The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

SMR/1845-17

Conference on Structure and Dynamics in Soft Matter and Biomolecules: From Single Molecules to Ensembles

4 - 8 June 2007

Dynamics of protein structure networks

Saraswathi VISHVESHWARA

Indian Institute of Science
Molecular Biophysics Unit
Karnataka
560012 Bangalore
INDIA

Dynamics of Protein Structure Networks

Saraswathi Vishveshwara

Molecular Biophysics Unit

Indian Institute of Science

Bangalore, India

ICTP, June 4, 2007

Outline

1. Introduction to protein structure networks
2. Biologically relevant information from graph/network analysis

3. Dynamics of structure networks
   a) Unfolding simulation of Lysozyme
      Network changes during folding/unfolding transition
   b) Equilibrium simulation of methionyl tRNA synthetase
      Network of communication pathways

Protein structures as Graphs

Concepts of Graph Theory for Protein Structure Analysis

Nodes	Edges	Graph Operation	Purpose	References
Secondary structure (α-helix, β-strand)	Spatially Close Secondary Structures	Identification of Subgraph Isomorphism	Fold & Pattern identification	Mitchell et al., 1989; Grindley et al., 1993;
Secondary structure (α-helix, β-strand)	Spatially Close Secondary Structures (dynamically arrived at)	Dynamical Matrix construction	Testing Folding Rules	Przytycka et al.,2002
Side Chain	Spatial Proximity	Identification of Subgraph Isomorphism	Functionally & Structurally important motif recognition	Artymiuk et al.,1994
Side Chain	Spatial Proximity decided by overlap cut off criterion (Weighted edge)	Graph Spectra, Identification of Clusters and Cluster centers	Identification of clusters important for function, structure and folding	Kannan & Vishveshwara, 1999
Backbone	Spatial neighbours within radius cut off (6.5 - 7.0 Å)	Graph Spectra, Identification of Clusters and Cluster centers	Identification of Proteins with similar folds	Patra & Vishveshwara., 2000
Backbone	Spatial neighbours within radius cut off (7.0 Å)	Graph Spectra	Protein Dynamics	Bahar, 1999
All Atoms	Defined based on Constraints (Weighted Edge)	Graph Spectra	Protein Dynamics	Jacobs et al., 2001

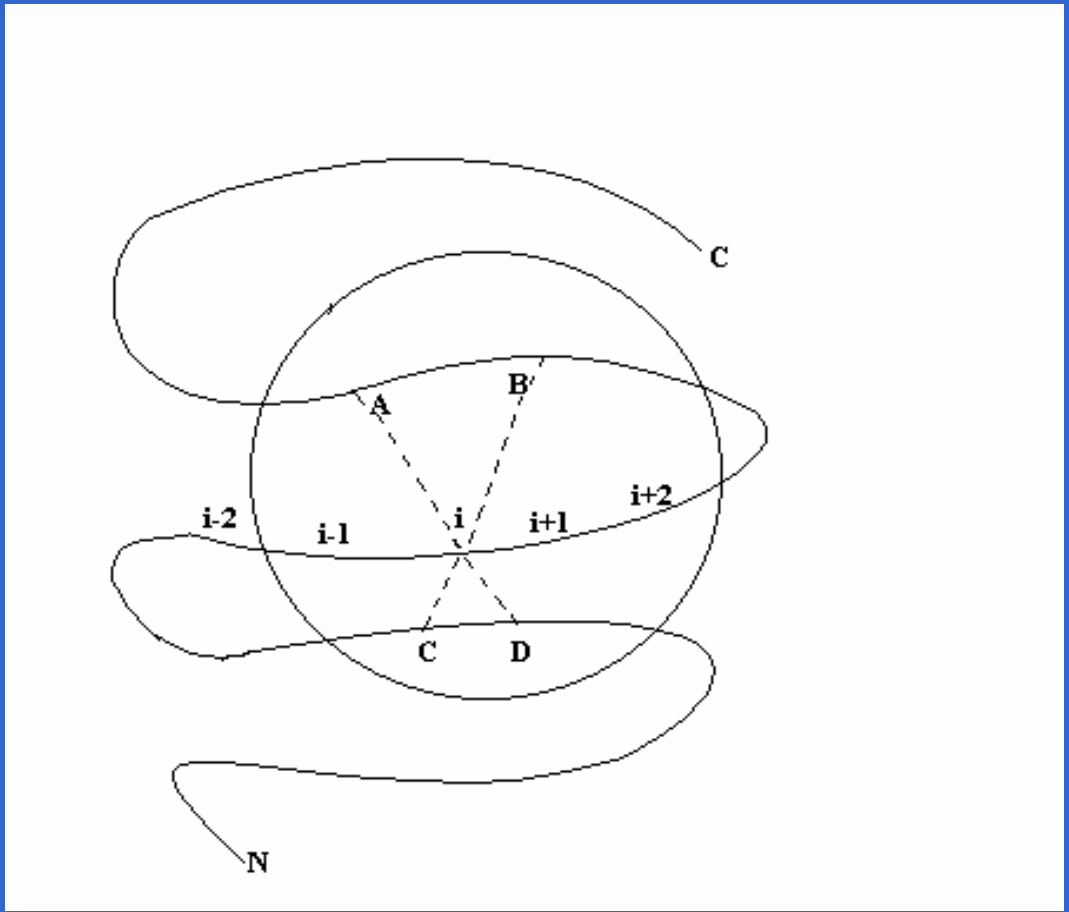
Graph Theory in Molecular Structure.....

