Protein Contact Network analysis and atomistic Molecular Dynamics simulation study of various thermostable mutant proteins

MILIND SANJAY KALE

A dissertation submitted for the partial fulfilment of the BS-MS dual Degree in Chemical Sciences.

Indian Institute of Science Education and Research Mohali

April 2020

Certificate of Examination

This is to certify that the dissertation titled **"Protein Contact Network analysis and Atomistic Molecular Dynamics Simulation study of various thermostable proteins."** submitted by **Milind Sanjay Kale** (Reg. No. MS15197) for the fulfilment of **BS-MS dual degree programme** of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Shashi Bhushan Pandit Dr Sabyasachi Rakshit Dr. Monika Sharma
(Supervisor)

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Monika Sharma at the Indian Institute of Science Education and Research Mohali.
This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Milind Sanjay Kale
(Candidate)

Dated: 4-5-2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Monika Sharma
(Supervisor)

Acknowledgements

Foremost, I would like to express my gratitude to my supervisor Dr. Monika Sharma for continuous support of my MS thesis and research, for her patience, motivation, enthusiasm and immense knowledge.

Besides my supervisor I would like to thank the rest of my thesis committee: Dr. Shashi Bhushan Pandit and Dr. Sabyasachi Rakshit for their insightful information.

Also, I would like to thank my friends Mr. Asish Swain and Mr. Nilesh Deokate for their support throughout my thesis and Mr. Jagdish Prakash Hajra for his insightful knowledge in MD simulations.

Last but not least, I would like to thank my family and IISER Mohali for supporting me throughout my thesis.

List of Table

List of Figures

Fig-3.4.3) RMSD vs Time plot of T4 lysozyme (Black-4S0W, Green-2OE4, Red-1TLA)

Fig-3.4.4) Radius of gyration vs time plot of T4 lysozyme (Black-4S0W, Green-2OE4, Red-1TLA)

Fig-3.4.5)-Fig-3.4.10) Hydrogen-bond analysis of T4 lysozyme odd number for intra-peptide H-bond and even number for peptide water H-bond.

Fig-3.5.1) community cluster analysis of GH10 xylanase

Fig-3.5.2) RMSD vs Time plot of GH10 Xylanase (Black-4QCE, Green-4QDM, Red-4QCF)

Fig-3.5.3) Radius of gyration vs time plot of GH10 xylanase (Black-4QCE, Green-4QDM, Red-4QCF)

Abbreviations

1. Wild Type- WT

2. Protein Contact Network- PCN

3. Residue Interaction Network- RIN

4. X-ray Diffraction- XRD

5. Molecular Dynamics Simulations- MDS

6. Root-Mean-Square-Deviations- RMSD

7. Radius of gyration- R_g

8. Protein Data Bank- PDB

9. Betweenness Centrality- B.C.

10. Clustering Coefficient- C.C.

11. yeast Cytosine Deaminase- yCD

12. Hydrogen-Bonds – H-B

13. Adenylate Kinase- ADK

14. 3-Isopropylmalate Dehydrogenase- 3ID

15. Community- C (used in appendix)

Abstract

Thermal stability of protein means to be resistant to changes in protein structure due to applied heat. in this thesis, our interest has been to study, those proteins and their mutants, having indistinguishable ordered crystallographic structures, but exhibiting large changes in functions mostly in thermostability. Here in our analysis we take five different wild type proteins and their thermostable mutants and we analyse residue specific network parameters of each one using Residue Interaction Network(RIN) to know role of such parameters in determining it's thermostability thereafter we have done Atomistic Molecular Dynamics Simulation study of three selected proteins to know RMSD, Radius of gyration and Hydrogen-Bond analysis which are factors that can give us comparative measure to know how small changes in protein can affect its function majorly its melting temperature.

Contents

Table of Contents

Chapter 1

Introduction

1.1. What is protein thermal stability

Proteins are large macromolecules, consisting of one or more long chains of amino acid residues. Proteins differ from one another primarily in their sequence of amino acids, which is governed by the nucleotide sequence of their genes, and which usually results in protein folding into a specific quaternary structure that determines its activity. Thermal stability of protein means to be resistant to changes in protein structure due to applied heat. most life forms in the earth survive at temperature less than 50^0 C, Within these organisms are macromolecules (proteins and nucleic acids) which form the three-dimensional structures essential to their enzymatic activity.[1]certain thermophilic organism can withstand temperature above 50^0C and have corresponding adaptations to preserve protein function at these temperatures[2]

in this thesis, our interest has been to study, those proteins and their mutants, having indistinguishable ordered crystallographic structures, but exhibiting large changes in functions mostly in thermostability.

There are several cases that are coming to light, where significant changes in function have been observed in some proteins due to mutations, which have negligible changes in the overall structures.[3]in the present work we have attempted to utilize this approach to study small conformational changes occurring in the protein due to mutations, by analysing the contact patterns and their positional information in all the mutants and the wild type(WT), with respect to their structural stability and functionality at higher temperatures of different proteins occurring in different organisms. earlier studies on thermostability have invoked the role of many structural factors that arise due to mutations and contribute to structural rigidity, such as higher number of hydrogen-bonds and salt-bridges, secondary structure stabilization, disulphide linkages, higher polar surface area, more number of Proline

residues, shortening and stabilization of loops, etc[4–8].

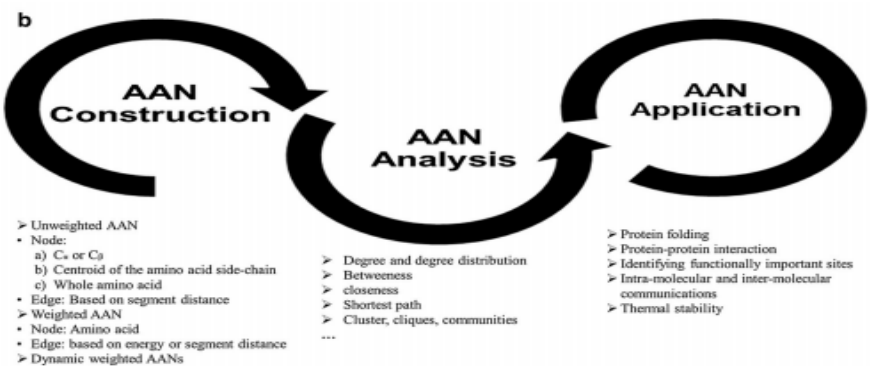
1.2. Protein Contact Network or Residue Interaction Network

The three-dimensional crystallographic structure of the protein can be represented using a network description – the "Protein Contact Networks" (PCNs) also known as Residue Interaction Network(RIN) – whose nodes are the amino acids (residues) and the chemical interactions between closely held residues are the links. Such complex network based approaches have been used for studying the structure and function of proteins.[1] RINs decrease the visual complexity of 3D protein structures and allow the user to focus on individual residues and their molecular interactions.[9] along with RIN analysis here we have used atomistic Molecular Dynamics Simulation(MDS) approach to know more about thermostability of mutants compared to wild type protein using RMSD Vs Time plot of mutants and wild type.
RINs can be applied to study residue interactions in a number of relevant application scenarios, for instance, with regard to protein dynamics and engineering, structure–function relationships, protein and ligand binding, and the impact of amino acid substitutions.[9]

Usually, while studying the effect of point mutations on protein functions, only the areas which are in the vicinity of mutation are analysed in detail. However, studies have shown that small conformational changes occurring in one region of the protein can have an effect in a region quite far from the point of perturbation without affecting the functionality[10].
Analysis of residue-specific network parameters offers several important clues towards the residues that may have a role to play in governing structural stability under thermal stress in the mutant proteins[3].
One important feature that is revealed through this analysis is that most changes in network parameters occur at the residues that are not part of the active site of the proteins, indicating the specific role of allosteric sites in these thermostable proteins whose overall conformations do not change due to the mutations[3].
Here in our analysis we take five different wild type proteins and their thermostable mutants and we analyse residue specific network parameters of each one to know role of such parameters in determining it's thermostability.
Generally, the global network parameters remain same in wild type and its mutant proteins
To know small conformational changes at the site of mutation, the variation in local, residue-specific network parameters were studied in all selected proteins. We show the detailed

analysis of the small conformational changes, occurring through-out the proteins, by analysing the local network parameters, positional information, contact patterns, and the changes in the community membership, with respect to their structural stability and functionality at higher temperatures. After that we have done MD simulation of four proteins including wild types and mutants to show differences in the RMSD value, radius of gyration and hydrogen bonds (both intramolecular and intermolecular) values which were comparative parameters to know more about thermostability.

Fig-1.1) overall workflow on Amino Acid Network or Residue Interactions Network



Chapter 2

Methods and computational details

2.1. System studied

In thesis, I have taken proteins from various organisms covering prokaryotic, eukaryotic also consisting of viral protein. All studied proteins are listed below with pdb id, organisms I which they are found, structure resolutions and by what method structure is obtained.

Table-2.1.1) System studied

Protein	PDB id	Organism/mutations	Resolution	Structure obtained with
1.Yeast cytosine deaminase		Saccharomyces cerevisiae		
	a)1P6O(WT)		1.14 Å	XRD
	b)1YSD	A23L/I140L	1.9 Å	XRD
	c)1YSB	A23L/V108I/I140L	1.7 Å	XRD
2.Adenylate kinase		Bacillus subtilis		
	a)1P3J(WT)		1.9 Å	XRD
	b)2OSB	Q16L/Q199R	1.8 Å	XRD
	c)2QAJ	Q199R/G213E	1.8 Å	XRD
3.3-Isoproylmalate dehydrogenase		Thermus thermophilus		
	a)1XAA(WT)		2.1 Å	XRD
	b)1G2U	A172V	2.1 Å	XRD
	c)1GC8	A172F	2.5 Å	XRD

	d)1GC9	A172G	2.3 Å	XRD
4.T4 lysozyme		Enterobacteria phage T4		
	a)4S0W		2.12 Å	XRD
	b)1TLA	S117F	2.0 Å	XRD
	c)2OE4	High pressure pseudo wild type	2.1 Å	XRD
5.GH10 xylanase		Bacillus sp. NG-27		
	a)4QCE		2.32 Å	XRD
	b)4QDM	V1L	1.96 Å	XRD
	c)4QCF	V1A	2.26 Å	XRD

2.2. Construction of Protein Contact Network

The crystallographic coordinates of all proteins are retracted from Protein Data Bank(PDB)[11]. PCN were controlled by using c-alpha atom in each residue in protein as a node and interaction within spatial proximity between any two residues as a link or edge. Here we use a coarse-grained approach[3,5,6] to construct the network by considering the spatial positions of only the protein backbone atoms (i.e. C-α atoms) of each residue. A pair-wise Euclidean distance matrix is computed between the C- α atoms of all residues. The distance matrix is converted into an Adjacency matrix, A using the following rule

A_{ij}=1; if D_{ij}<=5 Å; else; A_{ij}=0

where, D_{ij} is the Euclidian distance between C-α atoms of the i^{th} and j^{th} residues.For the RIN generation we use Cytoscape.[13] an established open-source software platform for the analysis and visualization of biological networks where RINalyser plugin is used for biological network generation. In General, this platform can be extended by plugins for any specific work. For instance, the Cytoscape plugin structureViz supports the structural analysis of protein–protein interaction networks. Like structureViz , RINalyser links cytoscape with UCSF Chimera[14] for 3D protein structure visualization also, Cluster Maker is another plugin available for cytoscape to find network modules and clusters.Our dataset from PDB comprising of five different wild type proteins and their thermostable mutants which are shown in listed tables for each protein.

Here all mutations were obtained in a manner that the previous mutation is not disturbed when there is incorporation of new mutation.

2.3. Network parameters

Brief introduction to network parameters discussed in thesis are given below

Network Density-Network Density is a measure of the proportion of possible binds which are among the members of a network[15].

Degree-The degree of node (i) in a network is the total number of edges pass through node(i) due to other nodes. greater the degree suggest residue is more interacting with other residues in protein.

Betweenness Centrality (B.C.) - B.C. of a node is number of shortest paths pass through that node. which measures the extent to which a vertex lies on paths between other vertices. B.C. is related to a networks connectivity so much as high betweenness vertices have the potential to disconnect graph if removed.[16]

It is relative quantity, betweenness centrality is directly proportional to the average rate at which traffic passes through any particular vertex. Higher betweenness centrality of a node means it has more control over the flow of information.

Clustering Coefficient (C.C.)-The clustering coefficient measures the average probability that two neighbours of a vertex are themselves neighbours.it also some time refer as density of triangles in a network. One can define coefficients similar to clustering coefficient to measure the density of different motifs.

Local Clustering Coefficient-Local clustering coefficient for a vertex is fraction of pairs of neighbours of vertex i that are themselves neighbours. we find that of average vertices of higher degree tend to have lower local clustering coefficients.

2.4. RIN of wild type protein and its mutants

All figures regarding RIN are on the appendix page. colour coding done there was according to community colour i.e. network graph and 3D structure coloured according to Glay cluster mapping.

2.5. Community Structure Analysis

To divide many network nodes into groups we have to find communities and modules in the network such that there is higher density of edges within the group than between them. Community structure analysis was performed for all the six Proteins using the Glay clustering in ClusterMaker plugin available for cytoscape which uses fast greedy algorithm.[17]
GLay clusters of wild type and mutant protein helps to comparatively analyse position of mutated residue. i.e. after single point mutation residue can change its previous cluster and goes to more compact clusters.

2.6. Atomistic Molecular Dynamics Simulation-

All MD simulation were carried out using GROMACS version 2019.3.[18] in conjunction with all-atom OPLS force field[19].The starting structures of 4 WT and its 2 thermostable mutants each were taken from the PDB entries 1P6O(WT), 1YSD, 1YSB, 1P3J(WT), 2OSB, 2QAJ, 4S0W(WT), 1TLA, 2OE4, 4QCE(WT), 4QDM, 4QCF.after adding the hydrogens, the protein structure were solvated with SPC water molecules in a cubic solvent box with a minimal distance between protein surface and box of 1.0 nm. All the crystallographic/structural water were removed. All the systems were minimized using the steepest descent algorithm until it converged to a force tolerance of 1000 KJ mol^{-1} nm^{-1} .After minimization each system was equilibrated to desired temperature through a stepwise heating protocol in the NVT ensemble using velocity-rescale thermostat followed by 100ps equilibration in the NPT ensemble using parinello-rahman barostat with position restrains on the protein molecule finally a production simulation was performed for each system at melting temperature of wild type for 10ns under periodic boundary condition(PBC) without any restraints on the protein. Nonbonded interactions were calculated using the particle-mesh Ewald method[20] with cut-off of 1.0 nm. The LINCS algorithm[21] was used to constrain all bonds including hydrogen atoms during simulation.

2.7. Parameters of MD Simulation trajectories-

Analysis such as root-mean-square deviation(RMSD), radius of gyration, hydrogen bonds were performed using the tools within the GROMACS simulation package. For hydrogen bond calculations, a donor-acceptor cut-off distance of 0.35nm and acceptor-donor-hydrogen bond angle cut-off of 30^0 were considered[22].The visual analysis of structures and preparation of figures was carried out using Pymol, VMD[23], Xmgrace[24].

Root-mean-square Deviation of atomic positions(RMSD)-RMSD is the measure of the average distance between the backbone atoms of superimposed protein. Lower RMSD value suggest that protein structure is stable i.e. smaller the deviations more stable the structure[25].

Radius of gyration(R_g)-radius of gyration of protein is used for measuring compactness of protein, if a protein is stably folded it will maintain stable value of R_g .

Hydrogen-Bond Analysis –in thesis we have calculated hydrogen-bonds formed intra-peptide and peptide-water throughout the simulation time which is 10000ps which is equal to 10ns.

