# Semester Project Documentation (DSA – CS–221)

**Semester Project Title:** BioStructure Explorer: C++ DSA Engine for Computational Biology Analysis

**Student Details**

| Student Name | Student Reg # | Student Degree |
|---|---|---|
| Atta Ur Rahman Sheikh | 2024122 | BSAI |

## 1. Main Features

1. DNA Pattern Search Engine using KMP Algorithm
2. Gene and Protein Interaction Graph Analyzer
3. Mutation Spread Simulation Engine
4. Protein Structure Parser and Analyzer

## 2. Types of Users & Requirements

### 1. Researchers / Biology Students

1. Will be able to upload biological datasets (FASTA, CSV, PDB).
2. Will be able to search DNA patterns efficiently.
3. Will be able to analyze gene and protein interaction networks.
4. Will be able to simulate mutation spread.
5. Will be able to visualize biological data.
6. Will be able to view algorithm execution statistics.

## 3. Requirements Breakdown

### 1. DNA Pattern Search Engine

1. System will allow uploading FASTA files.
2. System will preprocess DNA sequences.

3. System will implement KMP algorithm.
4. System will display all matched positions.
5. System will show execution time and comparisons.

**2. Gene/Protein Interaction Graph Analyzer**

1. System will allow uploading CSV files.
2. System will construct adjacency list graph.
3. System will perform BFS traversal.
4. System will perform DFS traversal.
5. System will calculate degree centrality.

**3. Mutation Spread Simulator**

1. System will configure grid size and probability.
2. System will represent environment as 2D grid.
3. System will simulate mutation using queue.
4. System will update grid at each time step.

**4. Protein Structure Parser**

1. System will upload PDB files.
2. System will parse ATOM records.
3. System will store atom coordinates.
4. System will display structure information.

## 4. Features to Coding Matrix

| Sr # | Feature Name | DSA Concept Used | Operation Performed | Complexity | Variables | Functions | LOC |
|---|---|---|---|---|---|---|---|
| 1 | DNA Pattern Search Engine | Arrays, KMP, Vectors | Pattern Matching | O(n+m) | 6 | 6 | 200 |
| 2 | Gene/Protein Graph Analyzer | Graphs, Queues, Hash Maps | BFS, DFS | O(V+E) | 12 | 8 | 300 |
| 3 | Mutation Spread Simulator | Queues, 2D Arrays | Simulation | $O(n^2)$ | 10 | 6 | 250 |
| 4 | Protein Structure Parser | Arrays, Vectors | File Parsing | O(n) | 8 | 3 | 200 |

