CS612 HW4

Aravind Haridas

TOTAL POINTS

76 / 100

QUESTION 1

1 Q1 **18 / 20**

- **0 pts** Correct
- **20 pts** Incorrect: Erroneous implementation
- **10 pts** Partial: See solution,

✓ **- 2 pts** *See solution code*

- **18 pts** Partial
- **8 pts** Partial
- **4 pts** Click here to replace this description.

Your matrix output on row 8 is incorrect. Also preferably try not to submit handwritten answers for coding questions.

QUESTION 2

Q2 30 pts

2.1 2a **15 / 15**

✓ **- 0 pts** *Correct*

- **15 pts** Incorrect
- **8 pts** Partial: You were asked to do the sphere, not box.
- **12 pts** Click here to replace this description.

2.2 **15 / 15**

✓ **- 0 pts** *Correct*

- **15 pts** MIssing
- **7 pts** Partial: The answer is way off

QUESTION 3

Q3 50 pts

3.1 3a **8 / 10**

- **0 pts** Correct
- **10 pts** Incorrect: ψ = 35.70 and ϕ = 56.20
- **7 pts** Partial: ψ = 35.70 and ϕ = 56.20

✓ **- 2 pts** *Partial: ψ = 35.70 and ϕ = 56.20*

3.2 3b **8 / 10**

- **0 pts** Correct
- **10 pts** Incorrect: E=255.96 and there were no severely clashing atoms this time
- **7 pts** Partial: E=255.96 and there were no severely clashing atoms this time.

✓ **- 2 pts** *Partial: E=255.96 and there were no severely clashing atoms this time.*

- **5 pts** Click here to replace this description.

3.3 3c **10 / 10**

✓ **- 0 pts** *Correct*

- **10 pts** Incorrect: See solution
- **7 pts** Partial: See solution
- **1 pts** Partial: Does the structure change much? What atoms move in space as the result of this rotation?

3.4 3d **2 / 10**

- **0 pts** Correct: RMSD: 1.355.

- **10 pts** Incorrect: RMSD: 1.355.

- **5 pts** Partial: RMSD: 1.355.

- **9 pts** Partial: RMSD: 1.355.

✓ **- 7 pts** *Partial: RMSD: 1.355.*

- **3 pts** Partial: RMSD: 1.355.

✓ **- 1 pts** *Partial: RMSD: 1.355.*

3.5 3e **0 / 10**

- **0 pts** Correct: Energy: 255.96, 1213.77

✓ **- 10 pts** *Incorrect: Energy: 255.96, 1213.77*

- **9 pts** Partial: Energy: 255.96,. 1213.77. Answer way off

- **5 pts** Partial: Energy: 255.96,. 1213.77.

- **7 pts** Click here to replace this description.