Chapter-3

Results

3.1. RIN and MDS data of Yeast cytosine deaminase(yCD) (PDB id-1P6O)

Classification**:** HYDROLASE; Residue count-161

Organism(s): Saccharomyces cerevisiae (strain ATCC 204508 / S288c)

It is protein present in the nucleus of *saccharomyces cerevisiae* for the function of converting cytosine to uracil or 5-methylcytosine to thymine by deaminating carbon 4 of cytosine.[26] yCD was chosen in the thesis because its high-resolution crystal structure is available, its catalytic mechanism is well studied, and it has potential use in antitumor suicide gene applications[27–30]. In the study performed by authors they have used computational redesign of series of point mutations in the enzyme core which leads to thermostabilization of enzyme without losing functionality[31].Using Site Directed Mutagenesis they made double mutant (PDB id-1YSD) consisting of A23L & I140L mutations and Triple mutant (PDB id-1YSB) consisting of A23L, I140L & V108I mutations. The isolated single mutations A23L, I140L, and V108I each slightly thermostabilized the enzyme by ~2°C in the T_m[31]. However, simultaneous incorporation of all three mutations increased T_m 10°C higher than that of the wild type. Here we see that the effects of mutation were synergistic that means T_m is increased compared to their individual effects.

PCN Analysis of yCD-

Increase in total number of edges in mutants PCN's can change Network Parameters. The four network parameters studied for all PCN's are Degree, Betweenness centrality, Shortest Path Length and clustering coefficients.

For Global-scale the average network parameters were computed for each PCN, the ranges of such parameter value for 3PCN's are-

Values here are wild type and most thermostable mutant structure.

1.Network density- 0.050-0.047

2.Betweenness Centrality- 0.0175-0.0180

3.Clustering Coefficient- 0.452-0.431

4.Shortest Path Length- 3.831-3.877

The very low variation in average network parameters in all PCN's shows low cross-structural RMSD of Proteins thereby having same 3-D structures.

For Finding any Local Changes in Network parameters were computed at individual node level for all 3 PCN's. Then we have done the comparative analysis of residue specific network parameters which was presented in tabulated manner.

Table-3.1.1) residue specific network parameters of yCD

colour code-Red for helix, Blue for sheet, yellow for loop

in g-lay cluster analysis numbers inside the bracket represents the total number of residues in the cluster

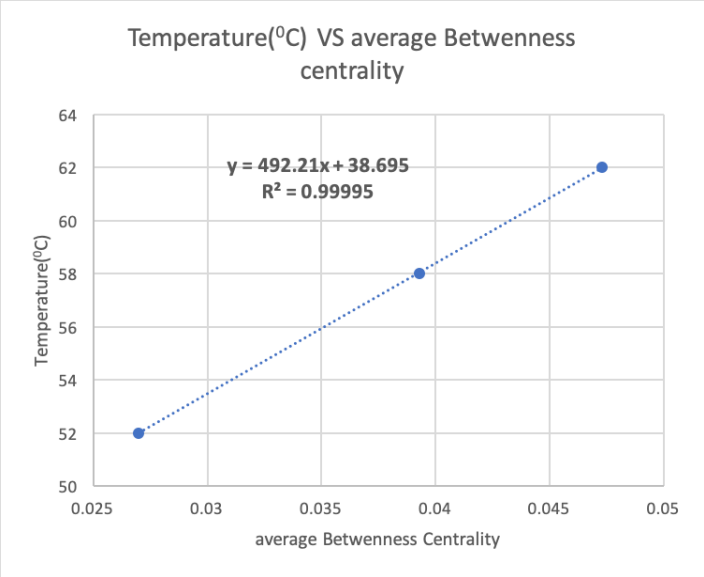
Sr.no.	PDB id	Residue	Degree	B.C.	C.C.	G-lay cluster
1	1P6O(wt)	ALA23	14	0.009	0.38	4(43)
2	1YSD	LEU23	16	0.0099	0.5	1(50)
3	1YSB	LEU23	16	0.0078	0.5	3(48)
1	1P6O(wt)	VAL108	21	0.059	0.2545	4(43)
2	1YSB	ILE108	23	0.1043	0.2527	2(28)
1	1P6O(wt)	ILE140	16	0.014	0.4	3(40)
2	1YSD	LEU140	15	0.02	0.3333	1(50)
3	1YSB	LEU140	17	0.03	0.3272	3(48)

.

Here we see that average betweenness centrality and clustering coefficient of residues increases with melting temperature.

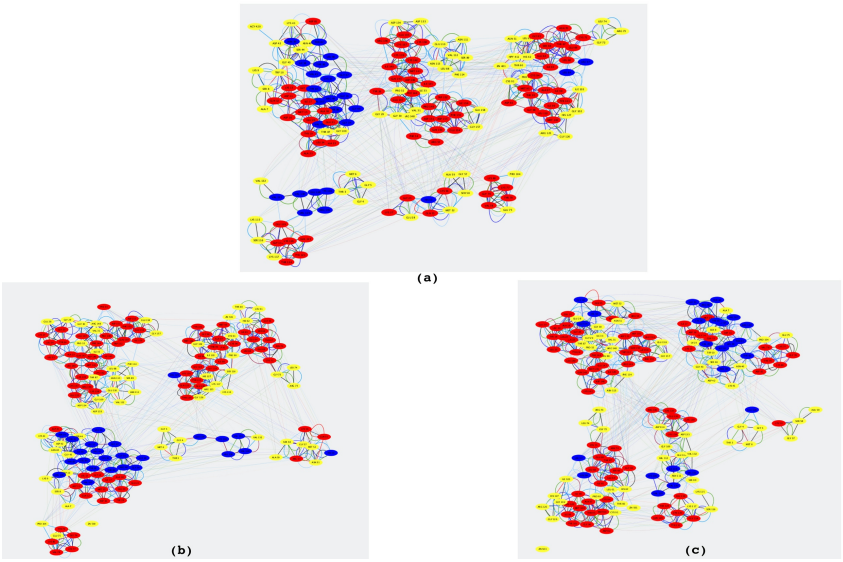
Then we observe that average betweenness centrality is directly proportional to melting temperature of protein. means it has more contribution than other residue specific parameters.

Fig-3.1.1) temperature(y-axis) vs average betweenness centrality(x-axis) graph



community cluster analysis of yCD-

Fig-3.1.2) community cluster analysis of yCD



a)1P6O (wild type) has formed 7 clusters.

T_m= 52^{0}C

b)1YSD (double mutant) has formed 6 clusters.

T_m= 58^{0}C

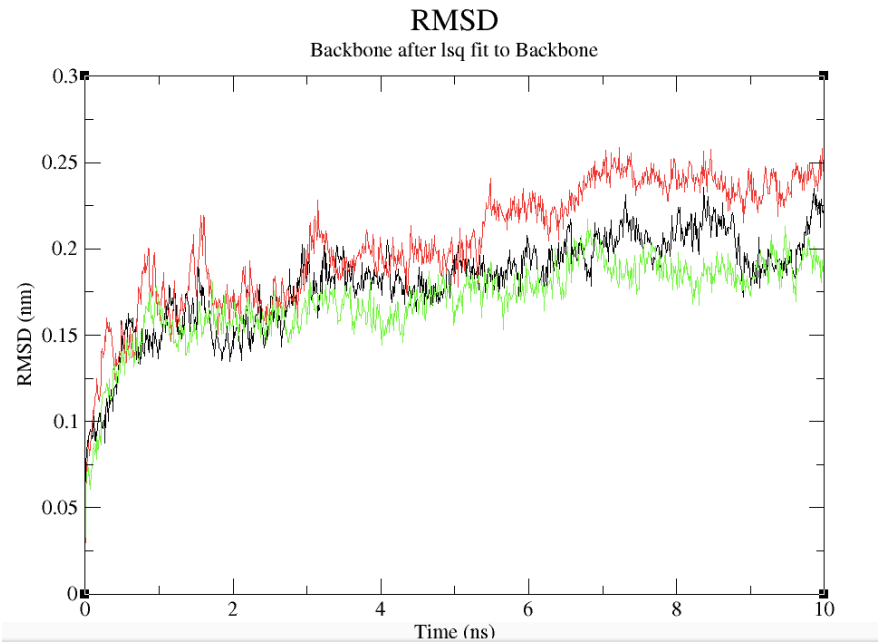
c)1YSB (triple mutant) has formed 6 clusters.

T_m= 62^{0}C

In G-lay cluster maker wild type, double mutant, triple mutant formed 7,7,6 communities respectively but in case of thermostable mutant proteins there is 3-4 larger communities (having residue count greater than 20) formed so it inferred that compactness of communities played major role in thermostabilization.

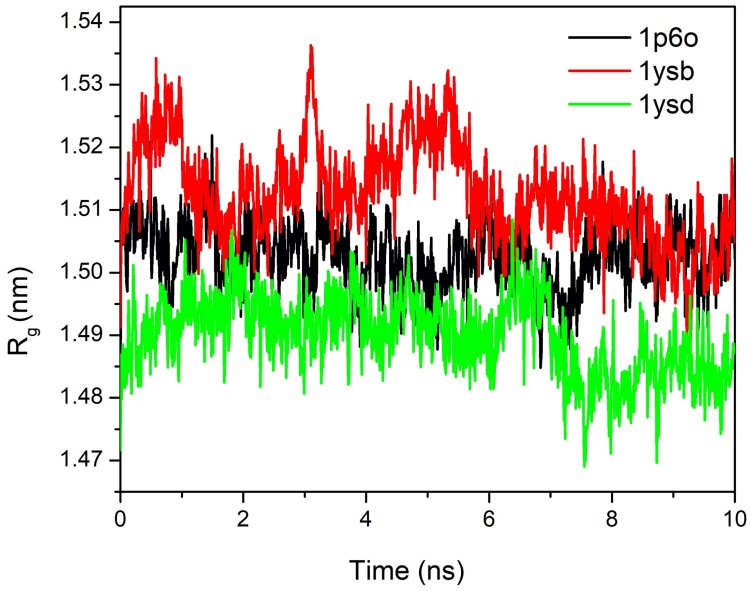
Molecular Dynamics Simulation analysis-

Fig-3.1.3) RMSD vs Time plot of yCD (Black-1P6O, Green-1YSD, Red-1YSB)



The RMSD simulation of 10ns at 300K showed that double and triple mutant showed stability in graph after 8ns but wild type showed slightly fluctuating behaviour after 8ns.

Fig-3.1.4) Radius of gyration vs time plot of yCD (Black-1P6O, Green-1YSD, Red-1YSB)

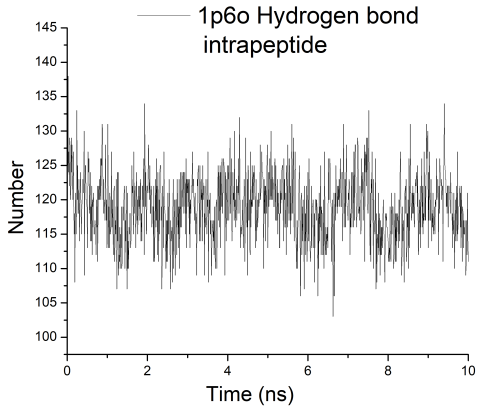


Here also we saw that R_g value of wild type is slightly increasing after 8ns, double mutant and triple mutant showed decreasing graph of R_g after 8ns.

Low value of R_g corresponds to tighter packing of structure.

H-B intra peptide

Fig-3.1.5)1P6O(wt)



H-B peptide-water

Fig-3.1.6)1P6O(wt)

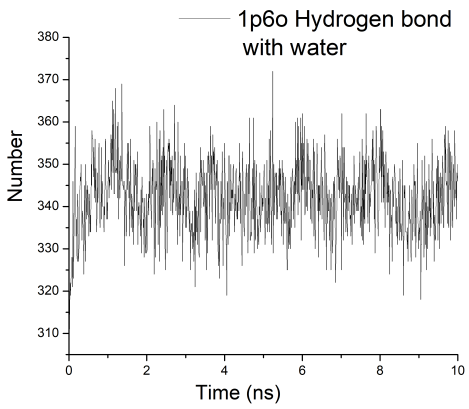


Fig-3.1.7)1YSD (double mutant)

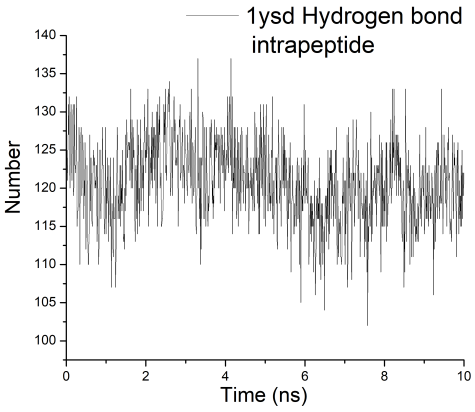


Fig-3.1.8)1YSD (double mutant)

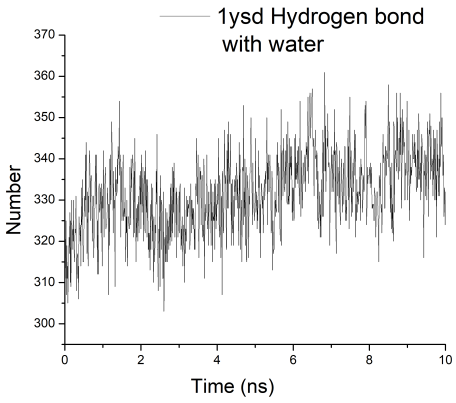


Fig-3.1.9)1YSB(triple mutant)

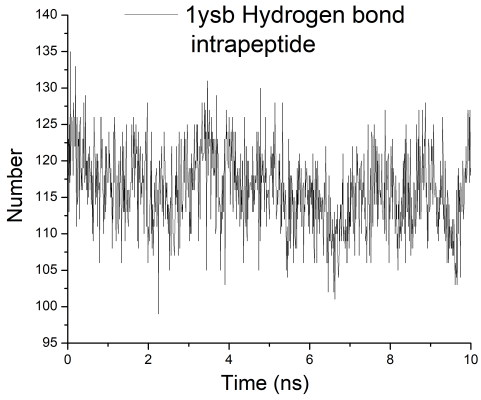
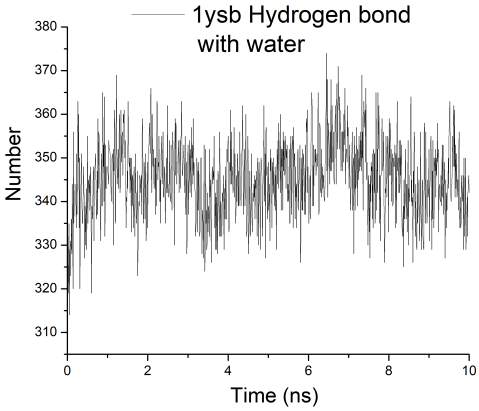


Fig-3.1.10)1YSB(triple mutant)



In H-B analysis there is increasing trend in intra-peptide hydrogen bonds and decreasing trend in peptide-water hydrogen bonds in thermostable mutant protein compared to wild type protein.

3.2.RIN and MDS data of Adenylate kinase(ADK) (PDB id-1P3J)

Classification**:** TRANSFERASE. Residue Count-217

Organism(s): Bacillus subtilis (strain 168)

ADK is phosphotransferase enzyme that catalyses the interconversion of adenine nucleotides namely AMP, ADP and ATP. ADK is important for cellular level homeostasis.

Wild type of ADK is 1P3J,Authors have generated total 5 different thermostable mutants but only 2 of them have better crystallographic data in the RCSB PDB Which was 2OSB(Q16L/Q199R) and 2QAJ(Q199R/G213E) Q199R mutation was necessary for defining the structural contexts that facilitated subsequent mutations[32].G213 decreases protein stability but when incorporated with Q199R it shows increase in stability by 3.4^0C which demonstrates that order of mutations plays important role for favouring particular molecular pathway[32].

PCN Analysis of ADK-

For Global-scale the average network parameters were computed for each PCN, The ranges of such parameter value for 3PCN's are-

Values here are of wild type and most thermostable mutant structure.

1.Network density- 0.033-0.033

2.Betweenness Centrality-0.020-0.018

3.Clustering Coefficient-0.412-0.428

4.Shortest Path Length-4.567-4.535

here also we saw less variation in global parameters so local changes in network parameters were computed at individual node level.

here we have done the comparative analysis of residue specific network parameters which was presented in tabulated manner.

