover, I also loved the fact that it is a meritocracy where I could start working on different issues and projects as soon as I could.

- ## Why I want to work on this project:

    - I am pursuing a B. Tech. in Computer Science which will be followed by an M.S. in Computational Natural Science where we study how Computer Science techniques like Machine Learning, probabilistic simulations can be applied in Natural Science Problems like Protein Folding which has baffled scientists for centuries now.

      My passion for this project comes from my passion for Bioinformatics which is, in my opinion, a beautiful amalgamation of the two fields. There are some amazing bioinformatics algorithms which help us gain insight into different biological problems like how we differ from other organisms on a DNA level and from the same data we can also find out how genetically close we are to different organisms.

When I first saw these algorithms in action, it blew my mind and I would like to give a similar experience to children but at a much earlier stage than I did.

- **My Expectations from Sugar Labs:**
  - During :
    - I understand that molecular biology may not be the field of expertise for my mentors but I would like that they read about a few very basic concepts like Phylogenetic Trees and the Ramachandran Plot which I can also help them with.
    - I would like the mentors to be supportive just like they have been till now but also not shy away from giving me constructive criticism since that would help me in making a great project.
  - After :
    - There are a lot of other molecular biology-based concepts which I won't be able to include in the activity during the course of GSoC 2019. I would like to keep working on this project even after GSoC ends and I would love to have the support of the organisation in doing the same.

Project Details

1. Overview

"BioShell" is a new activity for the Sugar platform through which I hope to help students get a better understanding of different molecular biology problems. As a student in school, I very quickly lost interest in molecular biology because when I studied the concepts of DNA sequences, Protein Structures, Transcription and Translation etcetera, I could not understand how they worked and of the significance, they had in the real world. Last year, I took a course in college which introduced me to these concepts from the perspective of a computer science student.

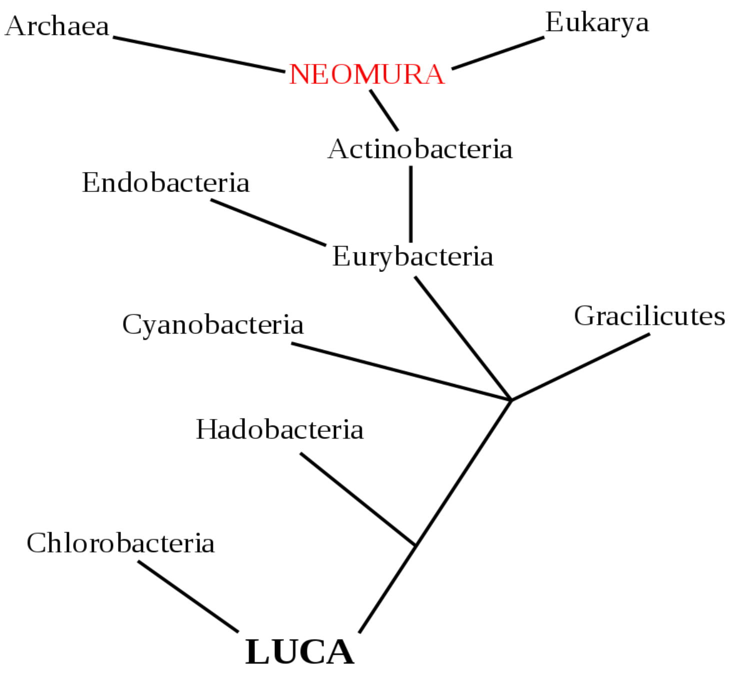
Of these concepts, the activity will include the construction on Phylogenetic Trees using DNA sequences of different organisms which are also available online, Analyzing protein structures described in .pdb files which are also available online and Translation process which involves creating a Protein in accordance with RNA which can be obtained from DNA.

The activity will open with a home screen where the user can choose between using the Phylogenetic Tree Construction and Protein Analysis modules.

This activity is aimed at students in and above the 10th grade who get an introduction to molecular biology and the concepts that I'm presenting.

2. Construction of Phylogenetic Trees

A phylogenetic tree is a tree showing evolutionary relationships between different species upon similarities and differences in their genetic characteristics.

