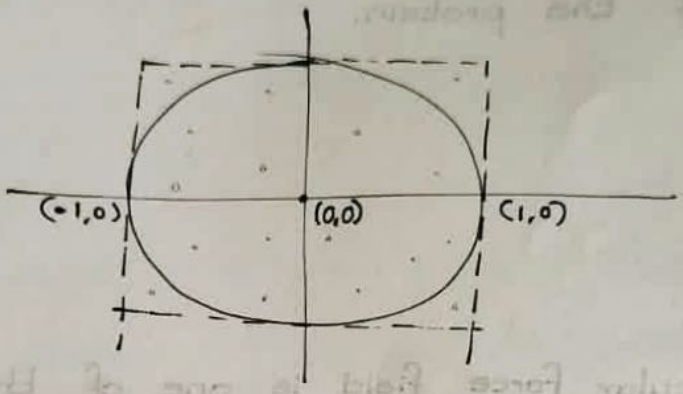
Q.1

The following is a computational algorithm to approximate the value of π. As the number of darts increase, the approximation tends to, or 'converges' on the actual irrational value of π.



Algorithm

```
for (i=0; i<m; i++){
    x = random (-1, 1)      // This simulates a dart landing on
    y = random (-1, 1)          some co-ordinates (x,y) which
                                lie inside square
    if (x²+y²<1) {          // distance from origin is less than 1
        n = n+1             // 'n' is no. of darts inside circle
    }
    m = m+1                 // 'm' is total number of darts
}
π = 4 · n/m                 // since n/m = πr²/(2r)²
```

Principle of Convergence

Just as by increasing number of steps we converge on the actual value of π, while generating stable conformations of a polypeptide chain we slowly converge on the actual 3-D structure of the protein.

Q.2

Force Field

A molecular force field is one of the 3 essential components of molecular Dynamic Simulation, (other 2 being protein structure & integration step)

A force field tries to approximate all the forces applied on the atoms in order to predict their motions

There are 2 major categories of force field components ;

1) Bonded Interactions
2) Non-Bonded interactions

Bonded : Electrostatic

Unbonded : $-\sum VDW + \sum_{Solvation} + \sum_{Hydration}$

Q.3

Integration step is the 3rd important component of MD simulation. We apply the force-field on the atoms of protein and calculate the position (r) & momentum (p) changes of atoms over infinitesimal time durations. This allows us to mimic how the molecules will behave in actual in-vitro conditions.

Verlet Algorithm :

We need to be able to predict the position and momentum of molecule after infinitesimal time 'dt'. But starting velocity can be anything, hence we need to eliminate velocity term

5½

We know,

$$r(t + dt) = r(t) + v(t) \cdot dt + a(t) \cdot dt^2 \ldots\ldots$$

$$r(t - dt) = r(t) + -v(t) \cdot dt + a(t) \cdot dt^2 \ldots\ldots$$

Adding,

$$r(t + dt) = 2 \cdot r(t) - r(t - dt) + 2 \cdot a(t) \cdot dt^2 + (0dd) \cdot dt^3$$

Hence if we know position at a moment, and position at time dt before it, we can predict future position at 't + dt', without need of explicit velocity

Q.4

The search space of a protein confirmation is of an astronomical order of magnitude. In a polypeptide sequence of just 100 residues, considering any of the 20 amino acids can be at a given point, the search space becomes 20^{100}. If we consider all of the rotamer conformations, it becomes 20^{500}.

In order to accomodate this, we allow the integration step to run for almost indefinite time. This allows us to parse through the large search space of possible confirmations, thereby increasing statistical probability of generating enough confirmations

Ergodic hypothesis

The ergodic hypothesis states that we can group atoms into 'ensembles'. The atoms inside an ensemble might have different microscopic properties, but have the same average macroscopic properties. Hence our confirmations get seperated into distinct buckets, which mimics how proteins behave in labs wherein they have few stable, functional conformations.

Q.5

A) In de-novo peptide design, we substitute different amino acids to a known poly-alanine 3-d structure and check if the energy is minimum.

1/3/2

The solvation energy depends upon the Solvent Accessible Surface Area (SASA) of the residue. We run a program such as naccess, which runs a 1 Å probe representing water molecule over the protein, thereby finding SASA.

Since Tryptophan is in the core, we can expect a low SASA, which is good since tryptophan is hydrophobic.

B)

i) Dead End Elimination

1/3/2

If we do reach a dead end on trans rotamer of phenylalanine, we can safely eliminate rotamers of a higher energy.

Hence all given rotamers except 'g-' conformation of Tryptophan are eliminated.

ii) Goldstein DEE

In G. DEE, we allow rotamers 'a' within a small ΔE of energy, all rest are eliminated.

Hence all rotamers except Trp 'g-' and Phe 'g-' will be eliminated.

Hence elimination algorithms converge on specific rotamer.