Table-3.2.1) residue specific network parameters of ADK

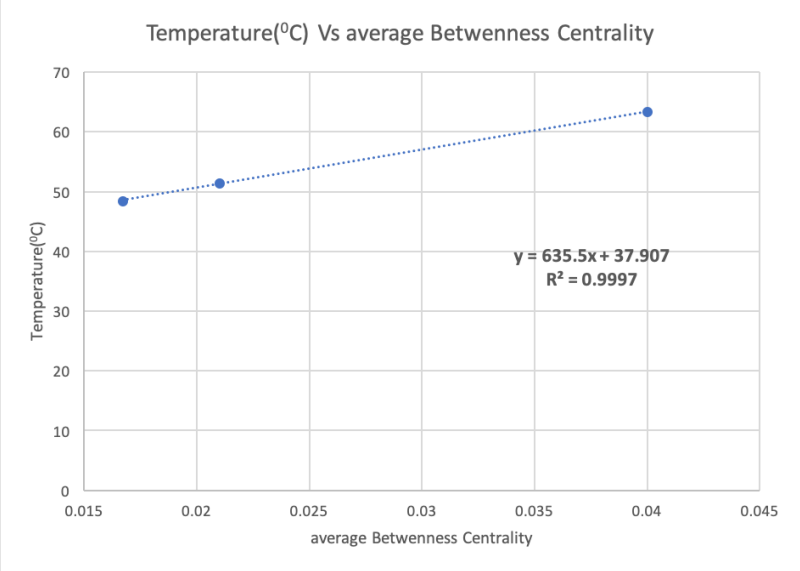
colour code-Red for helix, Blue for sheet, yellow for loop

in g-lay cluster analysis numbers inside the bracket represents the total number of residues in the cluster

Sr.no.	PDB id	Residue	Degree	B.C.	C.C.	G-lay cluster
1	1P3J(wt)	GLN16	19	0.048	0.32	3(63)
2	2OSB	LEU16	17	0.069	0.33	2(66)
1	1P3J(wt)	GLN199	6	0.002	0.33	4(9)
2	2OSB	ARG199	14	0.003	0.43	2(66)
3	2QAJ	ARG199	11	0.002	0.6	3(65)
1	1P3J(wt)	GLY213	7	0.0001	0.66	3(63)
2	2QAJ	GLU213	14	0.0369	0.30	3(65)

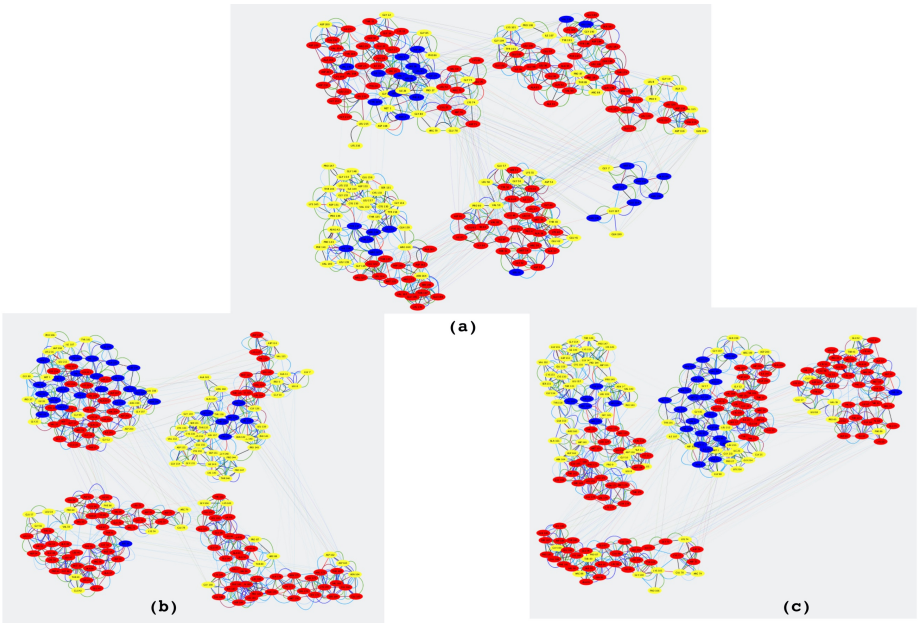
In ADK there is linear correlation between melting temperature and average betweenness centrality.

Fig-3.2.1) temperature(y-axis) vs average betweenness centrality(x-axis) graph



community cluster analysis of ADK-

Fig-3.2.2) community cluster analysis of ADK.



a)1P3J(wt) has 5 clusters.

T_m= 48.4^{0}C

b)2OSB has 4 compact structures.

T_m= 63.3^{0}C

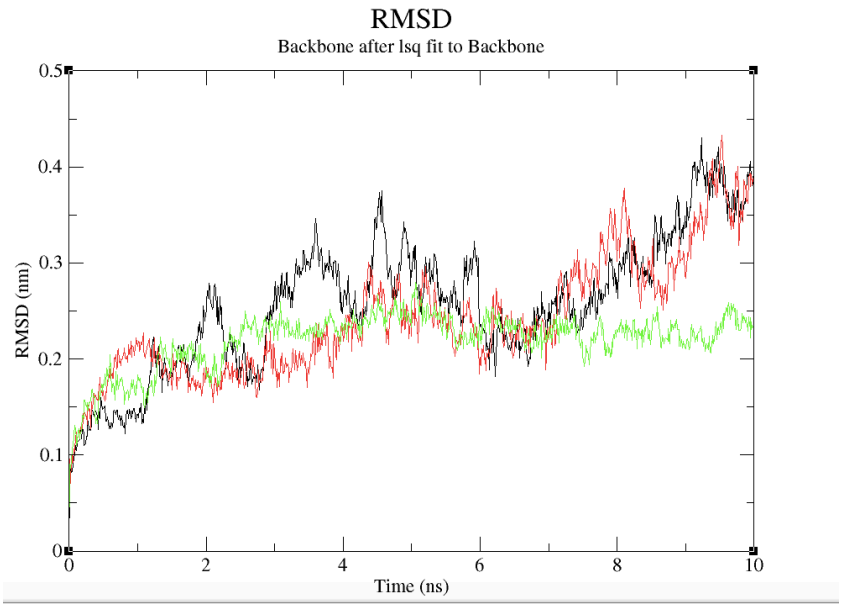
c)2QAJ has 4 compact structures.

T_m= 51.4^{0}C

Here also compactness of communities plays crucial role in the thermostabilization.

Molecular Dynamics Simulation analysis-

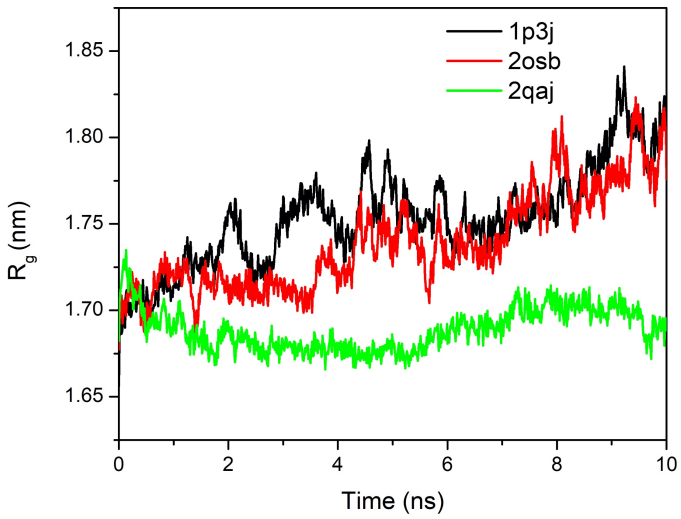
Fig-3.2.3) RMSD vs Time plot of ADK (Black-1P3J, Green-2OSB, Red-2QAJ)



The RMSD simulation of 10ns at 320K showed that mutant structure maintained overall stability throughout simulation while wild type displayed more fluctuations.

Striking result of 2OSB(Green) showed that RMSD shows very less fluctuation compared to both of the structures.

Fig-3.2.4) Radius of gyration vs time plot of ADK (Black-1P3J, Green-2OSB, Red-2QAJ)

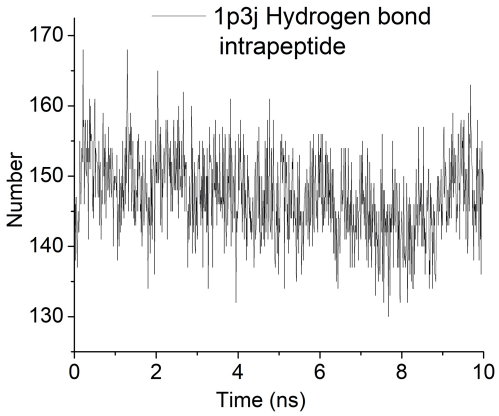


plot of R_g showed that mutant structures shows less R_g values compared to wild type. 2OSB(Green) showed very low R_g so it confirms that it has tighter packing compared to both structures.

Hydrogen-Bond analysis-

H-B intra peptide

Fig-3.2.5)1P3J(WT)



H-B peptide-water

Fig-3.2.6)1P3J(WT)

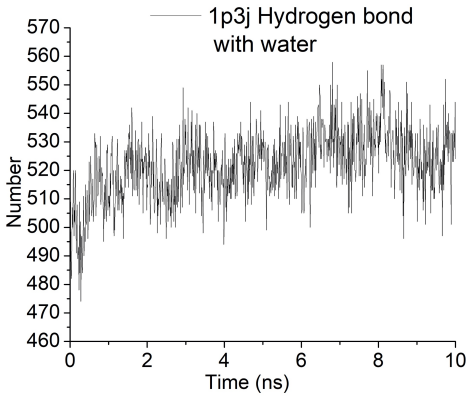


Fig-3.2.7)2QAJ

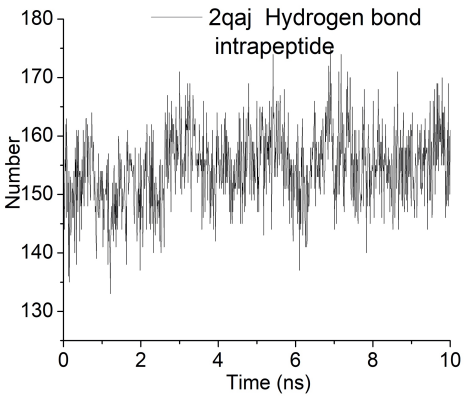


Fig-3.2.8)2QAJ

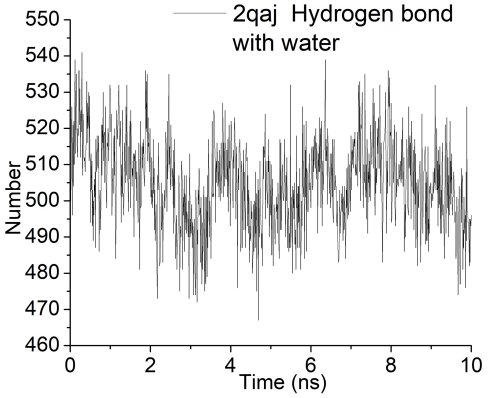


Fig-3.2.9)2OSB

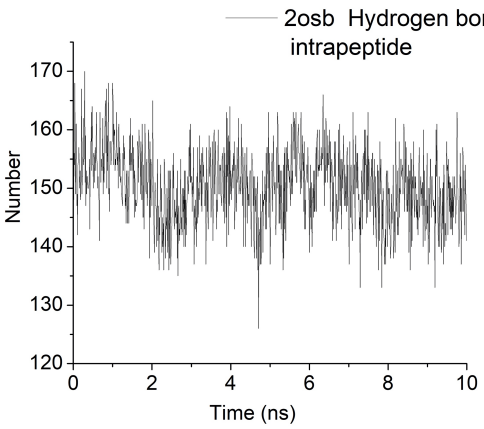
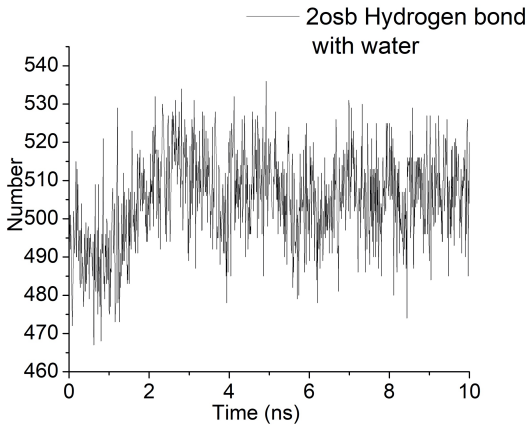


Fig-3.2.10)2OSB



In H-B analysis there is increasing trend in intra-peptide hydrogen bonds and decreasing trend in peptide-water hydrogen bonds in thermostable mutant proteins compared to wild type protein.

3.3. RIN data of 3-Isopropylmalate dehydrogenase(3ID)(PDB ID-1XAA)

Classification**:** OXIDOREDUCTASE**.** Residue count-345

Organism(s)**:** Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579)

3ID is enzyme in leucine biosynthetic pathway that catalyses dehydrogenation and decarboxylation of 3-isopropyl malate.

Here authors have done site directed mutation of single alanine residue located at domain interface of 172 position by valine(1G2U), glycine(1GC9) (short residue) shows very little increase or decrease in thermal stability but in one case they have changed by phenylalanine (1GC8)(long residue) and it shows increase in thermal stability[33].

PCN analysis of 3ID-

For Global-scale the average network parameters were computed for each PCN, the ranges of such parameter value for 3PCN's are-

Values here are of wild type and most thermostable mutant structure.

1.Network density- 0.023-0.021

2.Betweenness Centrality- 0.010-0.012

3.Clustering Coefficient- 0.450-0.426

4.Shortest Path Length- 5.428-5.559

Table-3.3.1) residue specific network parameters of 3ID

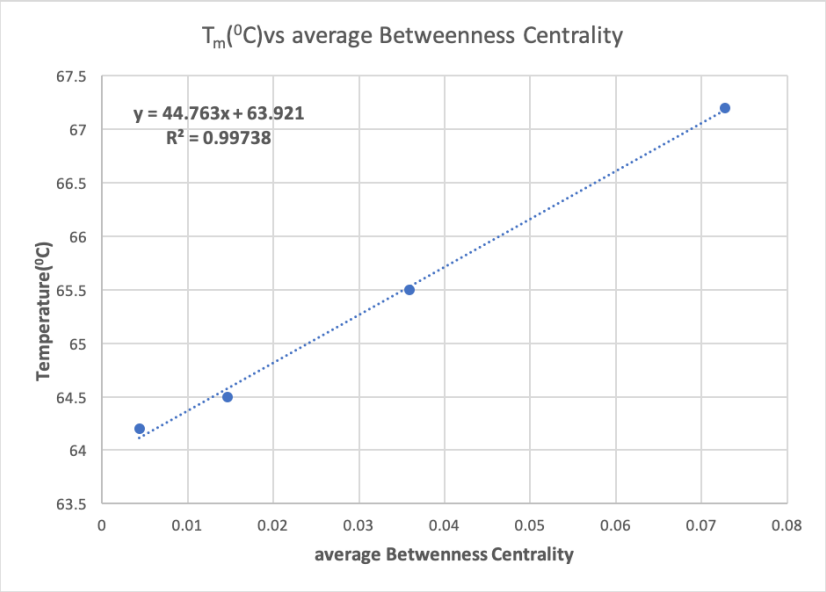
colour code-Red for helix, Blue for sheet, yellow for loop

in g-lay cluster analysis numbers inside the bracket represents the total number of residues in the cluster

Sr.no	PDB id	Residue	Degree	B.C.	C.C.	G-lay cluster	T_m (^{0}C)
1	1XAA	ALA172	18	0.0147	0.4722	4(40)	64.5
2	1G2U	VAL172	18	0.0359	0.3333	3(80)	65.5
3	1GC8	PHE172	19	0.0727	0.3181	2(48)	67.2
4	1GC9	GLY172	10	0.0044	0.4666	4(38)	64.2