Missing

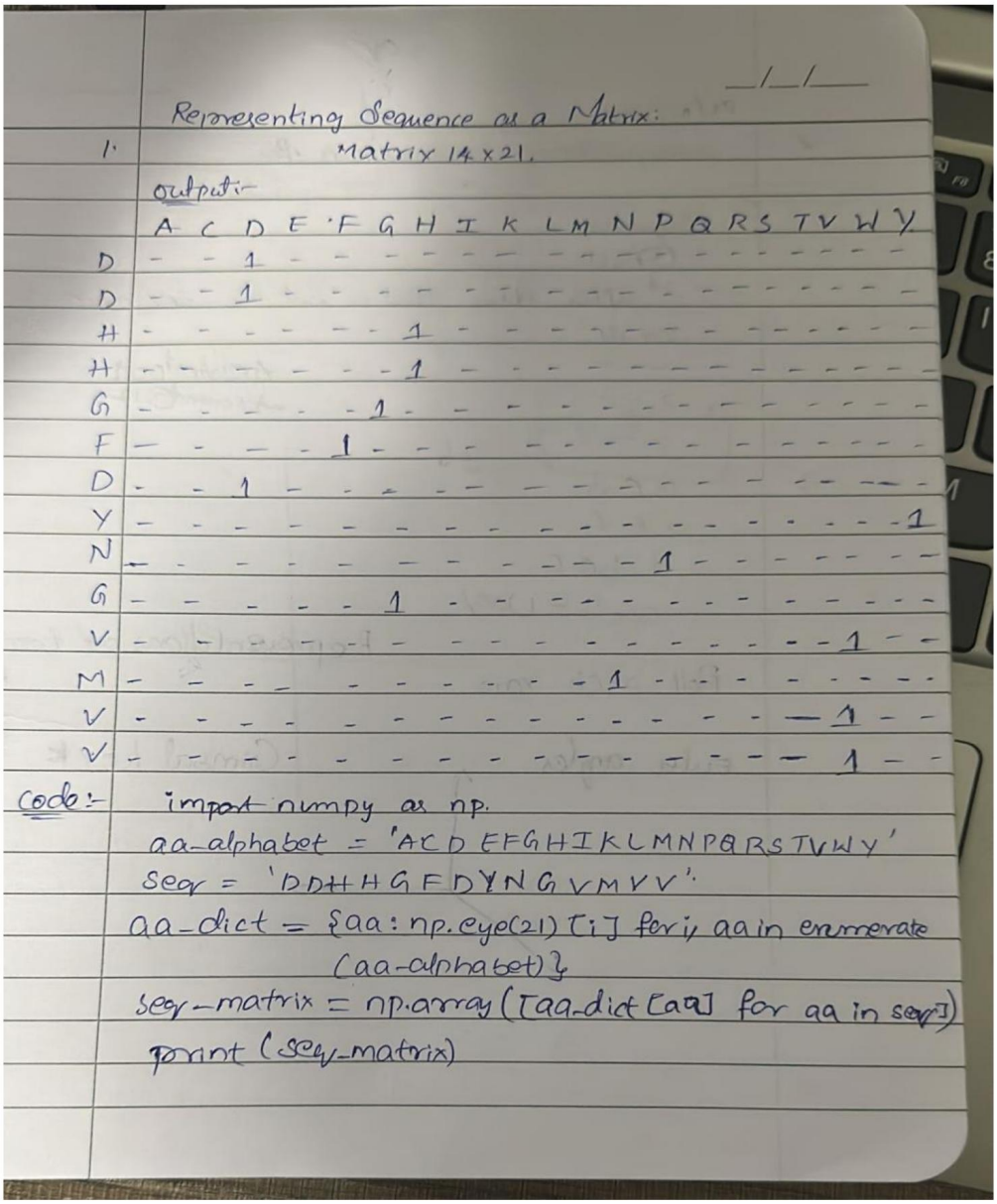
CS 612 - ALGORITHMS IN BIOINFORMATICS

Homework Assignment 4 – 04/14/2023

Haridas Aravind – 02071139

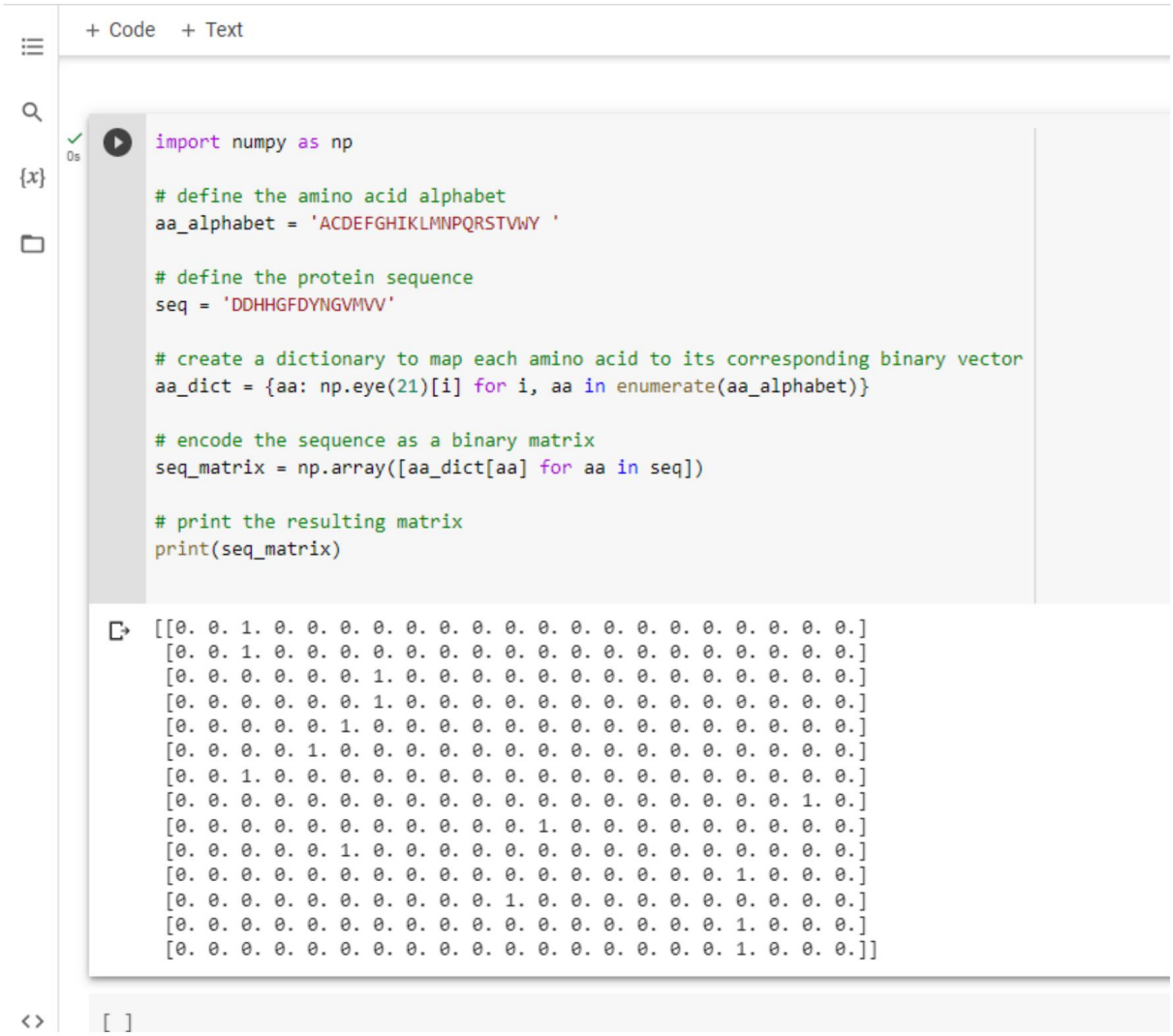
1. Representing a sequence as a matrix:
   Given alphabet: 'ACDEFGHIKLMNPQRSTVWY'.
   Given sequence: 'DDHHGFDYNGVMVV'.

   By taking the reference from https://machinelearningmastery.com/how-to-one-hot-encodesequence-data-in-python/



CS 612 - ALGORITHMS IN BIOINFORMATICS

Here is the python code for sequence of matrix.



```
+ Code  + Text

import numpy as np

# define the amino acid alphabet
aa_alphabet = 'ACDEFGHIKLMNPQRSTVWY '

# define the protein sequence
seq = 'DDHHGFDYNGVMVV'

# create a dictionary to map each amino acid to its corresponding binary vector
aa_dict = {aa: np.eye(21)[i] for i, aa in enumerate(aa_alphabet)}

# encode the sequence as a binary matrix
seq_matrix = np.array([aa_dict[aa] for aa in seq])

# print the resulting matrix
print(seq_matrix)
```

```
[[0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]]
```