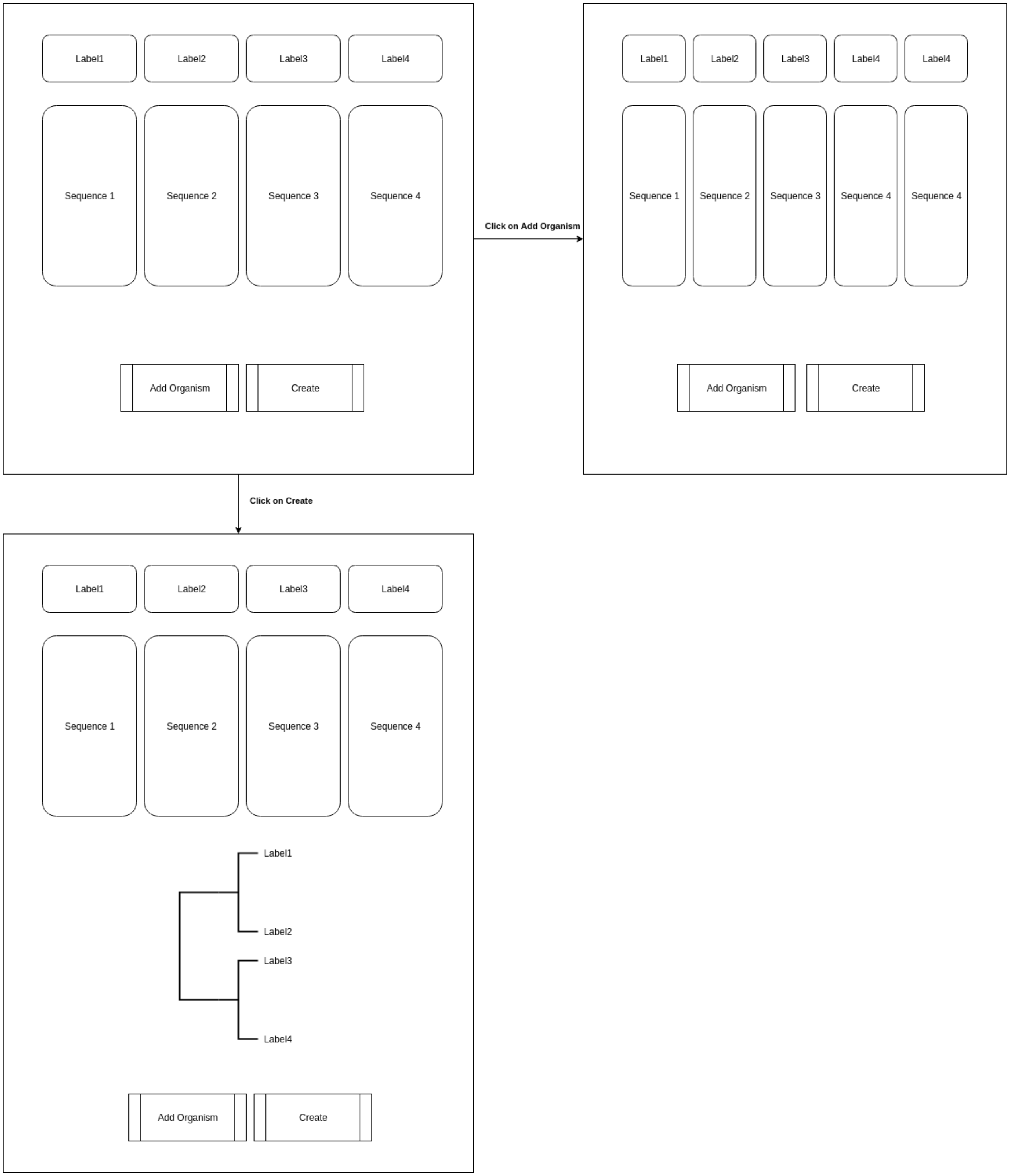


The DNA sequence of each known organism is comprised by repetition of four compounds namely Adenine, Guanine, Tyrosine and Cytosine which are simply known as A, G, T and C respectively. The DNA sequence for an organism can be represented as a string of A, T, G and Cs. For Example - The following string is the DNA sequence found in insulin in humans.

“ATGGCCCTGTGGATGCGCCTCCTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGCTTCTTCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACTAG”

Using the DNA sequences input by the user along with a few DNA sequences kept as a part of the database in the activity I plan to construct a phylogenetic tree using a distance-matrix method called UPGMA (unweighted pair group method with arithmetic mean) which calculates the genetic distance from multiple sequence alignments. In UPGMA, at each step, we create a similarity/dissimilarity matrix to show how close/far two organisms are. We then combine the two closest organisms to form a higher level cluster. This is a link to a working example of the algorithm.