## 5. Project Screenshots

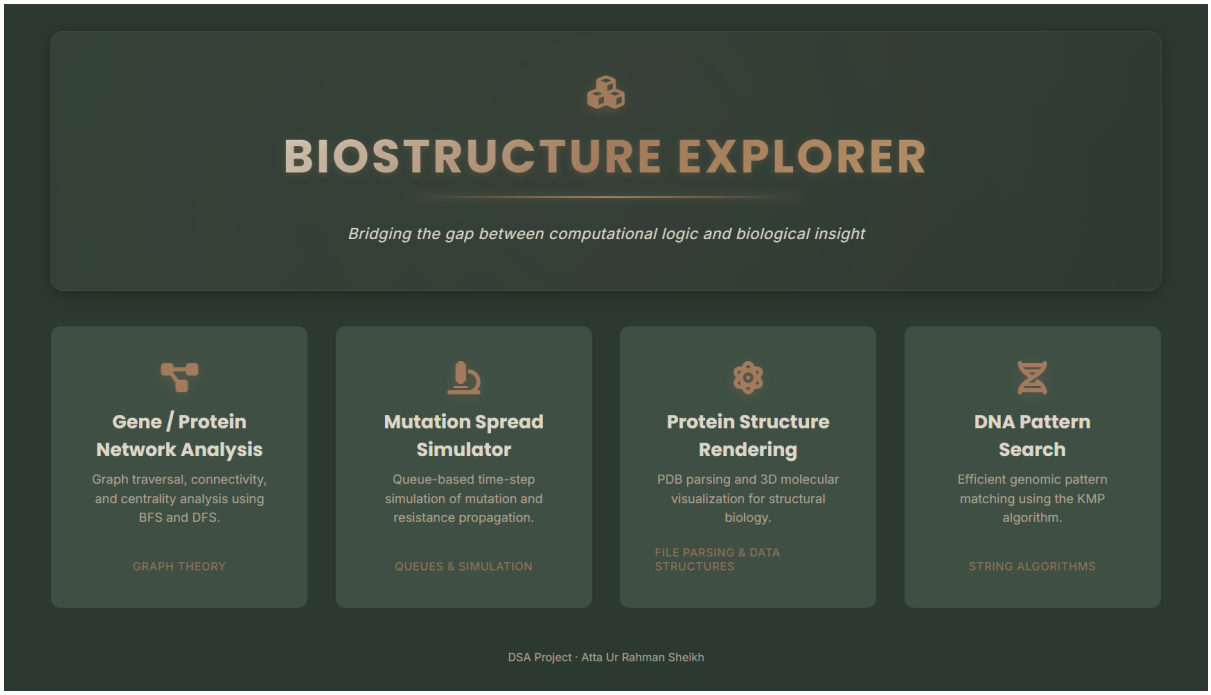


Figure 1: Project Dashboard

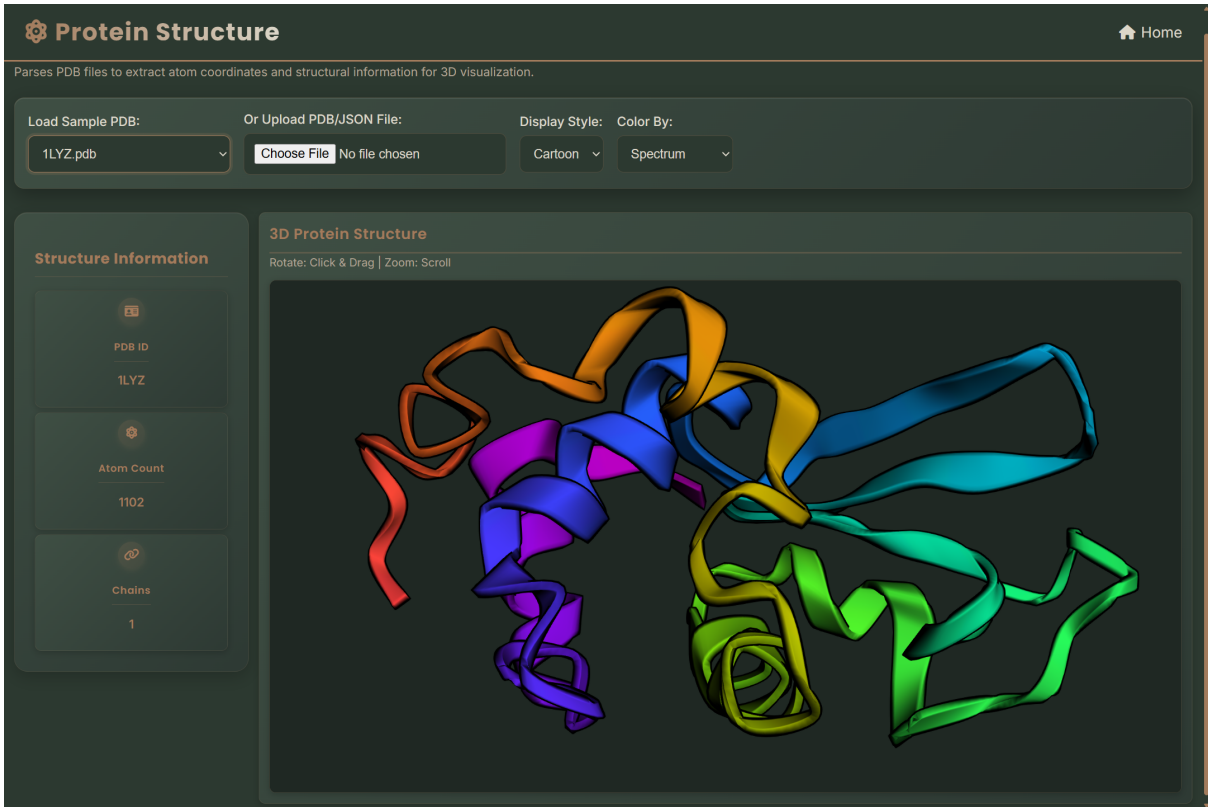


