Available online at www.easletters.com

ISSN No: 2349-2864

Entomology and Applied Science Letters, 2016, 3, 3:13-19

Coarse-grained Gö-model Simulation of Villin: Understanding the Folding Mechanism

Sanjana S.[1], Aaron James S.[2*] and Monisha M.[3]

[1]*Student, Department of EIE, Sathyabama University, Chennai, India*
[2]*Assistant Professor, Department of EIE, Sathyabama University, Chennai, India*
[3]*Independent scholar & Freelancer, Computer science Dept, Chennai, India*
Corresponding email:aaronjameszion@yahoo.com

ABSTRACT

Every biological protein has a specific function. Protein folding is the process by which a protein structure assumes its functional shape or conformation. Thus, studies on protein folding have been very essential and useful to understand the various intricacies in the function of a protein. To understand the folding mechanism, a 36 residue protein called 'Villin' (PDB ID -1VII) has been studied in silico in this research work. Computational methods have been used to determine the secondary structure of the protein while the process of folding happens. Coarse grained molecular dynamics has been used with GROMACS, so as to get a faster chemical simulation of the Villin molecule. The molecule was simulated at various temperatures with the same energy functions. Since Go model was used for the simulations, a reduced Kelvin temperature is used as most of the unwanted bonds are removed. Simulations were done for 30 different temperatures, thus creating a different protein folding trajectory for each temperature. Then, for each of the temperatures, RMSD and native contact Q values were found out by using appropriate commands. These values were plotted along with the Mean Q and Mean RMSD values to find out the exact temperature where there is equilibrium between folded and unfolded regions. This is found through the inflexion point in the plots. This temperature was found to be 116 rK (Reduced Kelvin). Now, at this temperature, we needed the time-steps where the actual transition occurs. This was found out by taking the Q values between 0.35 and 0.65. Therefore the time steps where the protein folding transition from unfolded to folded and vice versa occurred were computed. This data was then used to calculate the trajectories at those time steps. The trajectories were then used to compute the secondary structure (DSSP) using a software called STRIDE. Contact maps and Ramachandran plots were also computed for the required transitional folded paths. Therefore this thesis gives all the relevant information about the folding process in Villin molecule especially the secondary structure and gives an insight about the structural analysis in the intermediate regions of protein folding.

Keywords: Coarse grained molecule, molecular dynamics, villin molecule, GROMACS, STRIDE, contact map and Ramachandran plot.

INTRODUCTION

A description of the origin of interest in and the development of molecular dynamics simulations of biomolecules is presented with a historical overview including the role of interactions with Shneior Lifson and his group in Israel [1] and Modeling and studying proteins with molecular dynamics is explained in [2].Protein folding is the process by which protein molecules folds into its characteristic highly structured three dimensional functional shape or conformation. The amino acid sequence of a protein determines its structure that in turn determines its mechanism of action. [3] Shows the simplified representation of protein conformations for rapid simulation of protein folding.

and the existence of a theoretical framework based on the global properties of the energy landscape [6,7] have, in recent years, allowed a very fruitful collaboration between theory and experiment in the study of folding[4]shows the studies on protein folding, unfolding and fluctuations by computer simulation. A three-dimensional lattice model of lysozyme. [5] Shows ENM is used to compute the principal modes along a full-atom MD trajectory and amplify them to better explore the phase space. The method is applied to T4 lysozyme and to the refolding of an S-peptide analog.The easiest way to analyze temperature dependent protein folding is through computational methods. There is three ways by which modeling can be done computationally: Threading, Homology Modeling and Ab initio modeling. Molecular dynamics (MD) is an important tool for studying protein folding and dynamics in silico. Villin is a 92.5 kDa tissue- specific actin-binding protein that contains multiple gelsolin-like domains capped by a small headpiece which is the commonly studied and easiest way to study protein. Fig 1 shows the protein folding transition stage.