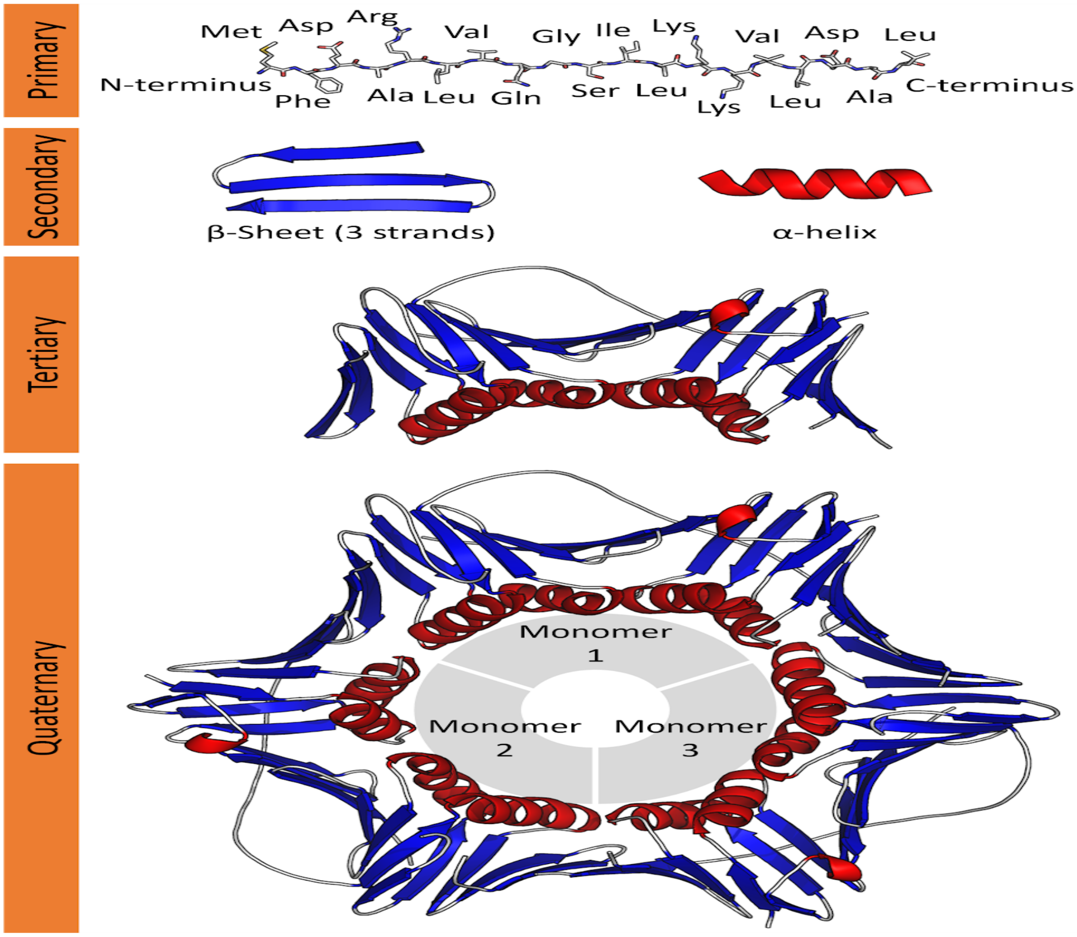
In the activity, there’ll be multiple text boxes along with labels in which the user can input the DNA sequences for different organisms along with their labels. On hitting the “Create” button, the activity will run UPGMA on the given sequences and construct the phylogenetic tree which will then be displayed on the screen. The following is a very basic model of what the layout would look like.