Figure 2: Protein Structure Visualization

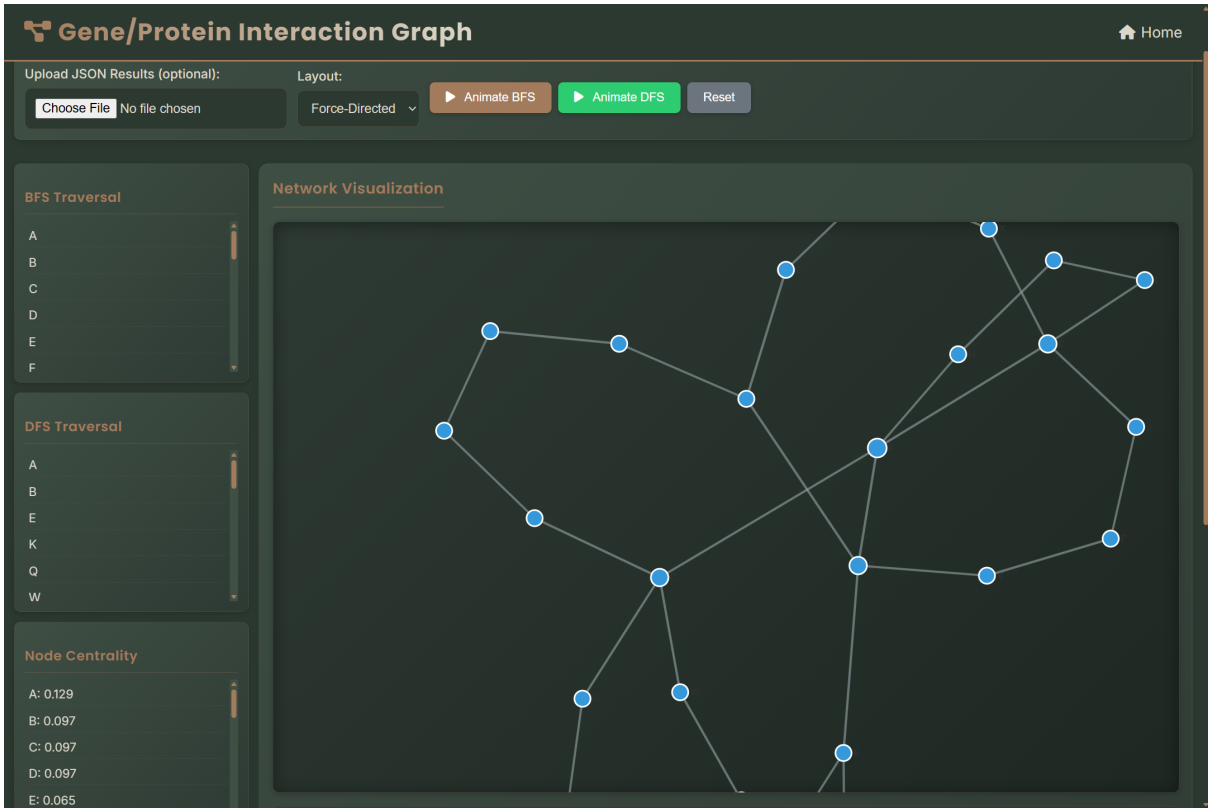


Figure 3: Gene and Protein Interaction Graph