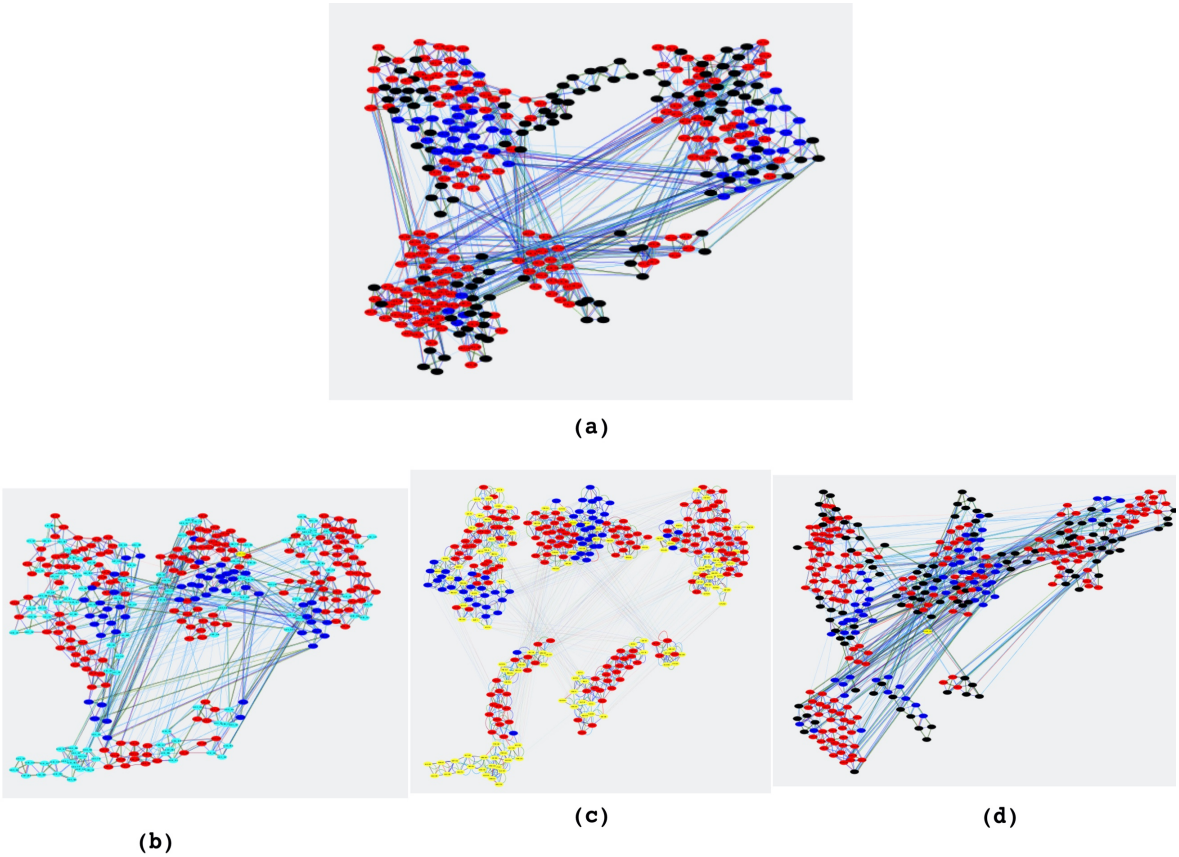
In 3ID there linear correlation of average betweenness centrality and melting temperature.

Fig-3.3.1) temperature(y-axis) vs average betweenness centrality(x-axis) graph



community cluster analysis of 3ID-

Fig-3.3.2) community cluster analysis of 3ID



a)1XAA(WT) has 8 clusters.

b)1G2U(A172V) has 8 clusters.

c)1GC8(A172F) has densely packed 5 clusters.

d)1GC9(A172G) has 9 clusters.

There is enough evidence from community cluster analysis that thermostable proteins are compactly packed so I didn't go for MD simulation analysis.

3.4.RIN and MDS data of T4 lysozyme (PDB ID-4S0W)

Classification**:** HYDROLASE (O-GLYCOSYL). Residue count-164

Organism(s): Enterobacteria phage T4

T4 lysozyme is enzyme that attacks bacterial cell wall, the enzyme particularly attacks peptidoglycans in the cell wall of bacteria and hydrolyses the ß-1,4 linkages between N-acetylmuramic acid and N-acetylglucosamine.it has several applications in industries like bacterial lysis for nucleic acid extraction.

The T4 lysozyme mutant SER117PHE(1TLA) was isolated and found to be more thermostable than wild- type by 1 -1.4 kcal/mol[34],there is also low thermostable pseudo wild type(2OE4) we will be comparing both with wild type(4S0W).

PCN analysis of T4 lysozyme-

For Global-scale the average network parameters were computed for each PCN, the ranges of such parameter value for 3PCN's are-

Values here are of wild type and most thermostable mutant structure

1.Network density- 0.046-0.044

2.Betweenness Centrality- 0.02-0.016

3.Clustering Coefficient- 0.487-0.465

4.Shortest Path Length- 7.444-7.160

comparative study of residue specific network parameters.

Table-3.4.1) residue specific network parameters of T4 lysozyme

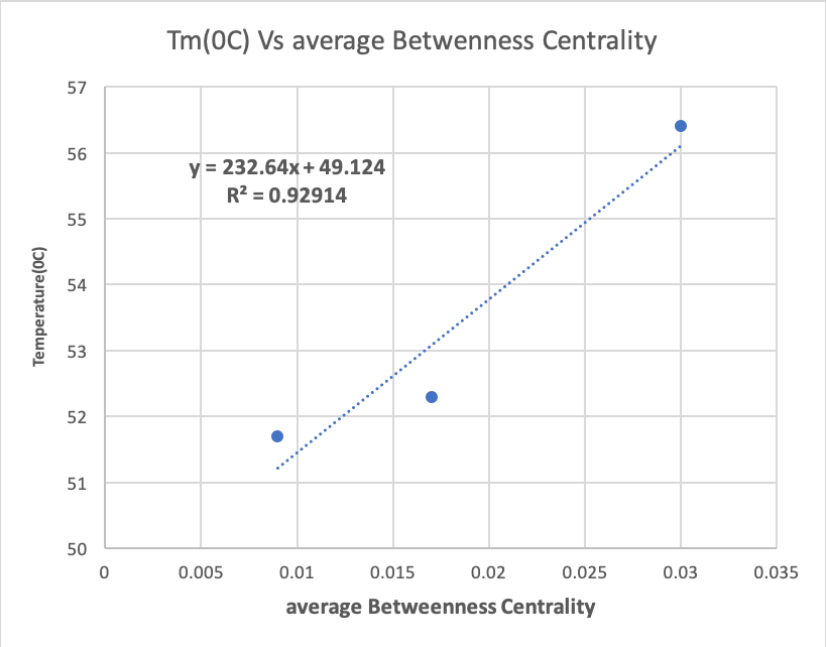
colour code-Red for helix, Blue for sheet, yellow for loop

in g-lay cluster analysis numbers inside the bracket represents the total number of residues in the cluster

Sr.no	PDB id	residue	Degree	B.C.	C.C.	G-lay cluster	T_m (^{0}C)
1	4S0W (wt)	SER117	15	0.017	0.43	1(51)	52.3
2	1TLA	PHE117	17	0.03	0.31	1(51)	56.4
3	2OE4 (pseudo Wild type)	SER117	15	0.009	0.41	2(48)	51.7

here also we see that there is linear relationship of B.C. with melting temperature.

Fig-3.4.1) temperature(y-axis) vs average betweenness centrality(x-axis) graph



117F mutant is buried in the hydrophobic core of the protein also Phe117 and Phe153 shows edge-face interaction[34], This might be reason for more thermostability than wild type.

community cluster analysis of T4 lysozyme-

fig-3.4.2) community cluster analysis of T4 lysozyme

a-c labelling from right to left.

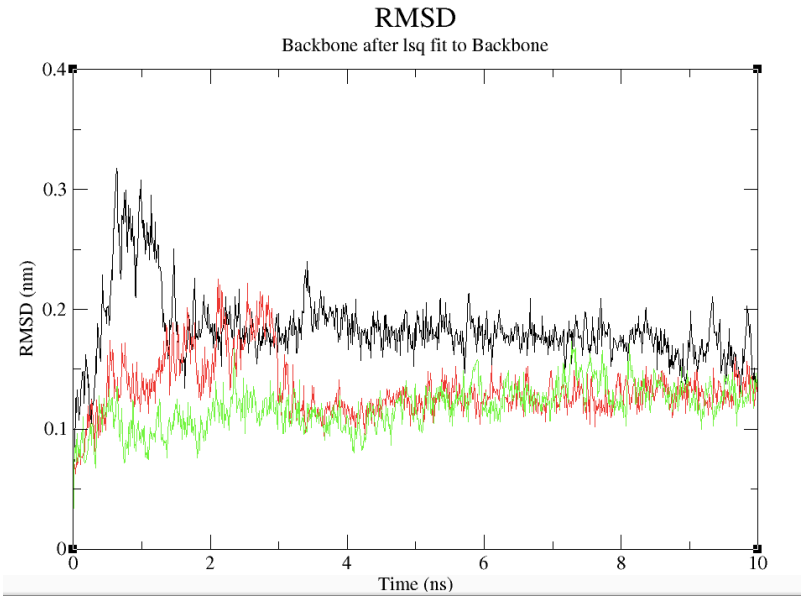
a)4S0W(wt) here we see community 1 doesn't interact with community 2 and widely spread.

b)1TLA(S117F) has good communication of community 1 with community 2.

c)2OE4(pseudo wt) here residue 117 shifts from community 1 to community 3 that have less members.

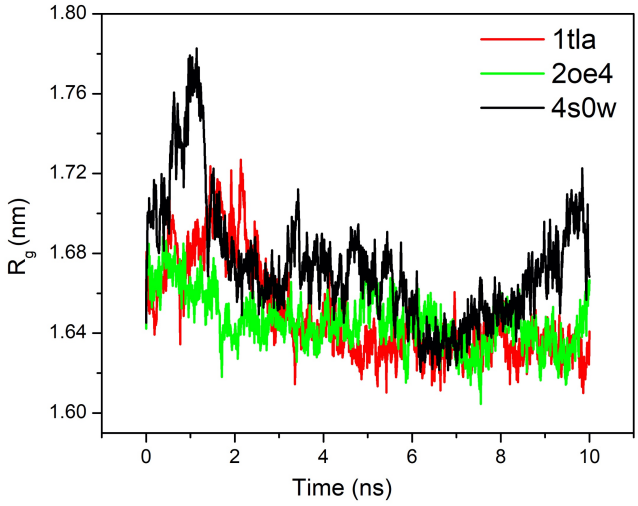
Molecular Dynamics Simulation analysis-

Fig-3.4.3) RMSD vs Time plot of T4 lysozyme (Black-4S0W, Green-2OE4, Red-1TLA)



The RMSD simulation of 10ns at 325K showed that the most thermostable mutant(1TLA-Red) and high-pressure pseudo wild type(2OE4-Green) showed less fluctuation compared to wild type. In wild type and thermostable mutant structure, we see sudden increase in RMSD value before 2ns and then it achieves stable graph until 10ns.

Fig-3.4.4) Radius of gyration vs time plot of T4 lysozyme (Black-4S0W, Green-2OE4, Red-1TLA)

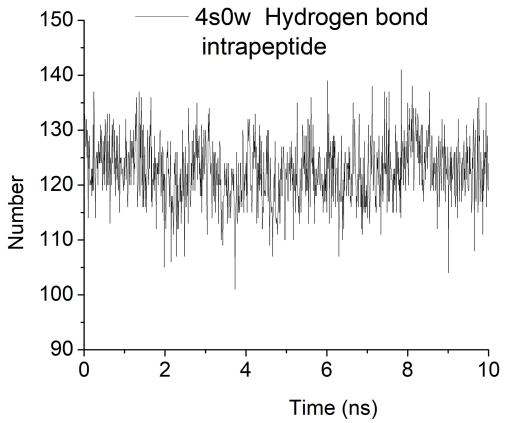


1TLA(Red) showed decreasing graph of Rg compared to both structure it means that 1TLA has tighter packing, Rg concerns with how can secondary structure compactly packed into 3D structure of protein. Wild type structure showed increasing value of Rg after 8ns due to deformity at melting temperature i.e. structure might be unfolding.

Hydrogen-Bond analysis-

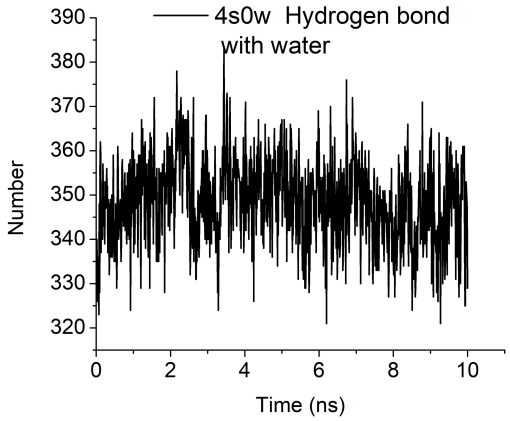
H-B intra peptide

Fig-3.4.5)4S0W(WT)



H-B peptide-water

Fig-3.4.6)4S0W(WT)



ig-3.4.7)1TLA

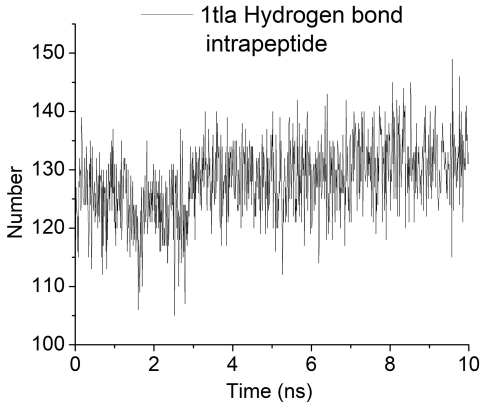


Fig-3.4.8)1TLA

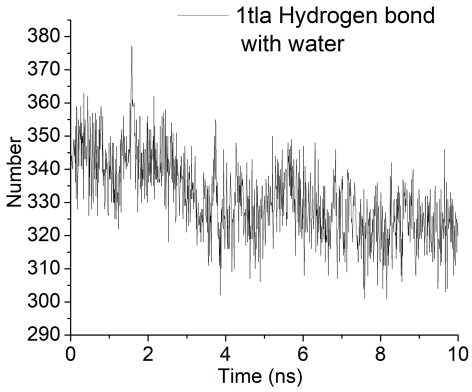


Fig-3.4.9)2OE4

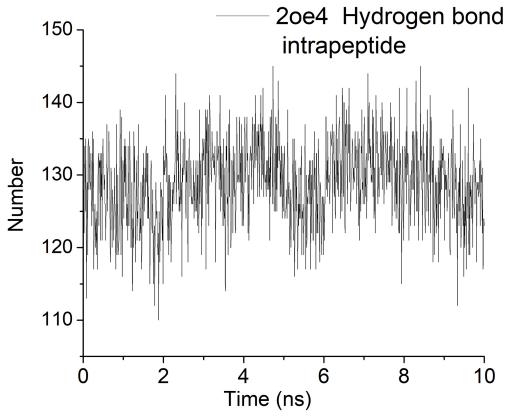
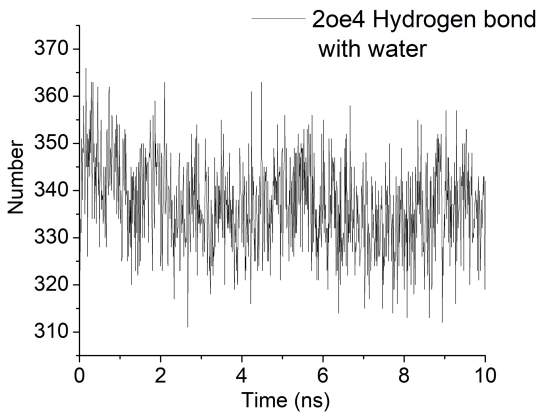


Fig-3.4.10)2OE4



in H-B analysis 2OE4 has least thermal stability which shows decreasing trend in intra-peptide hydrogen bonding and increasing trend in peptide-water hydrogen bonding.

1TLA which is most thermostable mutant of them shows exactly opposite behaviour compared to 2OE4 and 4S0W.

3.5.RIN and MDS data of GH10 xylanase (PDB ID-4QCE)

Classification: HYDROLASE

Organism(s): Bacillus sp. NG-27

Xylan is most abundant polymer in hemicellulose, which is considered as most important renewable sources on earth[35].the main function of GH10 xylanase is endohydrolysis of (1-4)-beta-D-glycosidic linkages in xylans[36].hot springs and sea bottom harbour are the source of enzyme. mainly it is present in thermophiles.