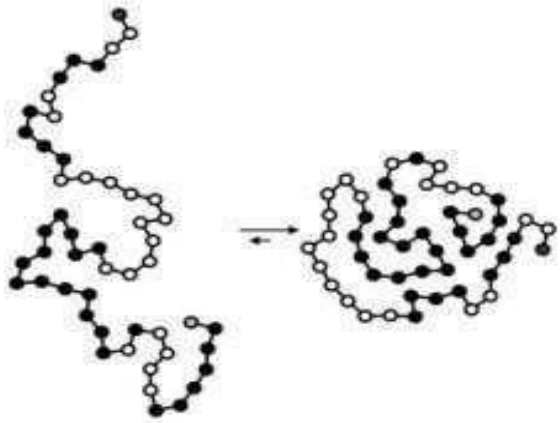

Fig 1) Protein folding transition stage

Coarse grained systems consists of fewer, larger subcomponents compared to the fine grained components. GROMACS is a molecular dynamic package that is used for the design of simulation of protein.

OBJECTIVES

OBJECTIVE 1:
Analyse and interpret FORTRAN codes of a structure based bio molecular wed stool simulator.

OBJECTIVE 2:
Simulate the effect of temperature on protein folding of villin molecule, using coarse models and molecular dynamics software-GROMACS

OBJECTIVE 3:
Analyse the 30 simulations of folding trajectories using several different structural descriptors to obtain a coherent view on folding mechanism.

OBJECTIVE 4:
Find the secondary structure of all transitional parts where the temperature where transition from folding to unfolding takes place.

SIMULATION SOFTWARE: STRIDE:

STRIDE is a program to recognize secondary structural elements in proteins from their atomic coordinates that performs same task as DSSP and utilizes the hydrogen bonds and maintain dihedral angles.

PYMOL:

PYMOL is an molecular visualization system that can produce high quality 3D images of small molecules such as

proteins. It can act as a graphical external MD program by displaying and animating a molecule undergoing simulation on remote computer.

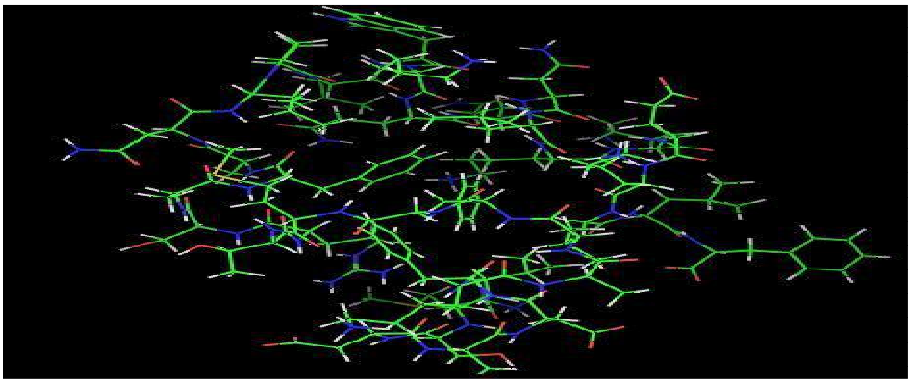

Fig 2) Stick Model Of Villin As Seen By Pymol

DSSP:
DSSP algorithm is a standard structure for assigning secondary structure to an amino acid of a protein, given the atomic resolution coordinates of a protein.

Q VALUES -NATIVE CONTACTS
In protein folding, native contact is the contact between the side chains of two amino acids that are not neighboring in the amino acid sequence but are spatially closed in the protein's native state tertiary structure.

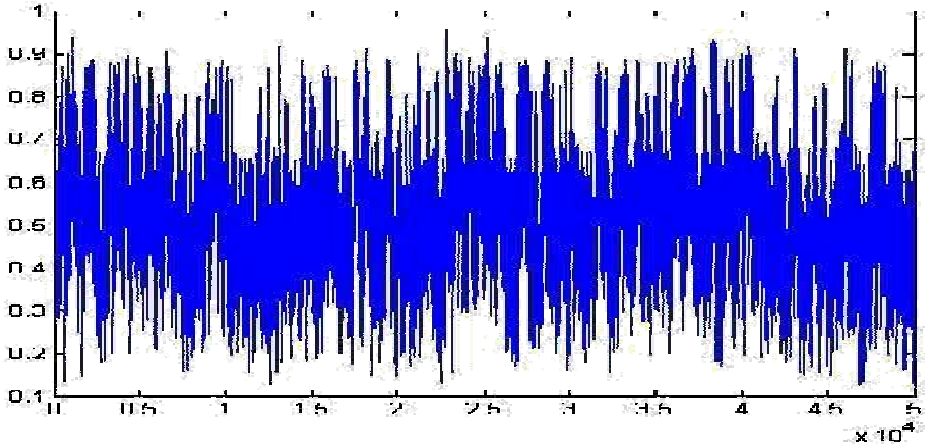

Fig 3) Q Plots

Q-PLOTS helps us to analyze the change in native contacts with respect to temperature. As we increase the temperature we can zigzag transition of protein folding at certain time intervals.