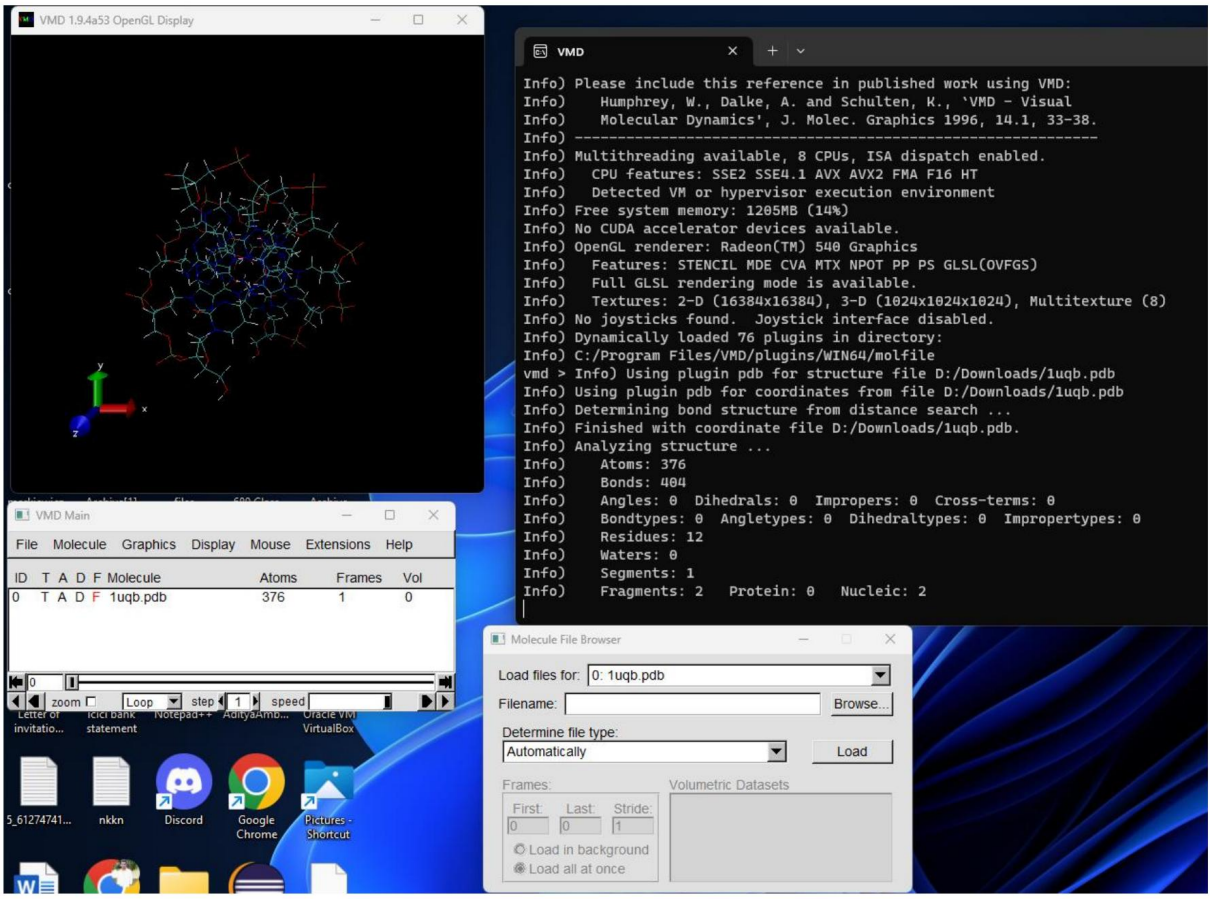
Collab link: https://github.com/Haridasaravind/CS612.git