Here we select this protein due to here mutation happens at N-terminus region i.e. at 1st position so it has 0 betweenness centrality and having 1.0 clustering coefficient in both wild(4QCE) and mutant(4QDM) having mutation at V1L, mutant2(4QCF) having mutation at V1A[25].

PCN analysis of GH10 xylanase -

For Global-scale the average network parameters were computed for each PCN, the ranges of such parameter value for 3PCN's are-

Values here are of wild type and most thermostable mutant structure.

1.Network density- 0.022-0.022

2.Betweenness Centrality- 0.015-0.018

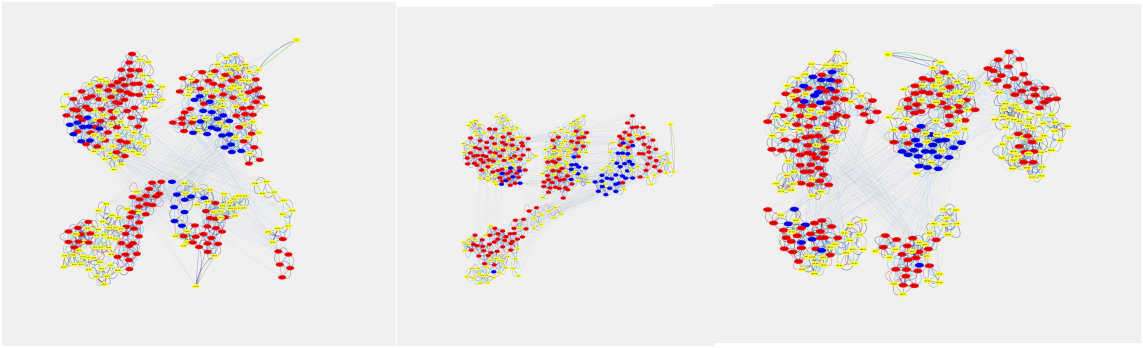
3.Clustering Coefficient- 0.417-0.430

4.Shortest Path Length- 7.667-7.791

Due to the fact that mutation only happens at N-terminus end, here we only done community cluster analysis which is comparable parameter.

Community cluster analysis-

Fig-3.5.1) community cluster analysis of GH10 xylanase



a)4QCE b)4QDM c)4QCF

a)4QCE(WT) here are total 5 communities in the clusters.

Valine has belonged to community 2 consisting of 78 community members.

T_m=70^{0}C

b)4QDM(V1L) here we see total 4 community clusters more compact than wt and 4QCF.

Leucine belonged to community 3 consisting of 80 community members.

T_m=75^{0}C

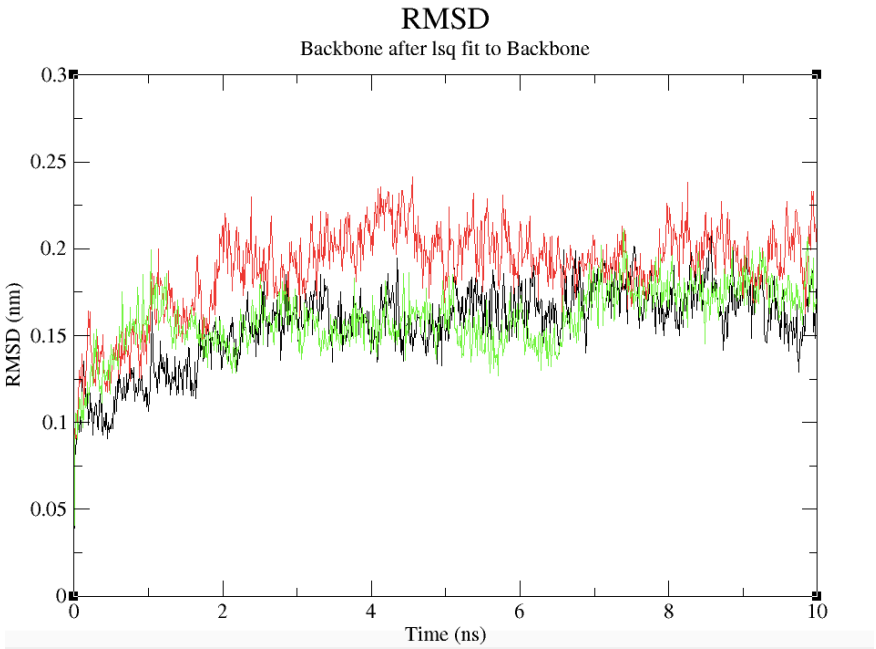
c)4QCF(V1A) here we see total 5 community clusters.

Alanine belonged to community 2 consisting of 76 community members.

T_m=68^0

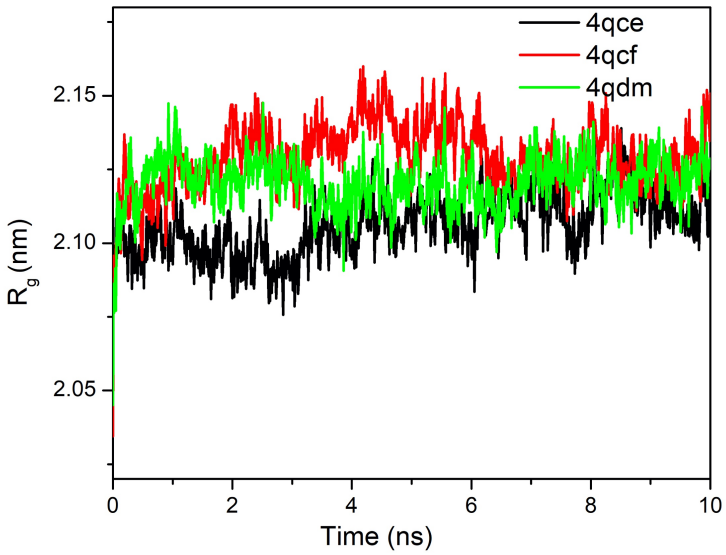
Molecular Dynamics Simulation analysis-

Fig-3.5.2) RMSD vs Time plot of GH10 Xylanase (Black-4QCE, Green-4QDM, Red-4QCF)



Here we can see that the less thermostable mutant(4QCF) and wild type(4QCE) shows more fluctuations compared to more thermostable mutant(4QDM), smaller the deviations more stable the structure so 4QDM shows more thermal stability.

Fig-3.5.3) Radius of gyration vs time plot of GH10 xylanase (Black-4QCE, Green-4QDM, Red-4QCF)



In Rg vs time plot all structures shows nearly similar graphs but 4QDM(Green) shows less fluctuations with time as compared to other structures while 4QCF(Red) which has low thermostability firstly showed high Rg value compared to wild type (Black).

Hydrogen-bond analysis-In GH10 xylanase mutation introduced at N-terminal so there is no significant change in hydrogen bond patterns.so there is no need for H-B analysis.

Conclusion

- In all cases of different thermostable mutant proteins betweenness centrality of nodes i.e. mutated residues have shown linear relationship with melting temperature.
- Closely packed and less number of G-lay cluster were seen in thermostable proteins.
- In mutants, high B.C. were seen but there is low C.C. i.e. High connectivity dominates ability to form groups with neighbours but in some cases like ADK both increases thereby having 30% increase in thermal stability.
- In some cases, like yCD at 23 position there is slight decrease in B.C. but there is increase in C.C. and member of clusters so all three factors are important to maintain thermal stability.
- In all thermostable mutant proteins, we see that mutated residue breaks community structure and goes to more dense community and when mutated residue goes to less dense community structure we see that there is no significant change in melting temperature.
- As unfolding of protein is cooperative process i.e. loss of structure in one part of the protein destabilizes the whole structure thereby loss of function so High connectivity between residues is important to maintain its native form.
- Molecular Dynamics Simulations results also confirms that thermostable mutant structure showed less fluctuation of RMSD value with Time, and radius of gyration of thermostable mutant structure showed decreased Rg value compared to wild type it conveys that tight packing is also necessary for the thermal stability.
- In Hydrogen bonds analysis, we observe that for thermostable mutant's intra-peptide hydrogen bonds shows increasing trend over time compared to wild type protein Also, we observe that for thermostable mutant's peptide-water hydrogen bonds shows decreasing trend over time compared to wild type protein.

Bibliography

(1) Kandhari, N.; Sinha, S. Complex Network Analysis of Thermostable Mutants of Bacillus Subtilis Lipase A. *Appl. Netw. Sci.* **2017**, *2* (1), 18. https://doi.org/10.1007/s41109-017-0039-y.

(2) Takami, H.; Takaki, Y.; Chee, G.-J.; Nishi, S.; Shimamura, S.; Suzuki, H.; Matsui, S.; Uchiyama, I. Thermoadaptation Trait Revealed by the Genome Sequence of Thermophilic Geobacillus Kaustophilus. *Nucleic Acids Res.* **2004**, *32* (21), 6292–6303. https://doi.org/10.1093/nar/gkh970.

(3) Srivastava, A.; Sinha, S. Thermostability of In Vitro Evolved Bacillus Subtilis Lipase A: A Network and Dynamics Perspective. *PLoS ONE* **2014**, *9* (8). https://doi.org/10.1371/journal.pone.0102856.

(4) Kannan, N.; Vishveshwara, S. Identification of Side-Chain Clusters in Protein Structures by a Graph Spectral Method11Edited by J. M. Thornton. *J. Mol. Biol.* **1999**, *292* (2), 441–464. https://doi.org/10.1006/jmbi.1999.3058.

(5) Bagler, G.; Sinha, S. Network Properties of Protein Structures. *Phys. Stat. Mech. Its Appl.* **2005**, *346* (1), 27–33. https://doi.org/10.1016/j.physa.2004.08.046.

(6) Bagler, G.; Sinha, S. Assortative Mixing in Protein Contact Networks and Protein Folding Kinetics. *Bioinformatics* **2007**, *23* (14), 1760–1767. https://doi.org/10.1093/bioinformatics/btm257.

(7) Barah, P.; Sinha, S. Analysis of Protein Folds Using Protein Contact Networks. *Pramana* **2008**, *71* (2), 369–378. https://doi.org/10.1007/s12043-008-0170-5.

(8) Vendruscolo, M.; Dokholyan, N. V.; Paci, E.; Karplus, M. Small-World View of the Amino Acids That Play a Key Role in Protein Folding. *Phys. Rev. E* **2002**, *65* (6), 061910. https://doi.org/10.1103/PhysRevE.65.061910.

(9) Doncheva, N. T.; Klein, K.; Domingues, F. S.; Albrecht, M. Analyzing and Visualizing Residue Networks of Protein Structures. *Trends Biochem. Sci.* **2011**, *36* (4), 179–182. https://doi.org/10.1016/j.tibs.2011.01.002.

(10) Tsai, C.-J.; Sol, A. del; Nussinov, R. Protein Allostery , Signal Transmission and Dynamics: A Classification Scheme of Allosteric Mechanisms. *Mol. Biosyst.* **2009**, *5* (3), 207–216. https://doi.org/10.1039/B819720B.

(11) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.

(12) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79* (16), 3090–3093. https://doi.org/10.1103/PhysRevLett.79.3090.

(13) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(14) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(15) Network Density - an overview | ScienceDirect Topics https://www.sciencedirect.com/topics/computer-science/network-density (accessed Apr 28, 2020).

(16) Liu, W.; Li, X.; Liu, T.; Liu, B. Approximating Betweenness Centrality to Identify Key Nodes in a Weighted Urban Complex Transportation Network https://www.hindawi.com/journals/jat/2019/9024745/ (accessed Apr 22, 2020). https://doi.org/10.1155/2019/9024745.

(17) Su, G.; Kuchinsky, A.; Morris, J. H.; States, D. J.; Meng, F. GLay: Community Structure Analysis of Biological Networks. *Bioinformatics* **2010**, *26* (24), 3135–3137. https://doi.org/10.1093/bioinformatics/btq596.

(18) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *Mol. Model. Annu.* **2001**, *7* (8), 306–317. https://doi.org/10.1007/s008940100045.

(19) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids | Journal of the American Chemical Society https://pubs.acs.org/doi/10.1021/ja9621760 (accessed Apr 22, 2020).

(20) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(21) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.

https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(22) Singh, B.; Bulusu, G.; Mitra, A. Understanding the Thermostability and Activity of Bacillus Subtilis Lipase Mutants: Insights from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119* (2), 392–409. https://doi.org/10.1021/jp5079554.

(23) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(24) Graphing with Gnuplot and Xmgr | Linux Journal https://www.linuxjournal.com/article/1218 (accessed Apr 22, 2020).

(25) Aier, I.; Varadwaj, P. K.; Raj, U. Structural Insights into Conformational Stability of Both Wild-Type and Mutant EZH2 Receptor. *Sci. Rep.* **2016**, *6* (1), 1–10. https://doi.org/10.1038/srep34984.

(26) Ko, T.-P.; Lin, J.-J.; Hu, C.-Y.; Hsu, Y.-H.; Wang, A. H.-J.; Liaw, S.-H. Crystal Structure of Yeast Cytosine Deaminase INSIGHTS INTO ENZYME MECHANISM AND EVOLUTION. *J. Biol. Chem.* **2003**, *278* (21), 19111–19117. https://doi.org/10.1074/jbc.M300874200.

(27) Ireton, G. C.; Black, M. E.; Stoddard, B. L. The 1.14 A Crystal Structure of Yeast Cytosine Deaminase: Evolution of Nucleotide Salvage Enzymes and Implications for Genetic Chemotherapy. *Struct. Lond. Engl. 1993* **2003**, *11* (8), 961–972. https://doi.org/10.1016/s0969-2126(03)00153-9.

(28) Kievit, E.; Bershad, E.; Ng, E.; Sethna, P.; Dev, I.; Lawrence, T. S.; Rehemtulla, A. Superiority of Yeast over Bacterial Cytosine Deaminase for Enzyme/Prodrug Gene Therapy in Colon Cancer Xenografts. *Cancer Res.* **1999**, *59* (7), 1417–1421.

(29) Black, M. E. Enzyme and Pathway Engineering for Suicide Gene Therapy. *Genet. Eng. (N. Y.)* **2001**, *23*, 113–127. https://doi.org/10.1007/0-306-47572-3_7.

(30) Greco, O.; Dachs, G. U. Gene Directed Enzyme/Prodrug Therapy of Cancer: Historical Appraisal and Future Prospectives. *J. Cell. Physiol.* **2001**, *187* (1), 22–36. https://doi.org/10.1002/1097-4652(2001)9999:9999<::AID-JCP1060>3.0.CO;2-H.

(31) Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L. Computational Thermostabilization of an Enzyme. *Science* **2005**, *308* (5723), 857–860. https://doi.org/10.1126/science.1107387.

(32) Miller, C.; Davlieva, M.; Wilson, C.; White, K. I.; Couñago, R.; Wu, G.; Myers, J. C.; Wittung-Stafshede, P.; Shamoo, Y. Experimental Evolution of Adenylate Kinase Reveals

Contrasting Strategies toward Protein Thermostability. *Biophys. J.* **2010**, *99* (3), 887–896. https://doi.org/10.1016/j.bpj.2010.04.076.

(33) Nagata, C.; Moriyama, H.; Tanaka, N.; Nakasako, M.; Yamamoto, M.; Ueki, T.; Oshima, T. Cryocrystallography of 3-Isopropylmalate Dehydrogenase from Thermus Thermophilus and Its Chimeric Enzyme. *Acta Crystallogr. D Biol. Crystallogr.* **1996**, *52* (4), 623–630. https://doi.org/10.1107/S0907444995016623.

(34) Anderson, D. E.; Hurley, J. H.; Nicholson, H.; Baase, W. A.; Matthews, B. W. Hydrophobic Core Repacking and Aromatic-Aromatic Interaction in the Thermostable Mutant of T4 Lysozyme Ser 117-->Phe. *Protein Sci. Publ. Protein Soc.* **1993**, *2* (8), 1285–1290. https://doi.org/10.1002/pro.5560020811.