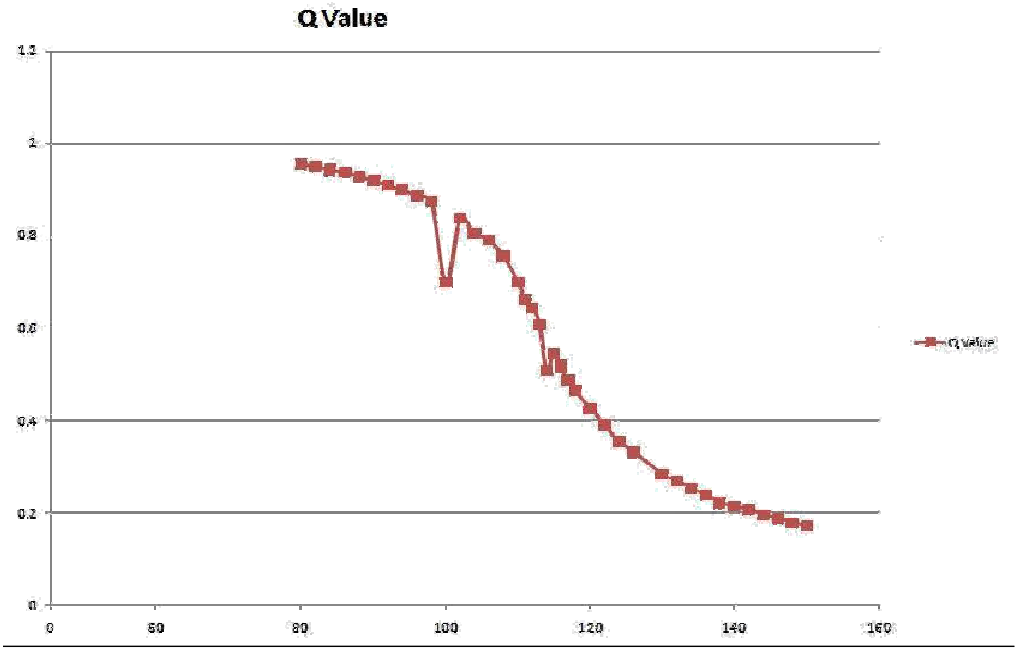

Fig 4) Mean vs temperature graph shows that inflextion point is located at 116rk which is used to find the Q values and plat the graph.

RMSD VALUES: The root-mean-square deviation (RMSD) is the measure of the average distance between the atoms of superimposed proteins.

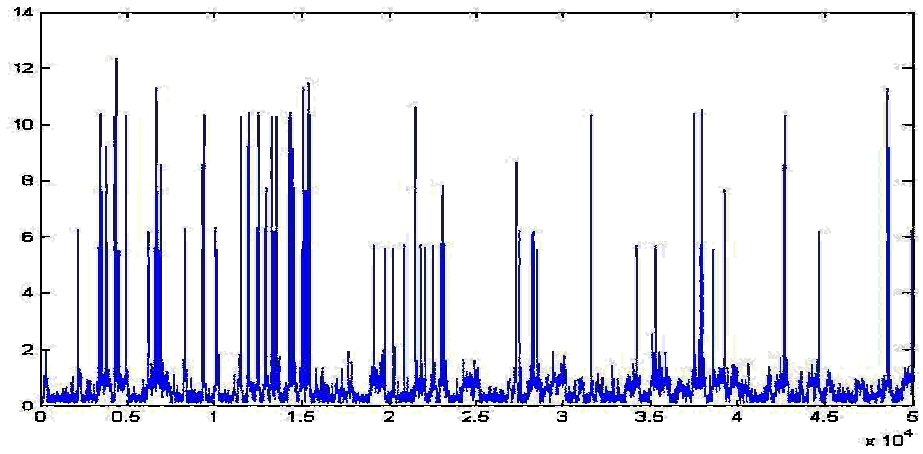

Fig 5) RMSD PLOTS are used to analyse the RMSD values with respect to increase in temperature that implies that the protein is unfolding