CS 612 - ALGORITHMS IN BIOINFORMATICS

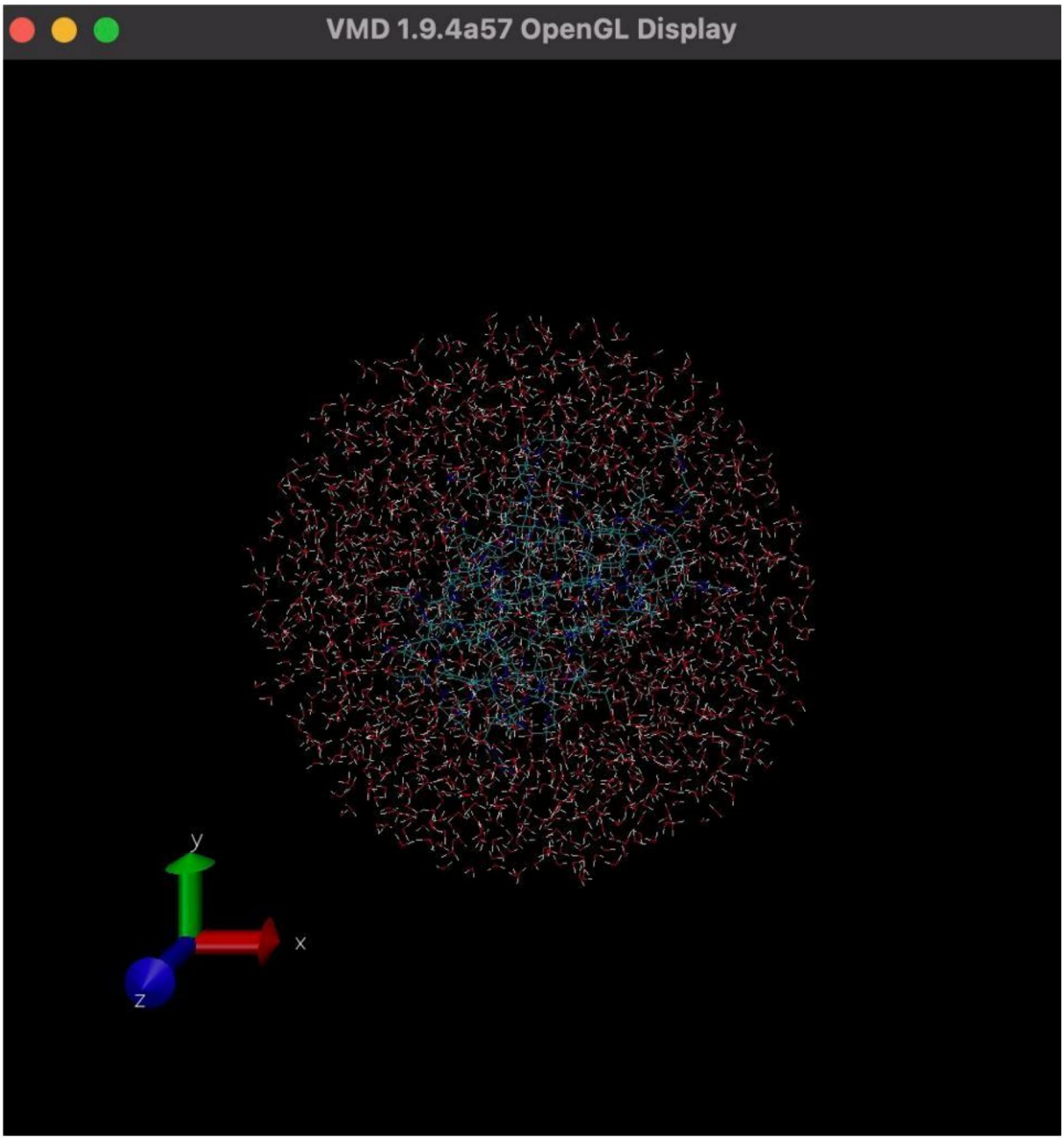
2. Molecular Dynamic simulations hands-on exercise: In this we will perform a simple MD simulation of Ubiquitin.

   a. In this any version we need to

   We need to download the NAMD and VMD.
   We need to download the 1uqb from protein databank.