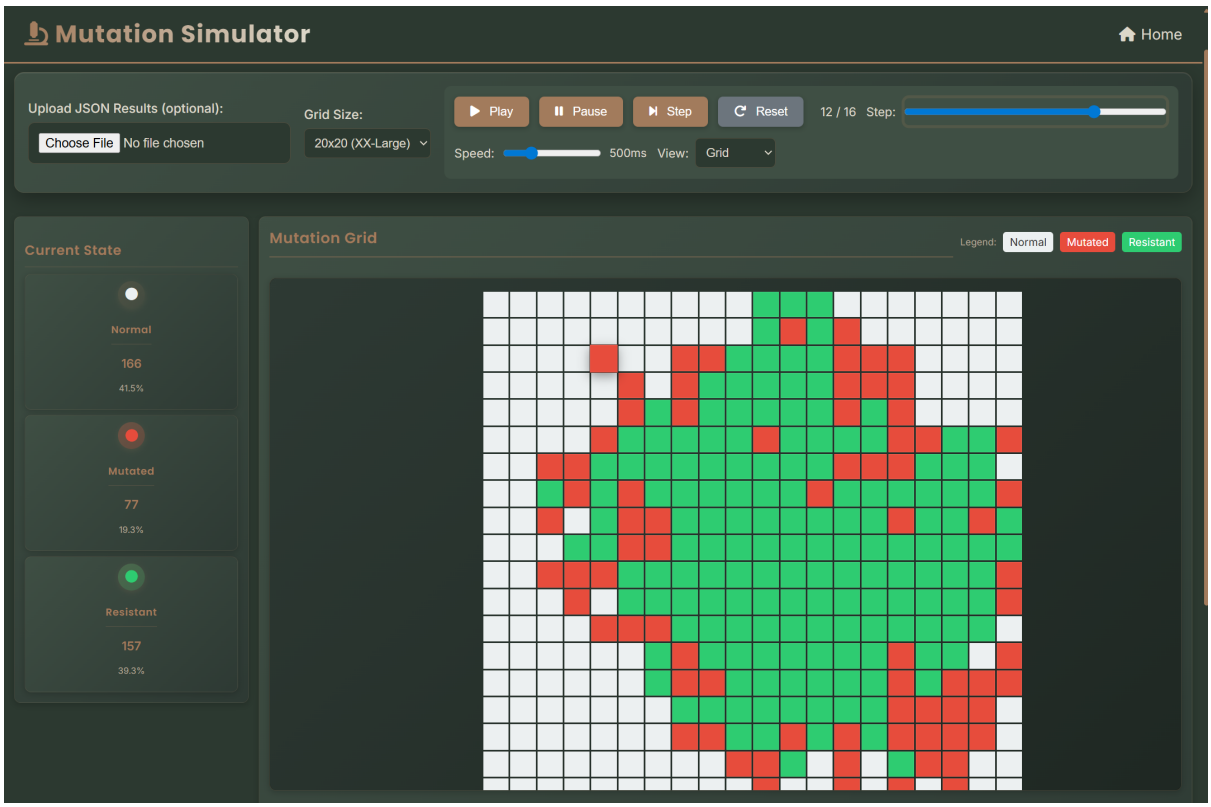


Figure 4: Mutation Spread Simulation

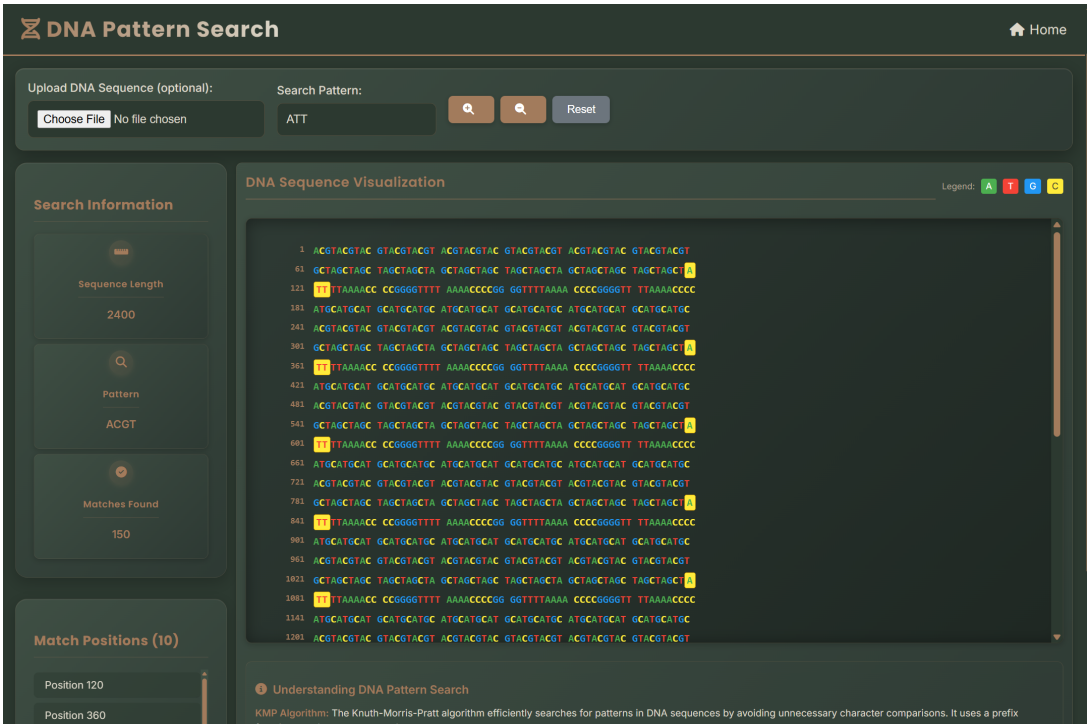


Figure 5: DNA Pattern Search Interface