Label1
Label2
Label3
Label4
Sequence 1
Sequence 2
Sequence 3
Sequence 4
Add Organism
Create
Click on Add Organism
Label1
Label2
Label3
Label4
Label4
Sequence 1
Sequence 2
Sequence 3
Sequence 4
Sequence 4
Add Organism
Create
Click on Create
Label1
Label2
Label3
Label4
Sequence 1
Sequence 2
Sequence 3
Sequence 4
Label1
Label2
Label3
Label4
Add Organism
Create

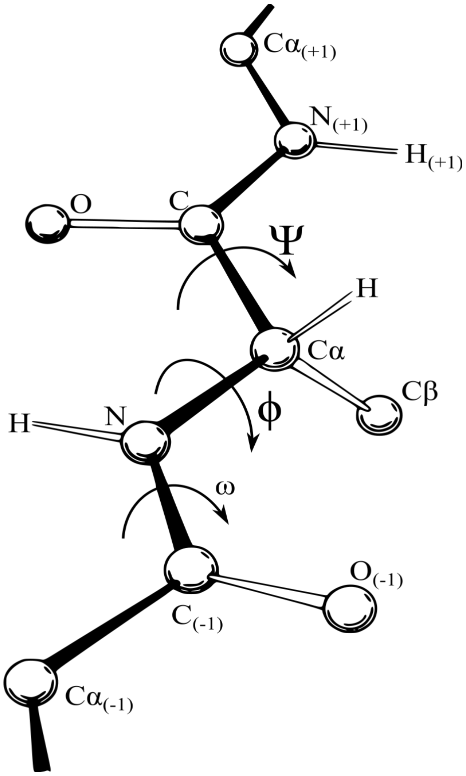
3. Protein Structures

The whole body of an organism is constituted by proteins and for any person working in practically any area of molecular biology, Proteins form some of the most interesting problems which still go unsolved. However, in spite of the fact that we can't make too many theoretical deductions, there are certain things that can be seen across different protein structures.

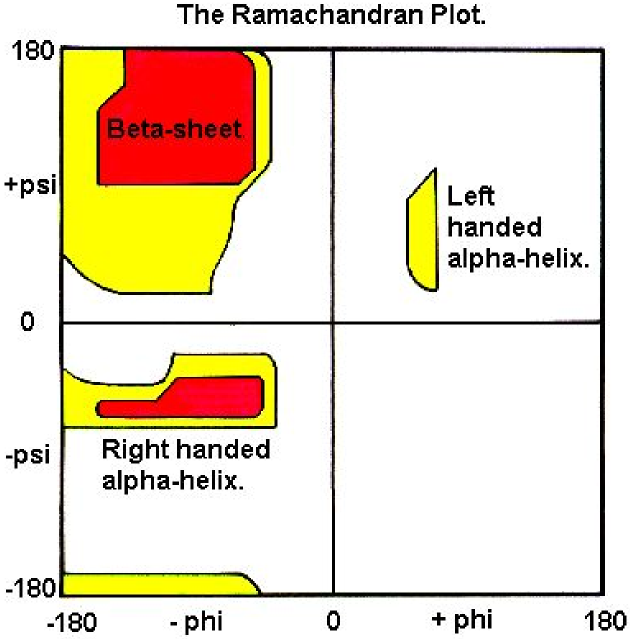
First off, we look at different kinds of protein structures -



The first of these being a pattern in dihedral angles for a protein which was first observed by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan using which they developed the Ramachandran Plot. The dihedral angles are named Φ, Ψ and ω.

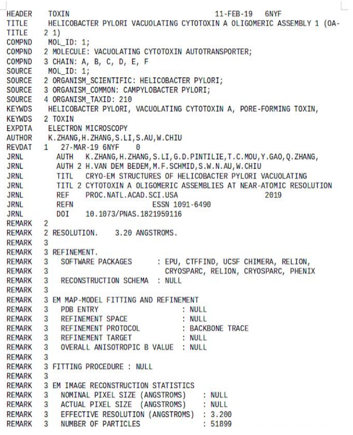


According to the plot, certain dihedral angles can exist only in specific secondary structures.

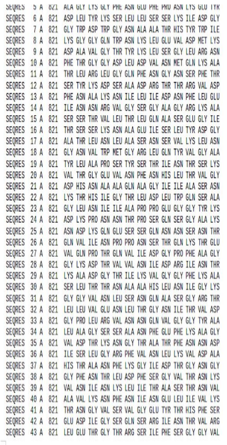


The description of different protein structures is openly available to all in the form of the Protein Data Bank(PDB) where each discovered protein has a unique code corresponding to which there exists a file which describes the position of all the atoms present in the protein.