(35) Fredriksen, L.; Stokke, R.; Jensen, M. S.; Westereng, B.; Jameson, J.-K.; Steen, I. H.; Eijsink, V. G. H. Discovery of a Thermostable GH10 Xylanase with Broad Substrate Specificity from the Arctic Mid-Ocean Ridge Vent System. *Appl. Environ. Microbiol.* **2019**, *85* (6). https://doi.org/10.1128/AEM.02970-18.

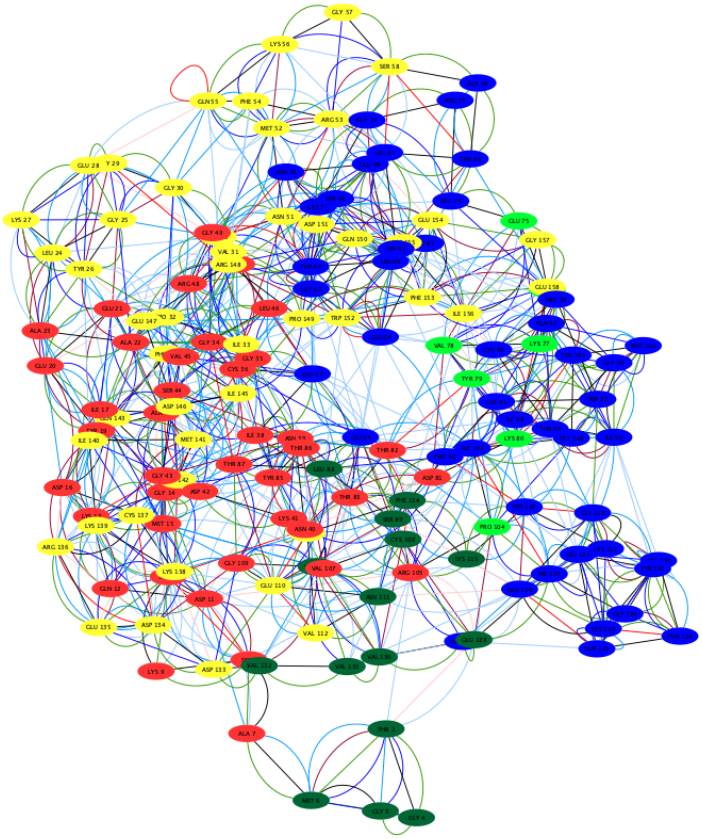
(36) Emami, K.; Nagy, T.; Fontes, C. M. G. A.; Ferreira, L. M. A.; Gilbert, H. J. Evidence for Temporal Regulation of the Two Pseudomonas Cellulosa Xylanases Belonging to Glycoside Hydrolase Family 11. *J. Bacteriol.* **2002**, *184* (15), 4124–4133. https://doi.org/10.1128/jb.184.15.4124-4133.2002.

Appendix

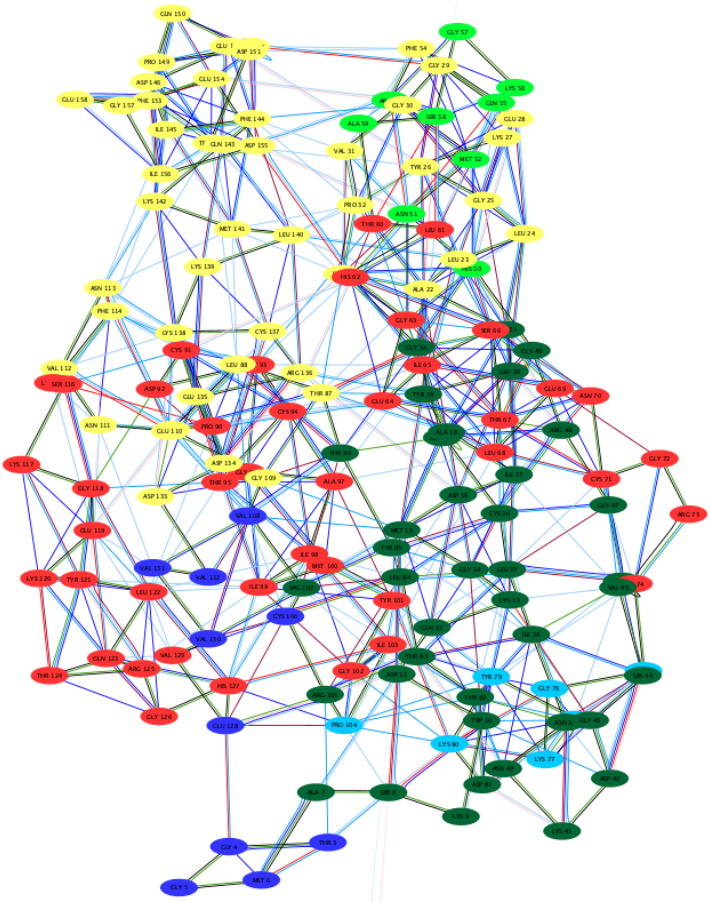
Here all community structure comparison done with respect to its wild type protein.

1.RIN of yeast Cytosine Deaminase-

a)1P6O(WT)

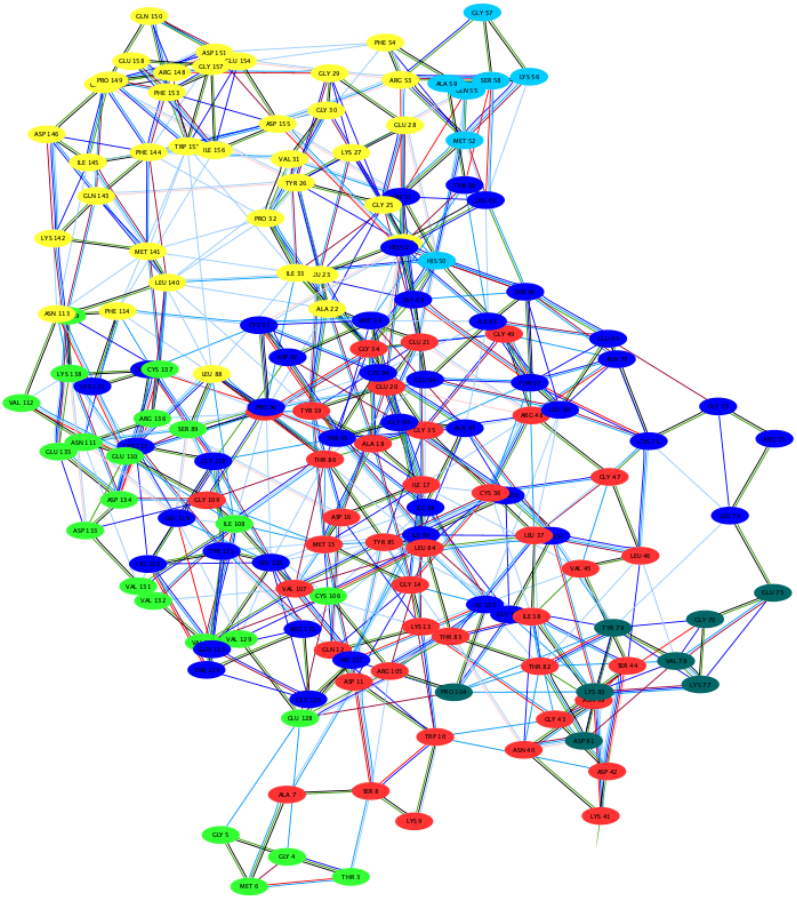


b)1YSD(A23L/I140L)



In 1YSD A23L residue breaks the community structure and goes to C1 where it has more community members, while I140L resides on same community structure as of wild type.

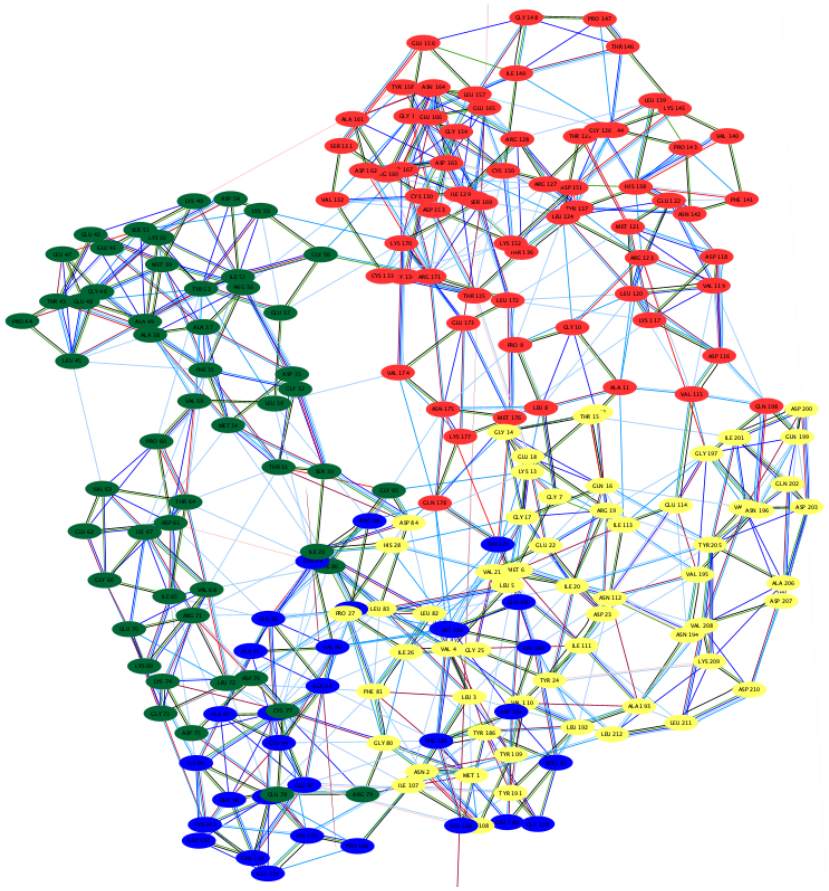
c)1YSB(A23L/V108I/I140L)



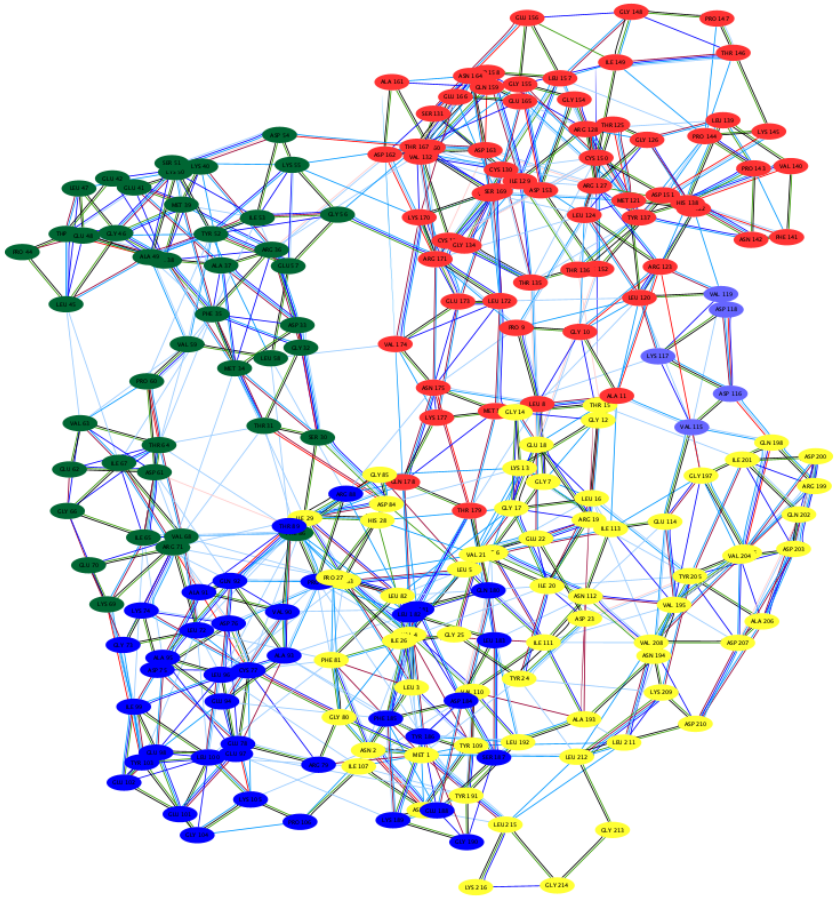
in 1YSB A23L residue changes community structure and goes to C3 where community members are more, V108I residue also changes community structure and goes to less crowded C2 community but there average betweenness centrality is more compared to other community members.

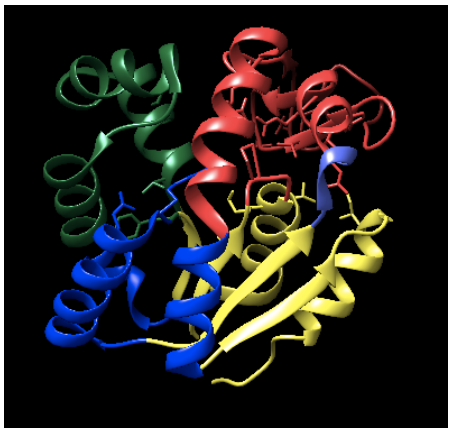
2.RIN of Adenylate Kinase(ADK)

a)1P3J(WT)



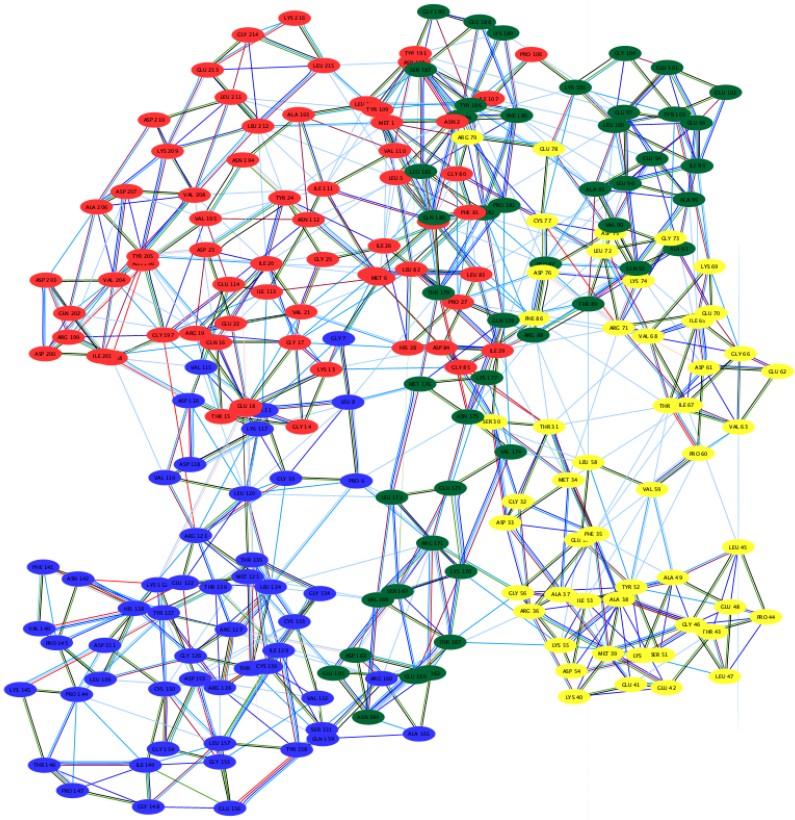
b)2OSB(Q16L/Q199R)

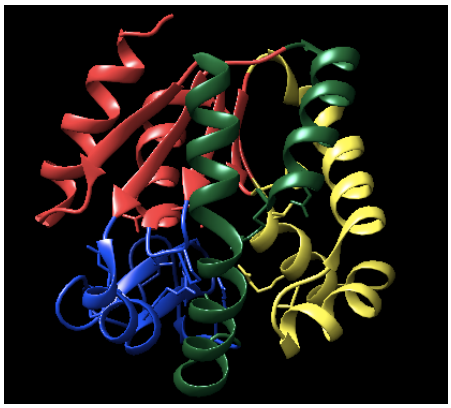




In 2OSB Q16L and Q199R both breaks community structure and goes to more crowded community C2 as compared to wild type protein.

c)2QAJ(Q199R/G213E)