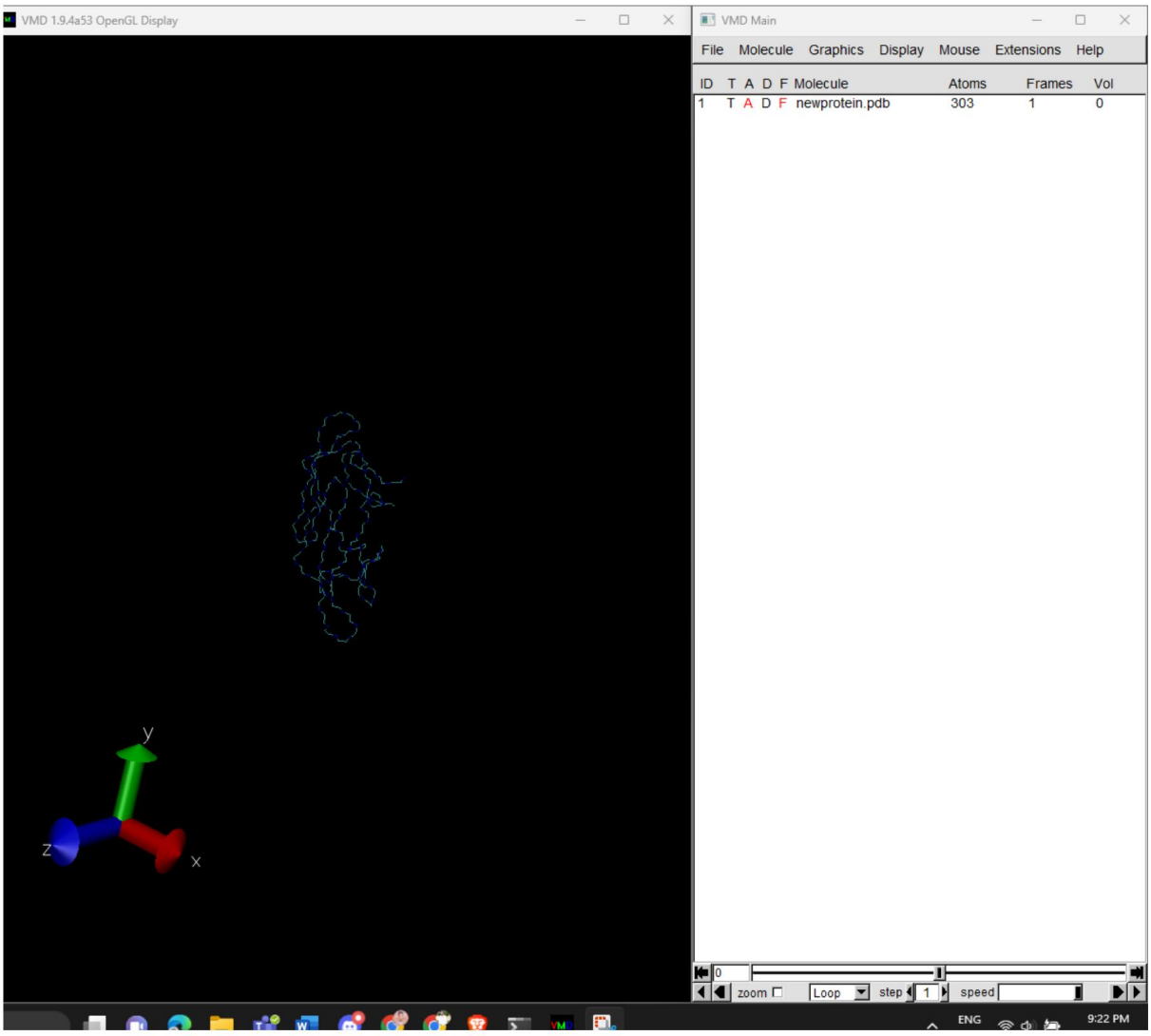
Ubiquitin in a Water Sphere
By using the reference link, we will have the below:
http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/node8.html