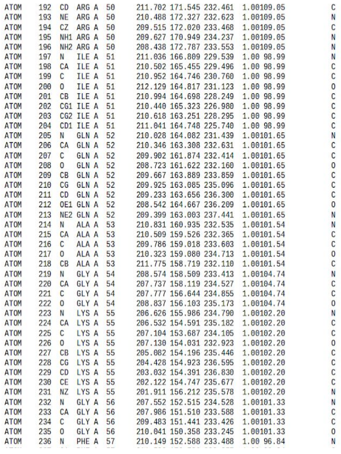
These are a couple of screenshot of the PDB file for 6YNF -



```
HEADER    TOXIN                                   11-FEB-19   6NYF
TITLE     HELICOBACTER PYLORI VACUOLATING CYTOTOXIN A OLIGOMERIC ASSEMBLY 1 (OA-
TITLE    2 1)
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: VACUOLATING CYTOTOXIN AUTOTRANSPORTER;
COMPND   3 CHAIN: A, B, C, D, E, F
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HELICOBACTER PYLORI;
SOURCE   3 ORGANISM_COMMON: CAMPYLOBACTER PYLORI;
SOURCE   4 ORGANISM_TAXID: 210
KEYWDS    HELICOBACTER PYLORI, VACUOLATING CYTOTOXIN A, PORE-FORMING TOXIN,
KEYWDS   2 TOXIN
EXPDTA    ELECTRON MICROSCOPY
AUTHOR    K.ZHANG,H.ZHANG,S.LI,S.AU,W.CHIU
REVDAT   1   27-MAR-19 6NYF    0
JRNL        AUTH   K.ZHANG,H.ZHANG,S.LI,G.D.PINTILIE,T.C.MOU,Y.GAO,Q.ZHANG,
JRNL        AUTH 2 H.VAN DEM BEDEM,M.F.SCHMID,S.W.N.AU,W.CHIU
JRNL        TITL   CRYO-EM STRUCTURES OF HELICOBACTER PYLORI VACUOLATING
JRNL        TITL 2 CYTOTOXIN A OLIGOMERIC ASSEMBLIES AT NEAR-ATOMIC RESOLUTION
JRNL        REF    PROC.NATL.ACAD.SCI.USA                     2019
JRNL        REFN                   ESSN 1091-6490
JRNL        DOI    10.1073/PNAS.1821959116
REMARK   2
REMARK   2 RESOLUTION.    3.20 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   SOFTWARE PACKAGES      : EPU, CTFFIND, UCSF CHIMERA, RELION,
REMARK   3                            CRYOSPARC, RELION, CRYOSPARC, PHENIX
REMARK   3   RECONSTRUCTION SCHEMA  : NULL
REMARK   3
REMARK   3 EM MAP-MODEL FITTING AND REFINEMENT
REMARK   3   PDB ENTRY                    : NULL
REMARK   3   REFINEMENT SPACE             : NULL
REMARK   3   REFINEMENT PROTOCOL          : BACKBONE TRACE
REMARK   3   REFINEMENT TARGET            : NULL
REMARK   3   OVERALL ANISOTROPIC B VALUE  : NULL
REMARK   3
REMARK   3 FITTING PROCEDURE : NULL
REMARK   3
REMARK   3 EM IMAGE RECONSTRUCTION STATISTICS
REMARK   3   NOMINAL PIXEL SIZE (ANGSTROMS)    : NULL
REMARK   3   ACTUAL PIXEL SIZE  (ANGSTROMS)    : NULL
REMARK   3   EFFECTIVE RESOLUTION (ANGSTROMS)  : 3.200
REMARK   3   NUMBER OF PARTICLES               : 51899
```



```
SEQRES   5 A  821  ALA GLY LYS GLY PHE ASN GLU PHE PRO ASN LYS GLU TYR
SEQRES   6 A  821  ASP LEU TYR LYS SER LEU LEU SER SER LYS ILE ASP GLY
SEQRES   7 A  821  GLY TRP ASP TRP GLY ASN ALA ALA THR HIS TYR TRP ILE
SEQRES   8 A  821  LYS GLY GLY GLN TRP ASN LYS LEU GLU VAL ASP MET LYS
SEQRES   9 A  821  ASP ALA VAL GLY THR TYR LYS LEU SER GLY LEU ARG ASN
SEQRES  10 A  821  PHE THR GLY GLY ASP LEU ASP VAL ASN MET GLN LYS ALA
SEQRES  11 A  821  THR LEU ARG LEU GLY GLN PHE ASN GLY ASN SER PHE THR
SEQRES  12 A  821  SER TYR LYS ASP SER ALA ASP ARG THR THR ARG VAL ASP
SEQRES  13 A  821  PHE ASN ALA LYS ASN ILE LEU ILE ASP ASN PHE LEU GLU
SEQRES  14 A  821  ILE ASN ASN ARG VAL GLY SER GLY ALA GLY ARG LYS ALA