Q.6

1. **Ensemble** :

Ensemble is a group of molecules with different microscopic properties but same macroscopic properties

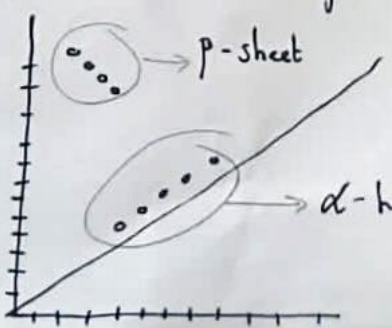
$$\langle A_{vs} \rangle_{micro} = \langle A_{vs} \rangle_{macro}$$

(2)

2. **Contact map**

Drawn with the primary sequence as distinct points on X & Y axis, contact map plots which amino acids contact which other amino acids. We can interpet secondary structures from it.

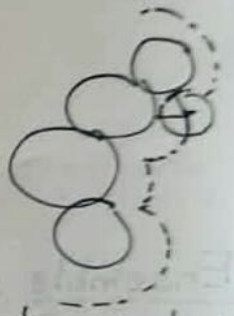
e.g.) α-helix, β-sheet (2)



3. United atom approximation

4. Solvent Accessible Surface Area

It is theoretical area where solvent can go, calculated by rolling an imaginay probe (1A°) representing solvent over the surface of molecule.



Programs such as naccess approximate it by slicing the molecule into z-slices and rolling a circle on contour.

5. Hydrogen Bond

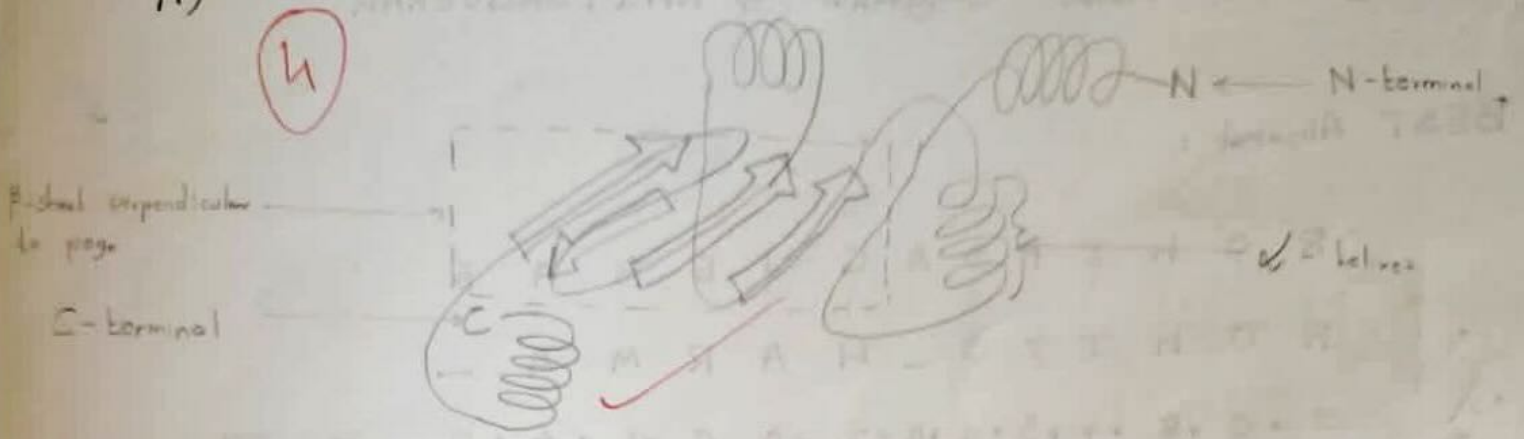
The attraction b/w a hydrogen connected to a highly electronegative atom such as F, O, N is called H-bond.

It is especially useful in holding the protein in a stable conformation.

Q.7

A)

4



7)

1	C				
2	H	1	109.5		
3	H	1	109.5		
4	H	1	109.5	180°	
5	C	1	109.5	2	-60°
6	H	5	109.5	2	-60°
7	H	5	109.5	2	180°
8	H	5	109.5	2	+60°

3½

Q.8.

My name : Rohil Sagvekar ⇒ RHITSAGVEKAR

1) BEST Alignment :

R	O	H	I	T	S	A	G	V	E	K	A	R
R	O	H	I	T	S	_	H	A	R	M	A	_
5	+0	+8	+4	+5	+4	+0	-2	-0	-0	-1	+4	+0	= 27

7 1/2

2)

S	M	R	I	T	I	M	A	N	D	A	N	A
R	_	H	I	T	S	A	G	V	E	K	A	R
-1	+0	+0	+4	+5	-2	-1	+0	-3	+2	-1	-2	-1	= 1

3)

S	R	Y	A	K	M	A	R	Y	A	D	A	V
R	H	I	T	S	A	G	V	E	K	A	R	
-1	-1	-1	-1	+4	-2	-1	-1	-1	+4	-3	= 4	

Decreasing order = 1 > 3 > 2

If we reverse sign, order will get reversal