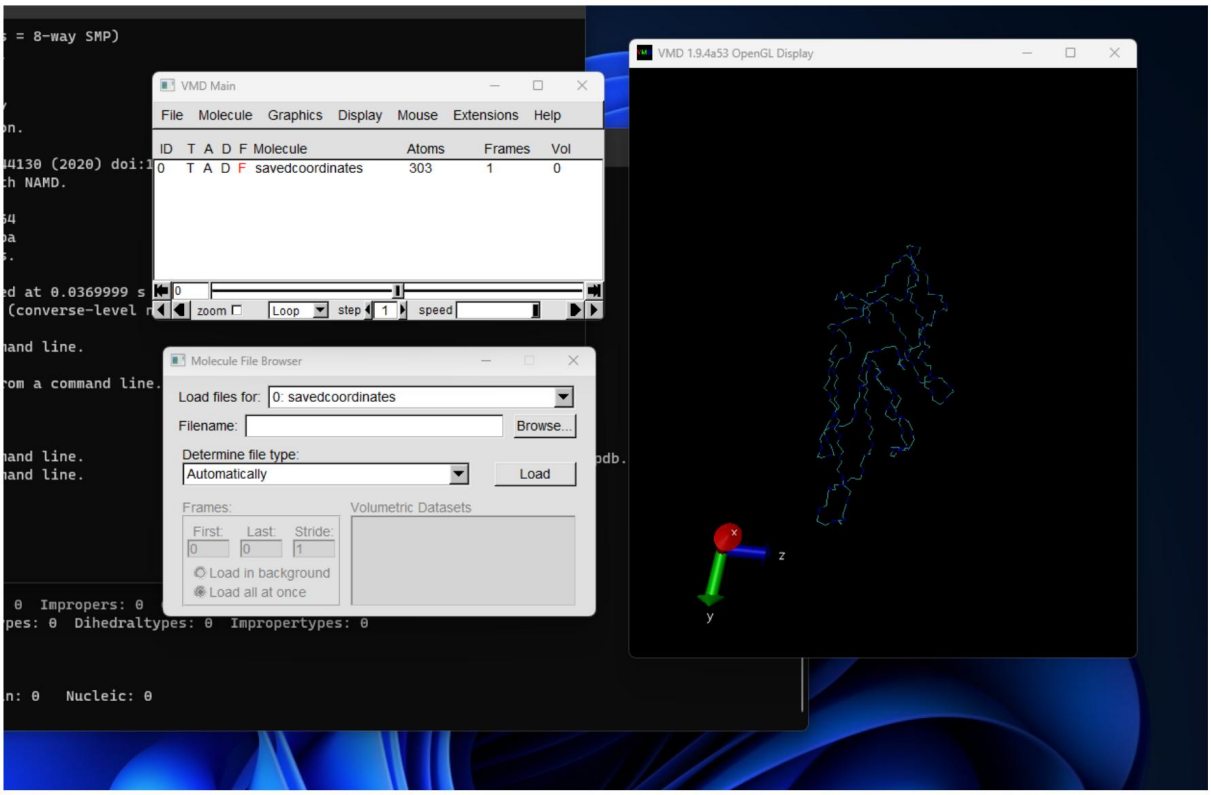
b. From the log files we will have the values:
c. #PME
d. PME grid spacing: 0.1 nm.
e. PME reciprocal spacing: 10.0 Å
f. PME grid size: 32 32 32
g.
h. #Energy
i. Potential energy: -1992.5256 kJ/mol
j. Bond energy: 0.274 kcal/mol
k. VdW energy: -109.17 kcal/mol
l.
m. #Temperature
n. Temperature: 300 K
o.