SEQRES  15 A  821  SER SER THR VAL LEU THR LEU GLN ALA SER GLU GLY ILE
SEQRES  16 A  821  THR SER SER LYS ASN ALA GLU ILE SER LEU TYR ASP GLY
SEQRES  17 A  821  ALA THR LEU ASN LEU ALA SER ASN SER VAL LYS LEU ASN
SEQRES  18 A  821  GLY ASN VAL TRP MET GLY ARG LEU GLN TYR VAL GLY ALA
SEQRES  19 A  821  TYR LEU ALA PRO SER TYR SER THR ILE ASN THR SER LYS
SEQRES  20 A  821  VAL THR GLY GLU VAL ASN PHE ASN HIS LEU THR VAL GLY
SEQRES  21 A  821  ASP HIS ASN ALA ALA GLN ALA GLY ILE ILE ALA SER ASN
SEQRES  22 A  821  LYS THR HIS ILE GLY THR LEU ASP LEU TRP GLN SER ALA
SEQRES  23 A  821  GLY LEU ASN ILE ILE ALA PRO PRO GLU GLY GLY TYR LYS
SEQRES  24 A  821  ASP LYS PRO ASN ASN THR PRO SER GLN SER GLY ALA LYS
SEQRES  25 A  821  ASN ASP LYS GLN GLU SER SER GLN ASN ASN SER ASN THR
SEQRES  26 A  821  GLN VAL ILE ASN PRO PRO ASN SER THR GLN LYS THR GLU
SEQRES  27 A  821  VAL GLN PRO THR GLN VAL ILE ASP GLY PRO PHE ALA GLY
SEQRES  28 A  821  GLY LYS ASP THR VAL VAL ASN ILE ASP ARG ILE ASN THR
SEQRES  29 A  821  LYS ALA ASP GLY THR ILE LYS VAL GLY GLY PHE LYS ALA
SEQRES  30 A  821  SER LEU THR THR ASN ALA ALA HIS LEU ASN ILE GLY LYS
SEQRES  31 A  821  GLY GLY VAL ASN LEU SER ASN GLN ALA SER GLY ARG THR
SEQRES  32 A  821  LEU LEU VAL GLU ASN LEU THR GLY ASN ILE THR VAL ASP
SEQRES  33 A  821  GLY PRO LEU ARG VAL ASN ASN GLN VAL GLY GLY TYR ALA
SEQRES  34 A  821  LEU ALA GLY SER SER ALA ASN PHE GLU PHE LYS ALA GLY
SEQRES  35 A  821  VAL ASP THR LYS ASN GLY THR ALA THR PHE ASN ASN ASP
SEQRES  36 A  821  ILE SER LEU GLY ARG PHE VAL ASN LEU LYS VAL ASP ALA
SEQRES  37 A  821  HIS THR ALA ASN PHE LYS GLY ILE ASP THR GLY ASN GLY
SEQRES  38 A  821  GLY PHE ASN THR LEU ASP PHE SER GLY VAL THR ASN LYS
SEQRES  39 A  821  VAL ASN ILE ASN LYS LEU ILE THR ALA SER THR ASN VAL
SEQRES  40 A  821  ALA VAL LYS ASN PHE ASN ILE ASN GLU LEU ILE VAL LYS
SEQRES  41 A  821  THR ASN GLY VAL SER VAL GLY GLU TYR THR HIS PHE SER
SEQRES  42 A  821  GLU ASP ILE GLY SER GLN SER ARG ILE ASN THR VAL ARG
SEQRES  43 A  821  LEU GLU THR GLY THR ARG SER ILE PHE SER GLY GLY VAL
```



```
ATOM    192  CD  ARG A  50     211.702 171.545 232.461  1.00109.05           C
ATOM    193  NE  ARG A  50     210.488 172.327 232.623  1.00109.05           N
ATOM    194  CZ  ARG A  50     209.515 172.020 233.468  1.00109.05           C
ATOM    195  NH1 ARG A  50     209.627 170.949 234.237  1.00109.05           N
ATOM    196  NH2 ARG A  50     208.438 172.787 233.553  1.00109.05           N
ATOM    197  N   ILE A  51     211.036 166.809 229.539  1.00 98.99           N
ATOM    198  CA  ILE A  51     210.502 165.455 229.496  1.00 98.99           C
ATOM    199  C   ILE A  51     210.952 164.746 230.760  1.00 98.99           C
ATOM    200  O   ILE A  51     212.129 164.817 231.123  1.00 98.99           O