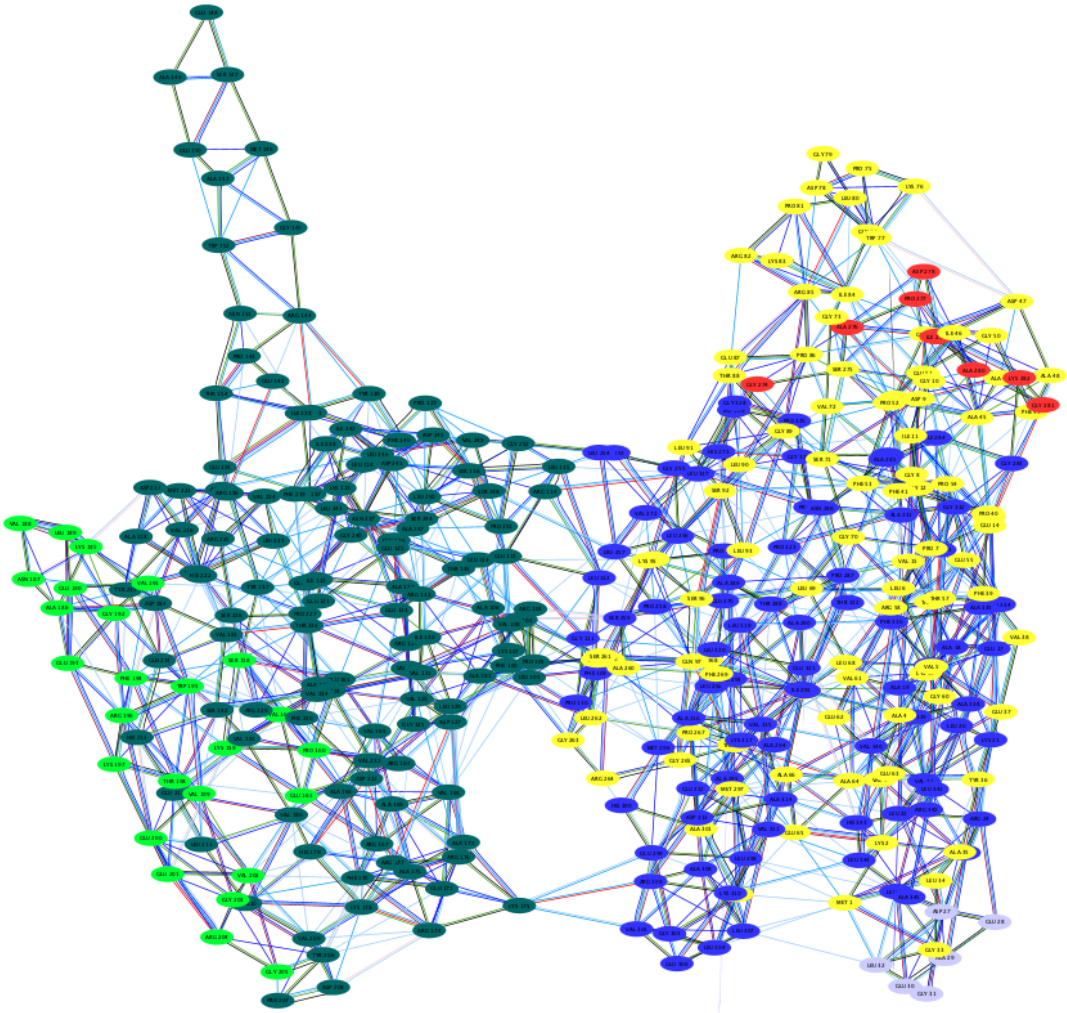


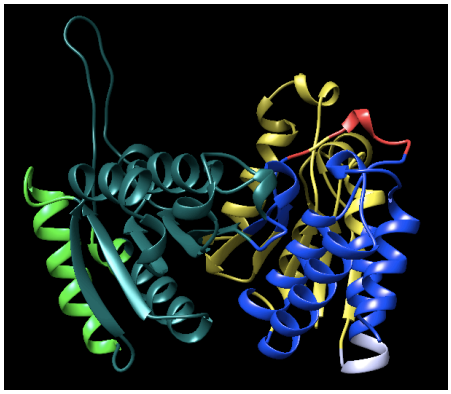


In 2QAJ Q199R residue breaks community structure and goes to more dense community C3 as in wild type Q199 resides on very less dense community consisting of only 9 members, and G213E doesn't change community structure.

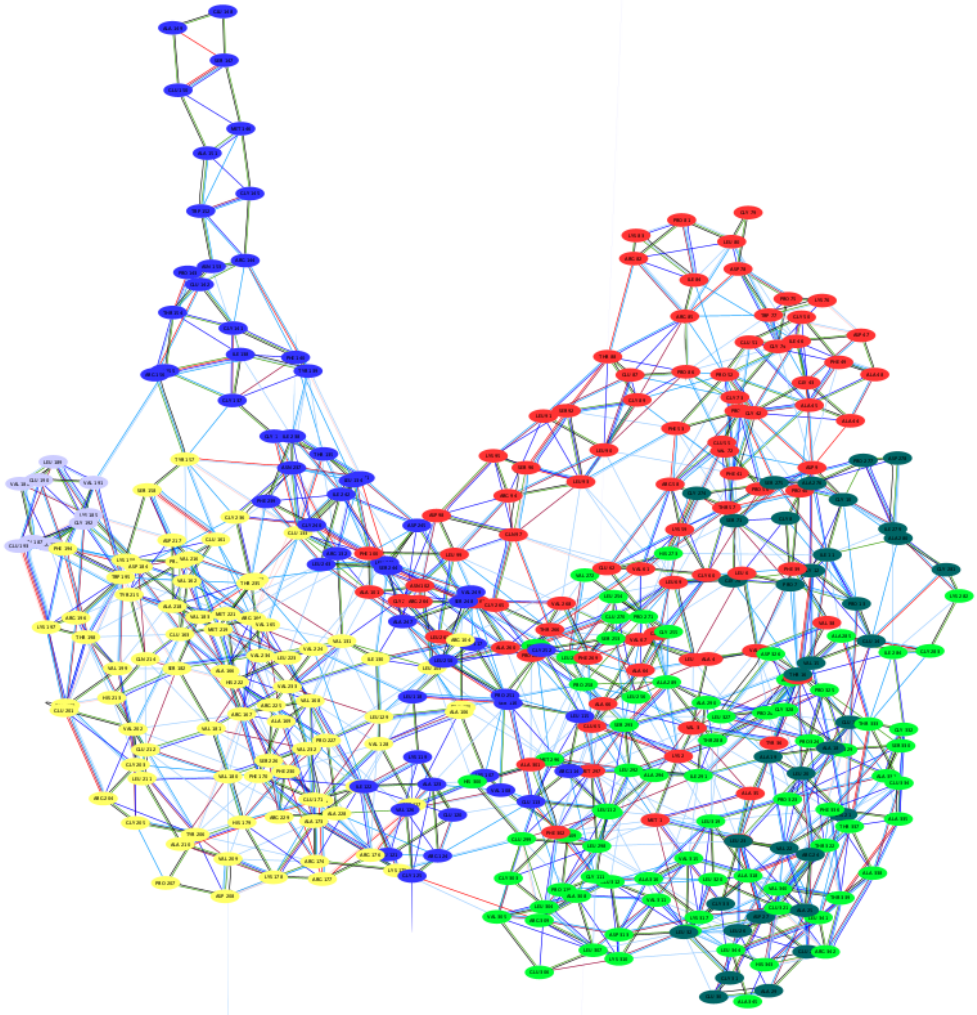
3.3-Isopropylmalate Dehydrogenase (3ID)

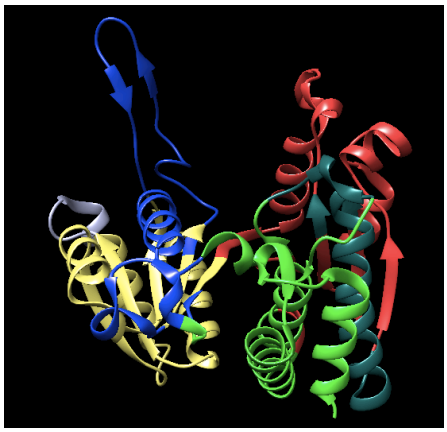
a)1XAA(WT)





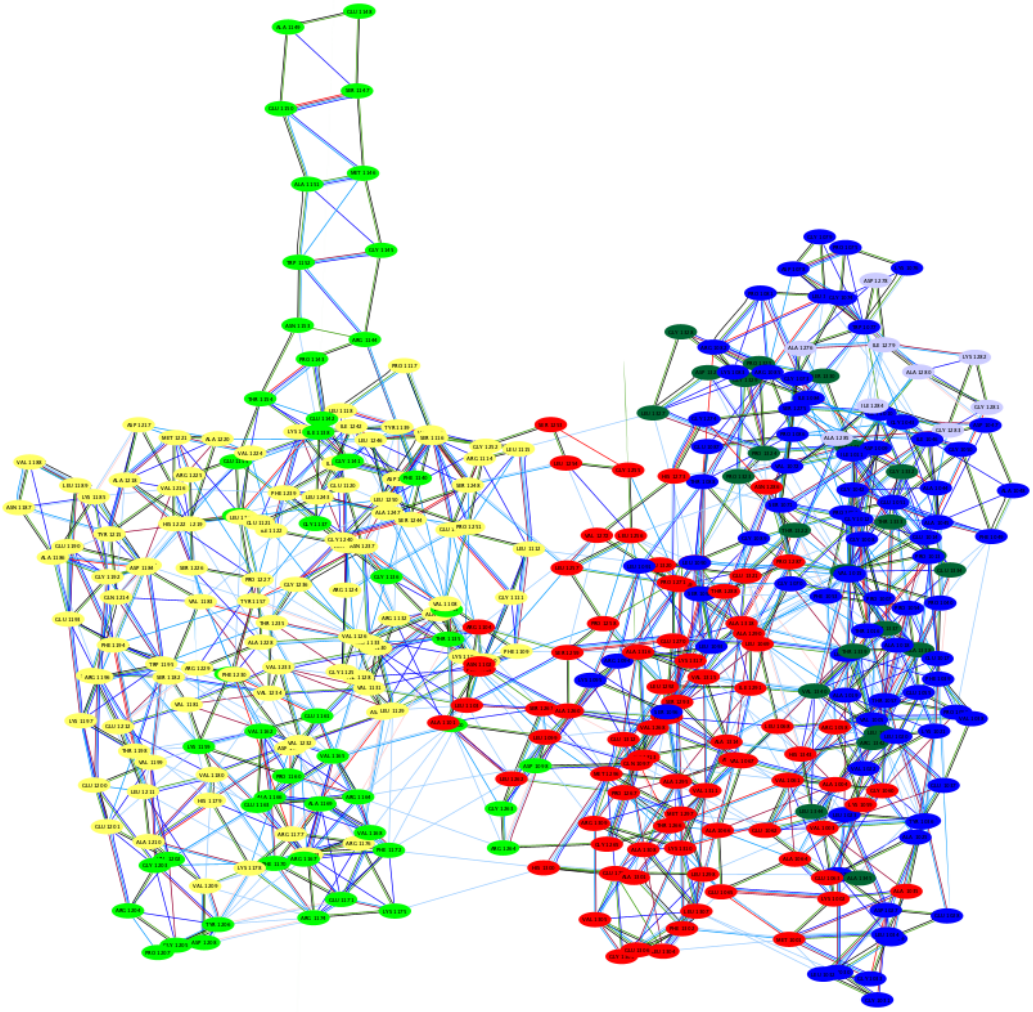
b)1G2U(A172V)

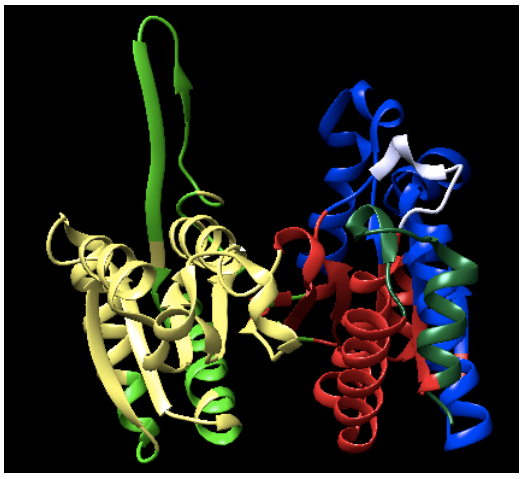




In 1G2U A172V residue breaks community structure and goes to more dense community C3 consisting of 80 community members.

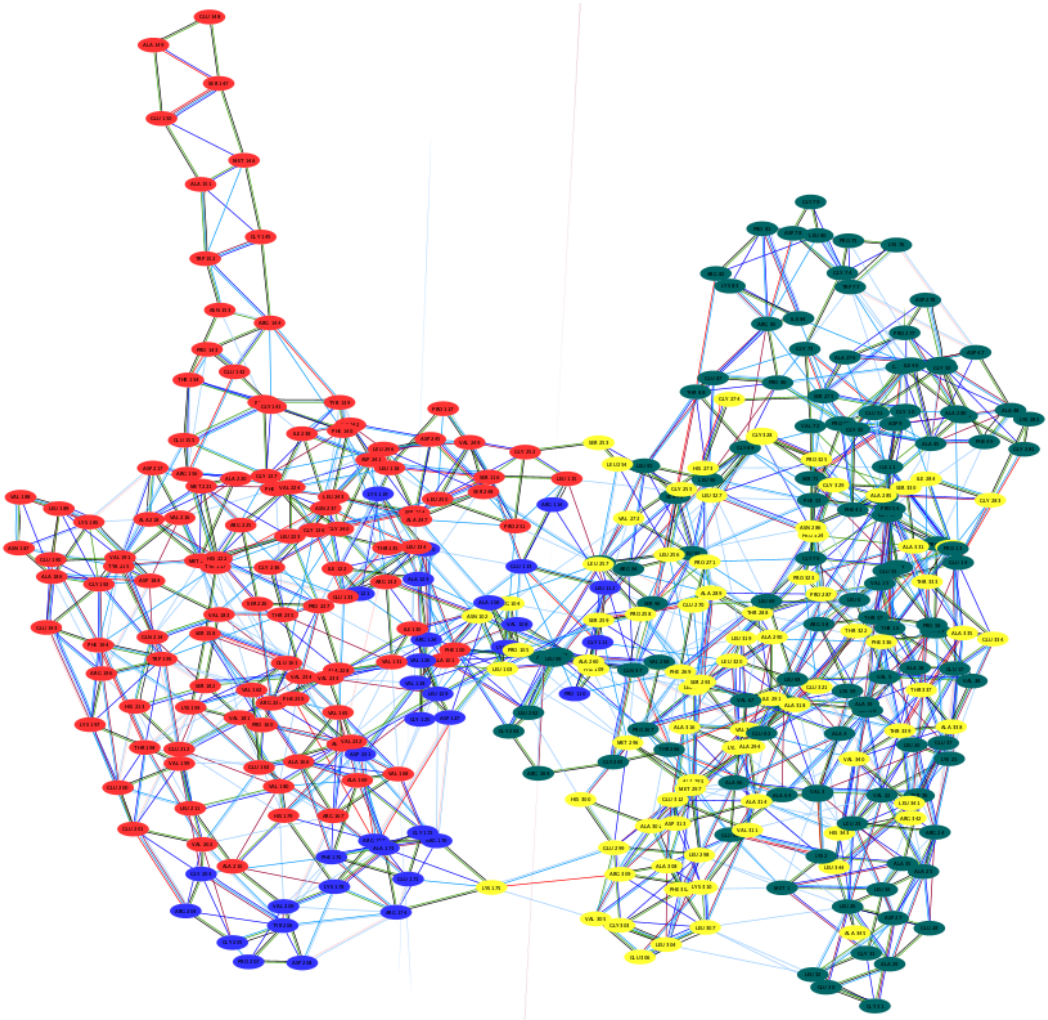
c)1GC8(A172F)