Conceptually superior since it takes into account the global topology of the structure unlike pair-wise interactions

Practical advantage: Interactions can be quantified by graph-spectral parameters and single numerical computation can yield the desired results

Protein Graphs

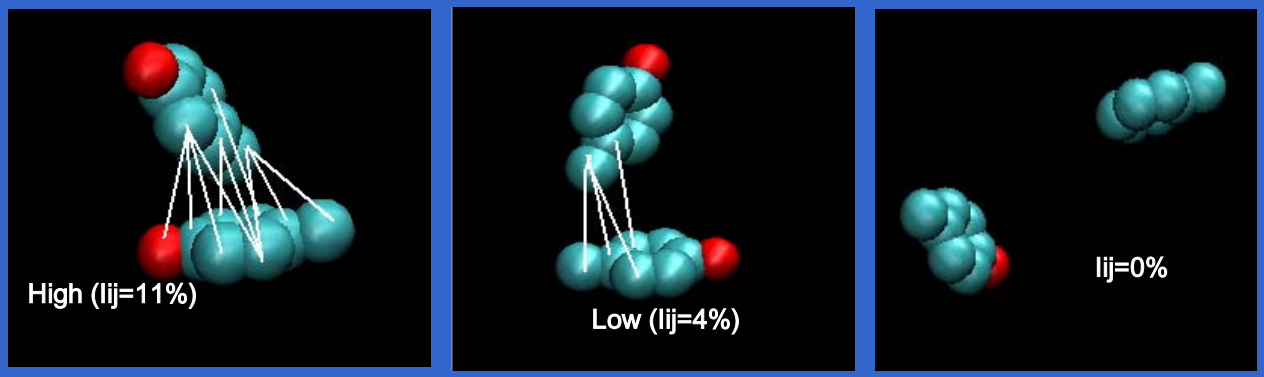
Main Chain Interaction (back bone level)



- Nodes : Amino acid Residues
- Edges : Spatial neighbours within fixed distance

S.M.Patra, Kannan, Vishveshwara, Biophys. Chem (2000); JMB (1999)

Side Chain Interaction



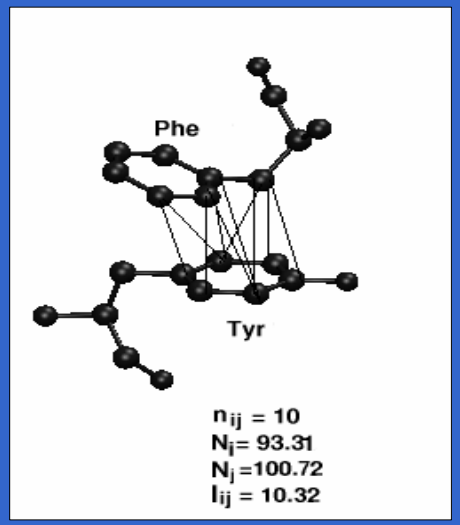
a) High Contact **b) Low Contact** **c) No interaction**

High and low contact criteria. A pair of phenylalanine rings interacting with each other are shown. The lines between the phenylalanines indicate the atoms that are within a distance of 4.5Å.

$$I_{ij} = (n_{ij} \div \sqrt{(N_i * N_j)}) \times 100$$

I_{min} is user defined interaction cutoff.

An (ij) residue pair with $I_{ij} > I_{min}$ is connected by an edge.