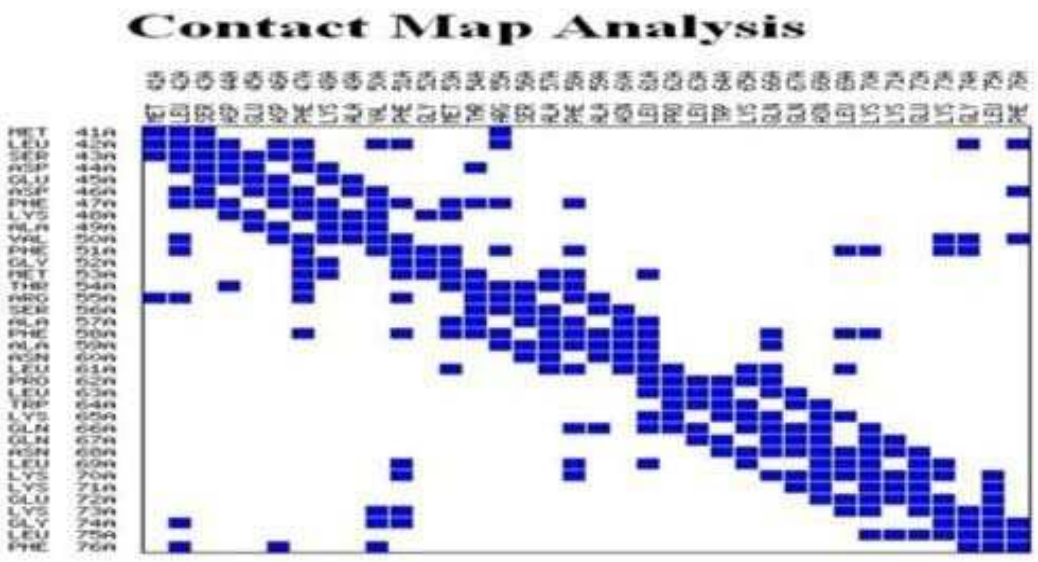

Fig5) Contact map analysis

A protein contact map represents the distance between all possible amino acid residue pairs of a three-dimensional protein structure using a binary two-dimensional matrix. For two residues and, the element of the matrix is 1 if the two residues are closer than a predetermined threshold, and 0 otherwise. Contact maps provide a more reduced representation of a protein structure than its full 3D atomic coordinates. The advantage is that contact maps are invariant to rotations and translations. This plot shows the distance between all amino acid residue pairs of Villin molecule which are 36 in number. This contact map changed over time, as the protein unfolded, therefore the distances of some of the amino acid pairs have changed.

RAMACHANDRAN PLOT:

A Ramachandran plot is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure.

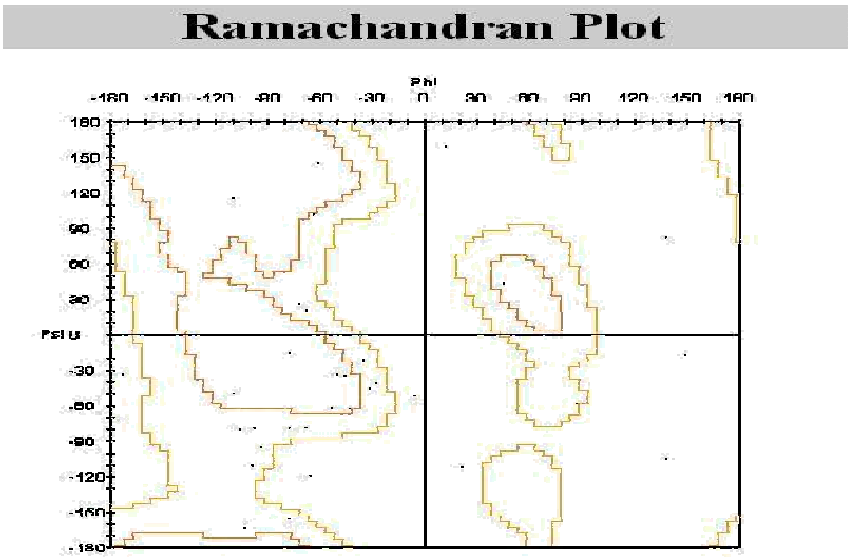

Fig 6) Ramachandran plot

DSSP STRUCTURE:

The main objective of this research work is to find out the secondary structure of the Villin molecule while it is unfolding. That is, to find the transitional change in the secondary structures of each residue.

The steps used for this were a) To find the temperature where the protein unfolding happens at equilibrium. b) Then, to find the time steps where the transition happens at that particular temperature. c) Then, to find the secondary structure of DSSP and STRIDE.

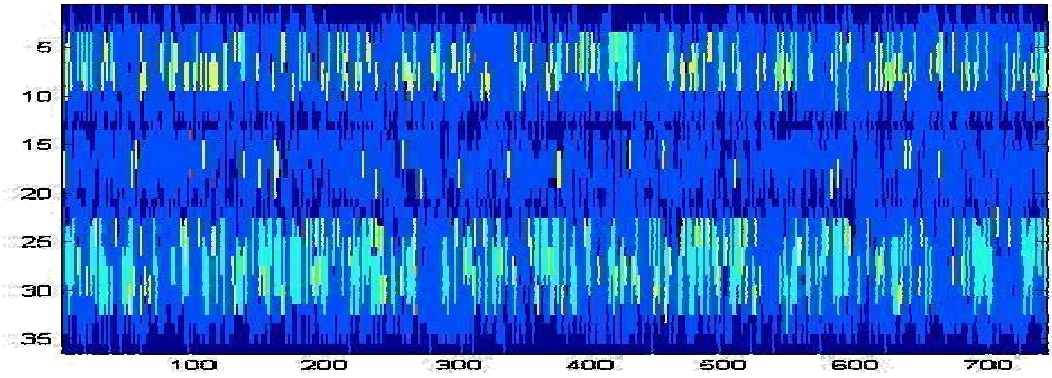

Fig 7) Secondary structure of villin molecule

The below table represents the characteristics of the villin molecule

Table 1) Characteristics of villin molecule

Symbol	Number Represented	Explanation	Colour
C	1	Coil – Random coils and loops	Dark Blue
T	2	Hydrogen Bonded Turn	Light Blue
H	3	Alpha Helix	Sky Blue
G	4	3/10 Helix	Green
B	5	Residue in isolated beta-bridge	Orange
I	6	Pi Helix	Yellow

CONCLUSION

From the study of the Villin molecule using Molecular Dynamics simulations, we had an interesting set of observations.

a) Parallel processing reduces the duration of the simulation by a significant factor. Thus, molecular dynamics must be done by a powerful computer.

b) Q values (Native Contacts) and RMSD values change with change in temperature.

c) A simulation for 50,000 time steps gives much more reliable information than one with fewer steps.

d) Peripheral softwares are very useful as they help in analyzing with simple commands. The secondary structure can be further used to study more properties of Villin molecules.

REFERENCES

[1] Karplus M: *Biopolymers* **2003**, 68:350-358.
[2] Schleif R: *Methods Enzymol* **2004**, 383:28-47.
[3] Levitt M. *J Mol Biol* **1976**, 104:59-107.
[4] Ueeda Y, Taketomi H, GoN: *Biopolymers* **1978**, 17:1531-1548.
[5] Zhang Z, Shi Y, Liu H: *Biophys J* **2003**, 84:3583-3593. ENM is used to compute the principal modes along a full-atom MD trajectory and amplify them to better explore the phase space. The method is applied to T4 lysozyme

and to the refolding of an S-peptide analog.
[6] Bryngelson JD, Wolynes PG: *Proc Natl Acad Sci* USA **1987**, 84:7524-7528.
[7] Leopold PE, Montal M, Onuchic JN: *Proc Natl Acad Sci USA* **1992**, 89:8721-8725.