ATOM    201  CB  ILE A  51     210.994 164.698 228.249  1.00 98.99           C
ATOM    202  CG1 ILE A  51     210.440 165.323 226.980  1.00 98.99           C
ATOM    203  CG2 ILE A  51     210.618 163.251 228.295  1.00 98.99           C
ATOM    204  CD1 ILE A  51     211.041 164.748 225.740  1.00 98.99           C
ATOM    205  N   GLN A  52     210.028 164.082 231.439  1.00101.65           N
ATOM    206  CA  GLN A  52     210.346 163.308 232.631  1.00101.65           C
ATOM    207  C   GLN A  52     209.902 161.874 232.414  1.00101.65           C
ATOM    208  O   GLN A  52     208.723 161.622 232.160  1.00101.65           O
ATOM    209  CB  GLN A  52     209.667 163.889 233.859  1.00101.65           C
ATOM    210  CG  GLN A  52     209.925 163.085 235.096  1.00101.65           C
ATOM    211  CD  GLN A  52     209.233 163.656 236.300  1.00101.65           C
ATOM    212  OE1 GLN A  52     208.542 164.667 236.209  1.00101.65           O
ATOM    213  NE2 GLN A  52     209.399 163.003 237.441  1.00101.65           N
ATOM    214  N   ALA A  53     210.831 160.935 232.535  1.00101.54           N
ATOM    215  CA  ALA A  53     210.509 159.526 232.365  1.00101.54           C
ATOM    216  C   ALA A  53     209.786 159.018 233.603  1.00101.54           C
ATOM    217  O   ALA A  53     210.323 159.080 234.713  1.00101.54           O
ATOM    218  CB  ALA A  53     211.775 158.719 232.110  1.00101.54           C
ATOM    219  N   GLY A  54     208.574 158.509 233.413  1.00104.74           N
ATOM    220  CA  GLY A  54     207.737 158.119 234.527  1.00104.74           C
ATOM    221  C   GLY A  54     207.777 156.644 234.855  1.00104.74           C
ATOM    222  O   GLY A  54     208.837 156.103 235.173  1.00104.74           O
ATOM    223  N   LYS A  55     206.626 155.986 234.790  1.00102.20           N
ATOM    224  CA  LYS A  55     206.532 154.591 235.182  1.00102.20           C
ATOM    225  C   LYS A  55     207.104 153.687 234.105  1.00102.20           C
ATOM    226  O   LYS A  55     207.130 154.031 232.923  1.00102.20           O
ATOM    227  CB  LYS A  55     205.082 154.196 235.446  1.00102.20           C
ATOM    228  CG  LYS A  55     204.428 154.923 236.595  1.00102.20           C
ATOM    229  CD  LYS A  55     203.032 154.391 236.830  1.00102.20           C
ATOM    230  CE  LYS A  55     202.122 154.747 235.677  1.00102.20           C
ATOM    231  NZ  LYS A  55     201.911 156.212 235.578  1.00102.20           N
ATOM    232  N   GLY A  56     207.552 152.515 234.528  1.00101.33           N
ATOM    233  CA  GLY A  56     207.986 151.510 233.588  1.00101.33           C
ATOM    234  C   GLY A  56     209.483 151.441 233.426  1.00101.33           C
ATOM    235  O   GLY A  56     210.041 150.358 233.245  1.00101.33           O
ATOM    236  N   PHE A  57     210.149 152.588 233.488  1.00 96.84           N
```

