Introduction to Biophysical Chemistry

Using Monte Carlo Simulations to unfold a 16-mer protein in a 2D lattice

- Sai Guha (BS17B009)
- Dasari Aswin (BS17B010)
- G Prashant (BS17B011)

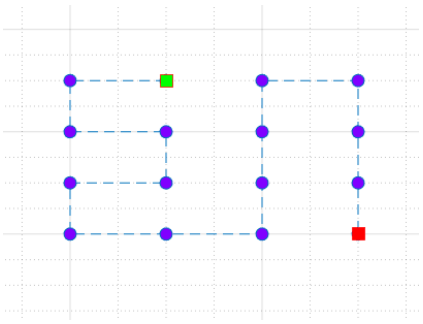
Introduction

Protein folding is the process by which a protein structure assumes its functional shape or conformation. Although any unique polypeptide may have more than one stable folded conformation, each conformation has its own biological activity and only one conformation is considered to be the active one. When this native structure of the protein is interrupted or denatured, it will not be able to carry out its specific function. Denaturation involves either partially or totally unravelling of the protein, or re-organizing the hydrogen bonds which gives the protein its native higher level of structure.

Monte Carlo Simulations

In order to understand the thermodynamic mechanisms behind the denaturation or unfolding of proteins, we use Monte Carlo simulations to compute the several possible pathways. Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. In this assignment, the Metropolis algorithm was used in the Monte Carlo simulations. Metropolis simulation allows the protein to change its conformation from a more stable state to a less stable state with some probability, which is determined by the probability condition parameter.

The simulation starts with an arbitrary initial native state conformation, which is shown below.



Here, the green residue is the N-terminus amino acid and the red residue is the C-terminus amino acid. All other amino acid residues of the peptide chain are labelled as violet squares. It can be noticed that there are 9 non covalent interactions in this native structure. If the energy of each interaction is *e* units, the total native state energy of the protein can be calculated as

$$E = 9 * e \text{ units}$$

Two simulations of 1 million steps each are run, substituting the value of *e* as -1 units in one case and -1.5 units in the other. Whether a move happens are not is determined stochastically by calculating the probability condition parameter ω is given by

$$\omega = \exp\left(-\left(E_f - E_i\right) / kT\right)$$

Here, Ef is the total energy of the conformation it would have after making the move, Ei is the total energy of the conformation before making the move, k is the Boltzmann constant and T is the temperature. For simplicity, we assume the value of kT as 1.

If the value of ω is greater than 1 (thermodynamically favourable move), we proceed forward to make the move. If the value of ω is less than 1 (thermodynamically unfavourable move), a random number between 0 and 1 is uniformly generated and the move is made if that number is less than ω.

The structure and the total energy of the protein conformation is displayed after each iteration.

Plots of the total interaction energy of the system versus the steps are made for both the values of interaction energy.

There are two kinds of Monte Carlo moves which are possible:

- The lattice points can undergo Corner move, where only 90^{o} move is possible.
- The residues between the lattice points can undergo Crank shaft move, which is possible only when the residues i-1, i, and i+1 form a right angle.

Both of the above-mentioned moves shift only one residue to another location, provided the corresponding location is not already occupied by another residue of the peptide chain.

Basic Documentation of the code

The code consists of a main function named `ProteinFold(e,kt,num_iter)` and two supplementary functions named `Energy(e,x,y)` and `EnergyNew(e,amino,x,y)`

`ProteinFold(e,kt,num_iter)`
This function initializes the native state of the protein and finds the energy of the current conformation, having arguments e (Energy of a single non-covalent interaction) and kT (product of k and temperature). Further, it simulates the unfolding of proteins using Monte Carlo Method. A plot is displayed after each iteration to show the current state of the protein. This function makes use of two functions `Energy()` to calculate the energy of the native state protein and to store the indices of native interactions and `EnergyNew()` to calculate the energy of the current state.

`Energy(e,x,y)`
Function to calculate the energy of the native state of the protein.

Arguments:
e - Energy of a single non-covalent interaction
x and y - coordinates of the native state protein

Returns:
E - The calculated total energy of the native state
new - A cell array which consists of the coordinates
amino - A 16x3 cell array which keeps track of the native state interactions.

`EnergyNew(e,amino,x,y)`
Function to calculate the energy of the current conformation of the protein.

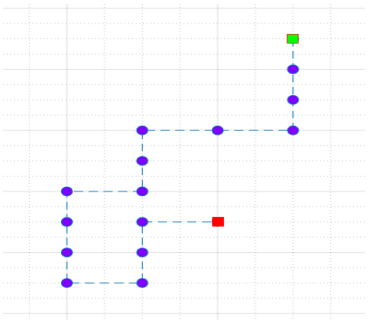
Arguments:
e - Energy of a single non-covalent interaction
x and y - coordinates of the native state protein
amino - A 16x3 cell array which keeps track of the native state
interactions

Returns:
E - The calculated total energy of the current conformation

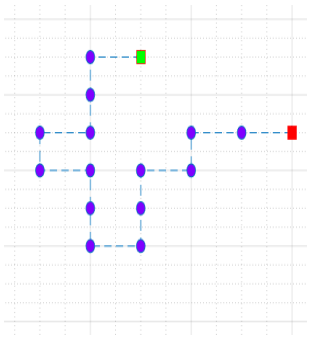
Observations

The conformation and the energy of the protein after 1 million iterations when *e*=-1 units and *e*= -1.5 units are given below:

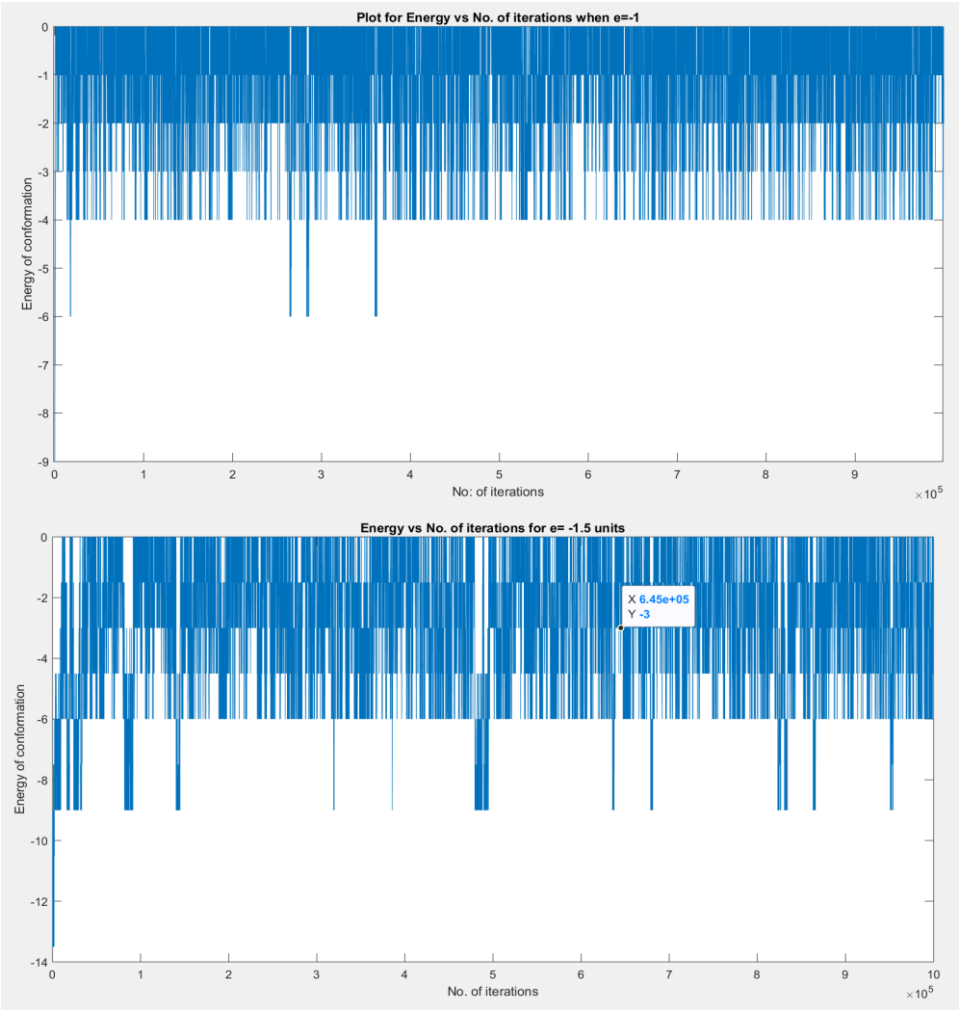
For e=-1 units, E=-2 units



For e=-1.5 units, E = 0 units



The plots for the Energy vs No. of iterations when e=-1 and e=-1.5 are given below:



Inferences

- When the simulation is run, the protein with lower interaction energy (-1 units) takes very few iterations to make its first move, while the protein with higher interaction energy (-1.5 units) takes comparatively greater number of iterations to make its first move.

- From the above plots of the Energy vs Number of iterations, it can be interpreted that if the energy of interaction is higher in magnitude, it takes a greater number of iterations to unfold and also has a higher chance of refolding to a more stable state.

- If the total energy of the conformation reaches 0 at any point, then it remains zero for some iterations and tries to refold back, thereby decreasing the energy.

- Furthermore, after plenty of iterations, the protein starts drifting away and moves out of the limits of the graph. Again, if the interaction energy if lesser in magnitude, it drifts away after a smaller number of steps when compared to the case when the interaction energy is higher in magnitude.

Conclusion

Any thermodynamic system without any constraint always has the drive to reach the minimum energy state. It also has a probability of transforming into a higher energy state, which depends of the temperature and the energy scale of the system. If the energy scale is lesser in magnitude, unfolding of protein happens faster because the non-covalent interactions are easier to break. Likewise, if the temperature is higher, the probability of denaturation is higher, as given by the probability equation, thereby favouring a higher energy state. Denatured proteins can be termed as protein structures with high energy state due to absence of hydrogen bonds and other Van der Waal's forces as compared to the native or folded state.

References

http://www.me.umn.edu/~dtraian/Project-Ramji.pdf
https://www.cse.iitb.ac.in/~anandb/mtech_thesis.pdf
https://en.wikipedia.org/wiki/Protein_folding