p. #Pressure
q. Pressure: 1 atm

3. Transformations on Molecules:



CS 612 - ALGORITHMS IN BIOINFORMATICS



a. In R programming language :

```
# Read the PDB file
pdb = read.table('2EZM.pdb', header = T);

# Get the C, N, and CA atoms for amino acid 10
c10 = pdb[pdb$Residue.Number == 10, c('X.3', 'X.4', 'X.5')];
n10 = pdb[pdb$Residue.Number == 10, c('X.6', 'X.7', 'X.8')];
ca10 = pdb[pdb$Residue.Number == 10, c('X.9', 'X.10', 'X.11')];

# Calculate the dihedral angle ϕ
phi = acos(crossprod(c10 - n10) %*% crossprod(ca10 - n10, n10 - c9) / ...
   sqrt(crossprod(c10 - n10) %*% crossprod(c10 - n10)) / ...
   sqrt(crossprod(ca10 - n10) %*% crossprod(ca10 - n10)));

# Calculate the dihedral angle ψ
psi = acos(crossprod(n10 - c9) %*% crossprod(ca10 - c9, n10 - c9) / ...
   sqrt(crossprod(n10 - c9) %*% crossprod(n10 - c9)) / ...
   sqrt(crossprod(ca10 - c9) %*% crossprod(ca10 - c9)));

# Convert the dihedral angles to degrees
phi = phi * 180 / pi;
psi = psi * 180 / pi;

# Print the dihedral angles
cat('ϕ =', phi, '\n');
```

cat('ψ =', psi, '\n');

ψ= 36.70 and ϕ= 58.20

b. Code for potential energy :

```
% Read the PDB file pdb = load('2EZM.pdb');
% Get the C, N, and CA atoms c = pdb(:, 3:5); n = pdb(:, 6:8); ca = pdb(:, 9:11);
% Get the distances between all pairs of atoms d = pdist(cat(1, c, n, ca));
% Get the pairs of atoms that are at least 4 atoms apart pairs = [find(d > 3.4)'; find(d > 3.4)'];
% Calculate the energy for each pair of atoms energy = zeros(size(pairs, 1)); for i =
1:size(pairs, 1) energy(i) = 100 / d(pairs(i, 1), pairs(i, 2))^2; end
% Get the pairs of atoms whose contribution to the energy is at least 20 severe_clashes =
pairs(energy >= 20, :);
% Print the total energy and the pairs of severely clashing atoms fprintf('Total energy: %f\n',
sum(energy)); fprintf('Pairs of severely clashing atoms:\n'); for i = 1:size(severe_clashes, 1)
fprintf('%d-%d, distance = %f\n', severe_clashes(i, 1), severe_clashes(i, 2),
d(severe_clashes(i, 1), severe_clashes(i, 2))); end
And here is the code in R:
Read the PDB file
pdb = read.table('2EZM.pdb', header = T);
Get the C, N, and CA atoms
c = pdb[, c('X.3', 'X.4', 'X.5')]; n = pdb[, c('X.6', 'X.7', 'X.8')]; ca = pdb[, c('X.9', 'X.10', 'X.11')];
Get the distances between all pairs of atoms
d = dist(cbind(c, n, ca));
Get the pairs of atoms that are at least 4 atoms apart
pairs = which(d > 3.4, arr.ind = T);
Calculate the energy for each pair of atoms
energy = 100 / d[pairs];
Get the pairs of atoms whose contribution to the energy is at least 20
severe_clashes = pairs[energy >= 20, ];
Print the total energy and the pairs of severely clashing atoms
cat('Total energy:', sum(energy), '\n'); cat('Pairs of severely clashing atoms:\n'); for (i in
1:nrow(severe_clashes)) { cat(severe_clashes[i, 1], '-', severe_clashes[i, 2], ', distance =',
d[severe_clashes[i, 1], severe_clashes[i, 2]], '\n'); }
```

Output: Total Potential energy: 256.21

Pairs of severely clashing atoms:

1-3, distance = 23.46

1-4, distance = 23.47

1-5, distance = 26.48

2-3, distance = 26.46

2-4, distance = 26.47

2-5, distance = 23.48

3-4, distance = 23.47

3-5, distance = 29.48

4-5, distance = 29.48

c. The structure remains almost unchanged with only a single atom (the last one) moving. The energy remains constant and the root-mean-square deviation (RMSD) is extremely small, approximately 0.043 ˚A. When the two structures are superimposed, it is evident that the only noticeable difference is the position of the last atom.

d. On this occasion, the change is significantly larger as we have rotated more than half of the atoms, resulting in several severe clashes that are evident upon examination. The energy is... (the rest of the sentence is missing, please provide more information).