My plan for the activity is that the user inputs the PDB code(in which case the activity will automatically fetch the file from the internet) for the protein or fetch it from a local file. Then the activity automatically opens the file and fetches data from it. This data will include the length of the protein molecule, the total number of chains present in the protein, ratio of the number of occurrences of a specific amino acid with the total number of amino acid residues, the ligand molecules in the protein and finally the dihedral angles at a few points in the structure and their position on the Ramachandran Plot to show that experimental data is in accordance with the hypothesis presented in the Plot.

4. Technologies Used
   - Language - Python
   - Framework for GUI - PyGtk+3
   - Framework for fetching PDB file - BeautifulSoup
   - Sugar Artworks

Timeline

Dates	Target
May 6 - May 27 (Community Bonding Period)	● Work with my mentor and the community in order to develop communication and find my way around the workings of the organisation. ● Setting up the repositories and creating the working environment. ● Researching and talking to the mentors and finding the best way to implement the features along with explaining the concepts used to the mentors.
May 28 - June 4 (Week 1)	● Creating the UI for the home screen which displays a menu for choosing between the different features in the activity. ● Connect each option with the required module which will be empty for now.

June 5 - June 12 (Week 2)	● Create the GUI for the Phylogenetic Tree Construction section of the activity including the textboxes, labels for each organism and buttons for adding an organism and Creating the tree. ● Connect the Create Tree button to the module in which UPGMA will be written. ● Write the logic for Adding a new label textbox and a DNA sequence textbox on pressing the Add Organism button.
June 13 - June 20 (Week 3)	● Implement the UPGMA module which will run the algorithm on the data input in the textboxes and construct the tree. ● Return the constructed tree to the parent module for drawing on the GUI.
June 21 - June 24 (Week 4)	● Implement the logic for drawing the tree on the canvas in a way which is most easily understandable by the user.
June 24 - June 28 (1st Evaluation)	● 1st Evaluation ○ Deliverable: A working GUI with working Phylogenetic Tree Construction with modules connected appropriately.
June 29 - July 6 (Week 5)	● Create a GUI for the Protein Structure Analysis module which will consist of a textbox which will take the PDB code for the protein. ● Also, add a browse button through which the user can open a local PDB file for analysis. ● Write a web scraping script which can fetch the PDB file according to the code provided from RCSB.
July 7 - July 14 (Week 6)	● Write the script for analyzing the protein file in order to fetch all the required data. ● Return the obtained data to the parent module for display to the user.
July 15 - July 22 (Week 7)	● Display the basic data to the user. ● Create the Ramachandran Plot on which I will plot some of the data points from the lost of dihedral angles.
July 22 - July 26 (2nd Evaluation)	● 2nd Evaluation ○ Deliverable: A standalone GUI with working modules for Phylogenetic Tree Construction,

	Protein Structure Analysis and seamless integration between the modules.
July 27 - August 3 (Week 8)	● Review the existing code for the first 3 days in order to find and fix bugs (if any). ● Conversion of Project to Sugar Activity which includes creating a toolbar and porting it to sugar.
August 4 - August 11 (Week 9)	● Testing and Code Review to check for bugs and if all use cases are being satisfied correctly. ● Use the Sugar Artwork to make the activity look like a part of the sugar environment.
August 12 - August 19 (Week 10)	● Test Run of the activity and fix bugs (if any). ● Final Code Review and writing the project report.
August 19 - August 26 (Final Submission)	● Final Evaluation: ○ Deliverables: Completed Activity -> BioShell along with the Project Report.
After August 26 (Post GSoC)	● Keep working on the activity to add more features and other sugar projects.

The time I will be Able to Give to The Project

I'll be able to give 30-35 hours a week of solid work to the project to create a neat and clean activity. I would also like to stay ahead of schedule so that I can research on how to improve the project while it is in progress and also work on other Sugar projects.

References

- Phylogenetic Tree Diagram - https://en.wikipedia.org/wiki/Phylogenetic_tree#/media/File:Neomuratree.svg
- UPGMA - https://en.wikipedia.org/wiki/UPGMA
- Protein Structures - https://en.wikipedia.org/wiki/File:Protein_structure_(full).png

- Dihedral Angle Diagram - https://en.wikipedia.org/wiki/Dihedral_angle#/media/File:Protein_backbone_PhiPsiOmega_drawing.svg
- Ramachandran Plot Diagram - http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.gif
- 6YNF.pdb - https://www.rcsb.org/structure/6NYF