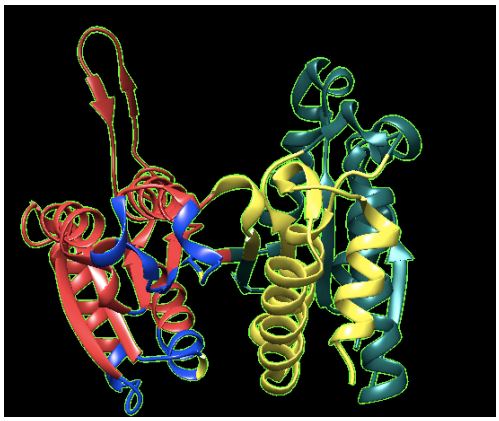




In 1GC8 A172F breaks community structure and goes to C2 which is more dense compared to wild type.

d)1GC9(A172G)

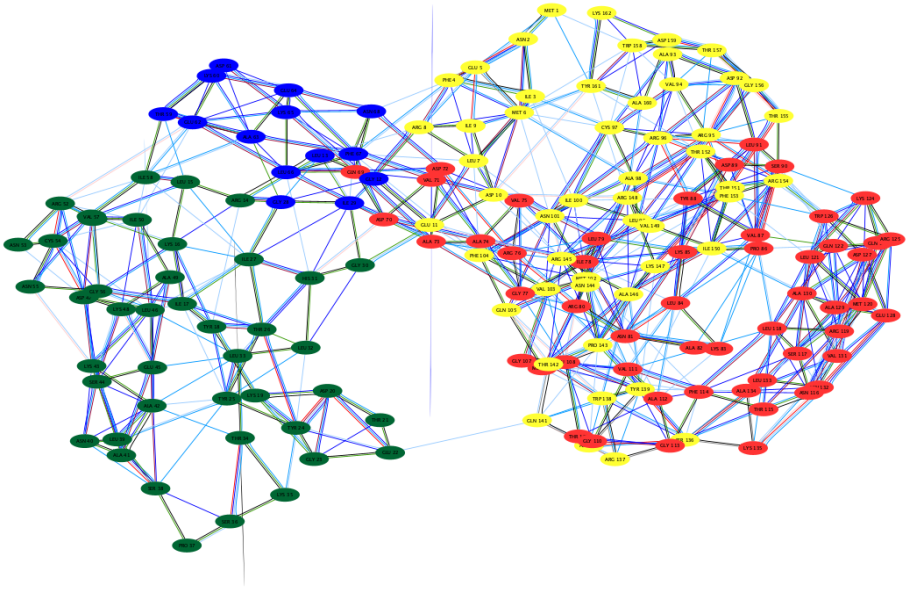




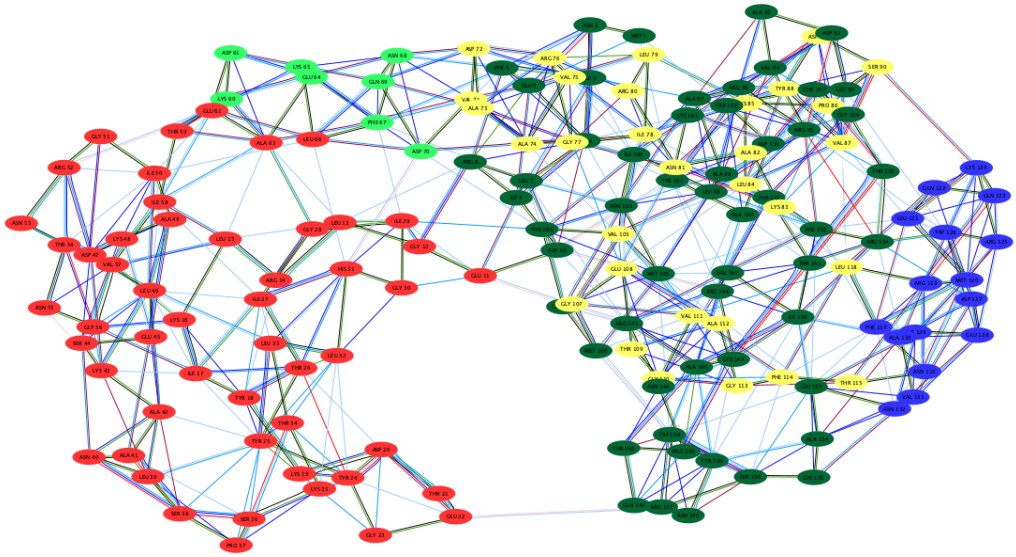
In 1GC9 A172F residue doesn’t change community structure so as we can also see that there is no significant increase in melting temperature compared to wild type protein.

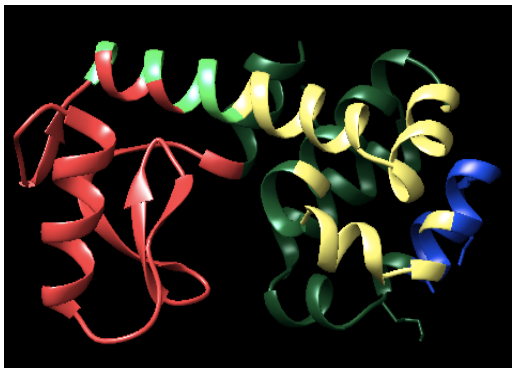
4.RIN OF T4 lysozyme

a)4S0W(WT)



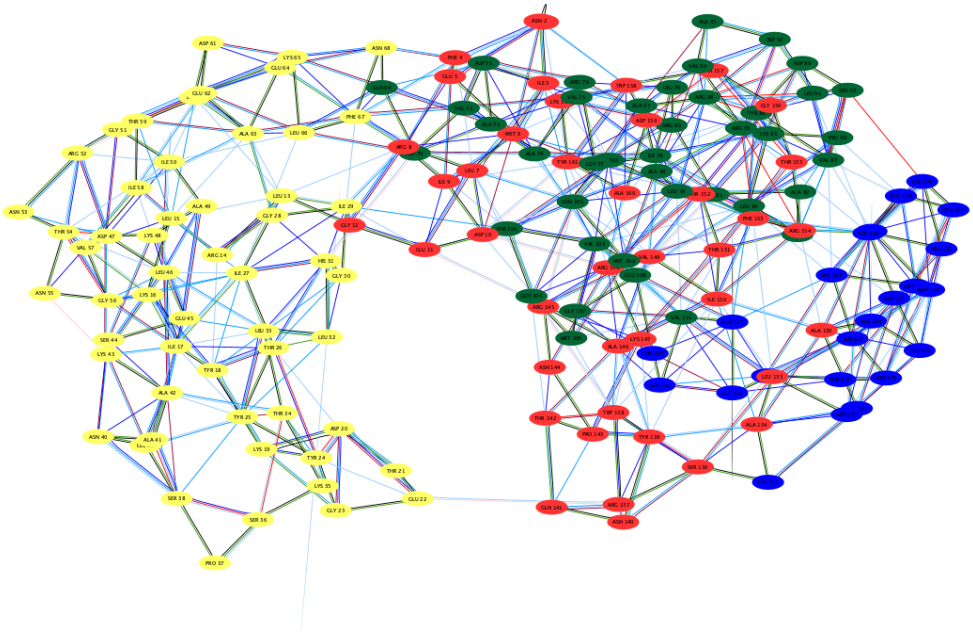
b)1TLA(S117F)

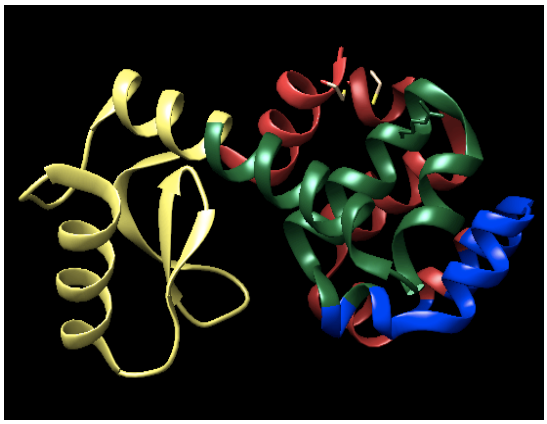




In 1TLA S117F residue doesn't breaks community structure it resides on same community structure as of wild type which is C1 consisting of 51 members.

c)2OE4(high pressure psedo wild type)

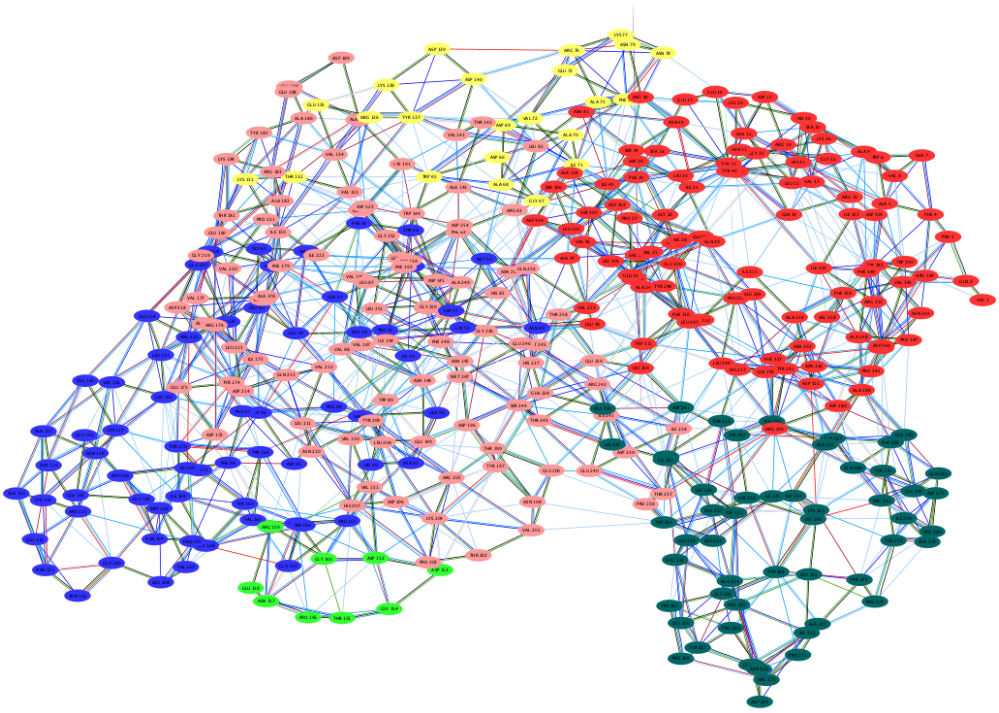


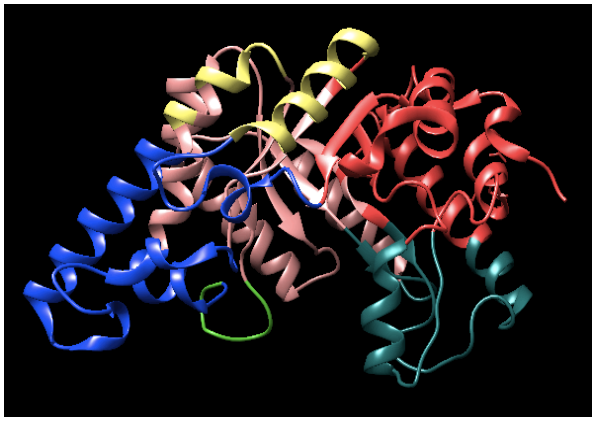


In 2OE4 S117 breaks community structure and goes to less dense community structure C2 consisting of 48 members so here also we see that its melting temperature is less than the wild type protein.

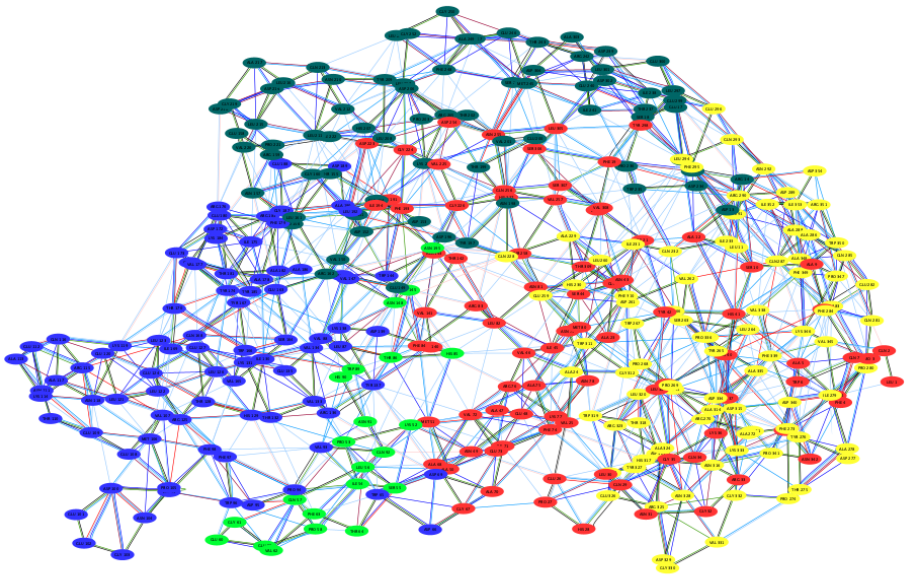
5.GH10 Xylanase

a)4QCE(WT)



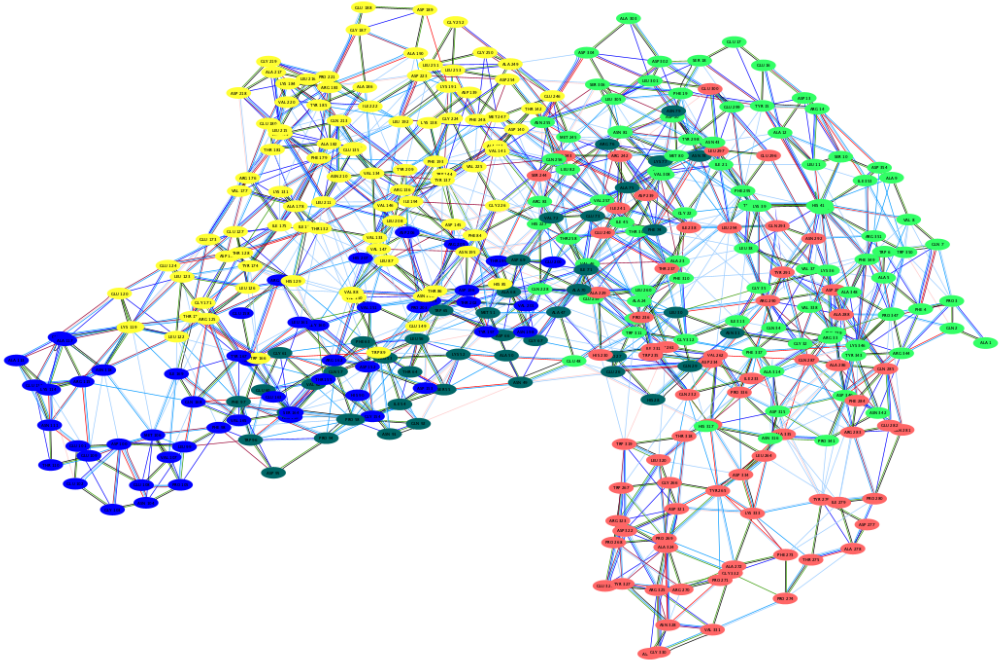


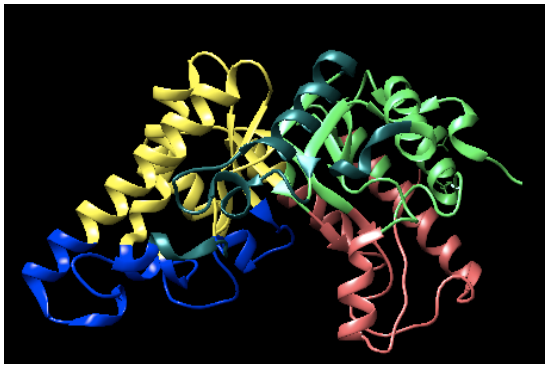
b)4QDM(V1L)



In 4QDM V1L residue doesn’t change community structure but it also has 3 more community members as compared to wild type protein.

c)4QCF(V1A)





In 4QCF V1A residue doesn’t breaks community structure and it has same community members as of wild type protein.