Backbone-based versus the Side-chain-based Protein Structure Graphs

Backbone-based (coarse grained)

Based on C-alpha-C-alpha distance

The extent of side-chain interaction is not considered

Side-chain-based

The interactions between sidechains are quantified, hence a weighted graph can be constructed or graphs can be constructed on the basis of the strength of interaction

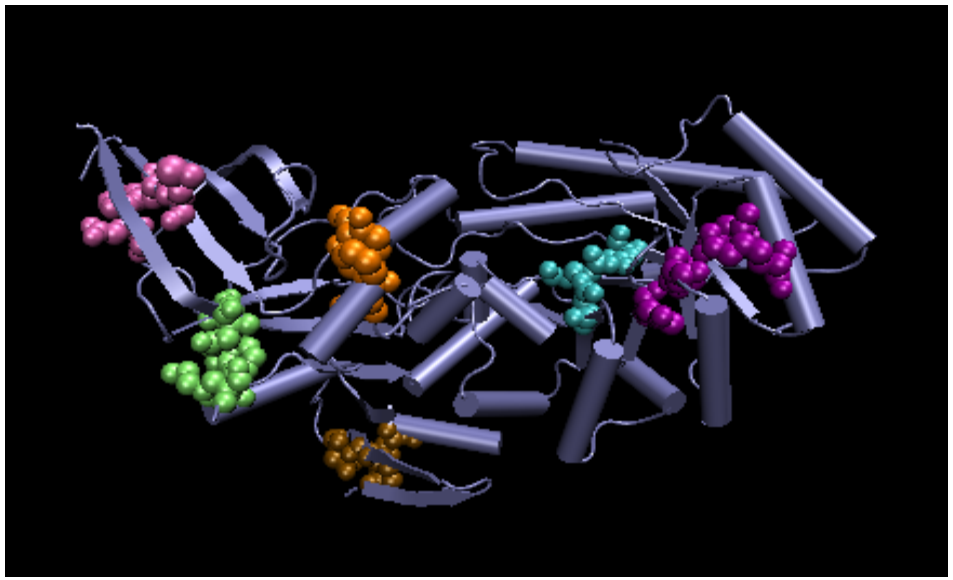
Graph spectral parameters

Provide information on the clusters of interacting residues

Detect cluster centres, which play a crucial role in the integrity of the cluster

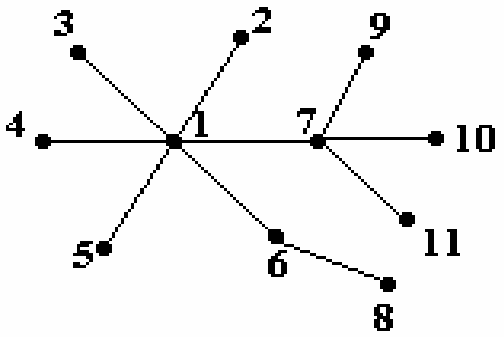
Advantages of Graph Spectral Analysis

- Solution to weighted graph
- Identification of Cluster Centre



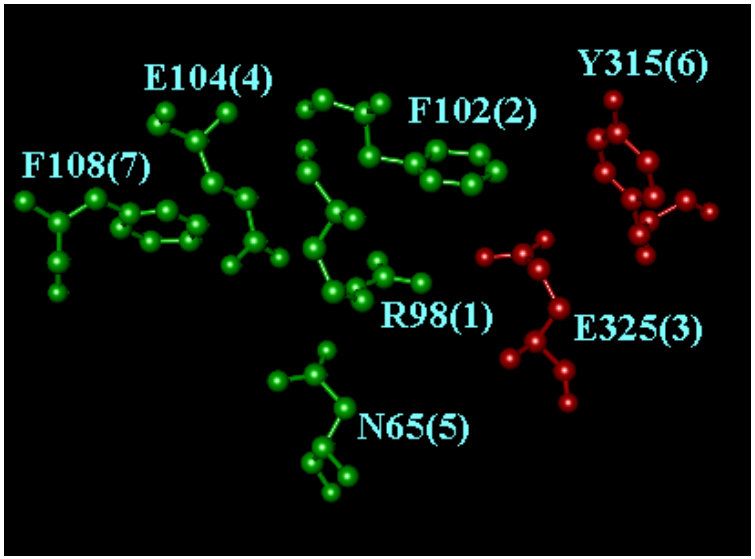
Clusters (at interaction strength Imin =6%)

in Ornithine Decarboxylase



Centre of a Graph

$$E(v) = \max_{v_i \in G} d(v, v_i)$$



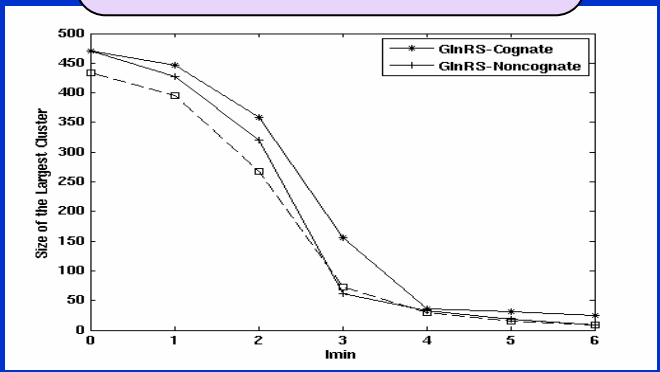
Construction of Graphs and Networks

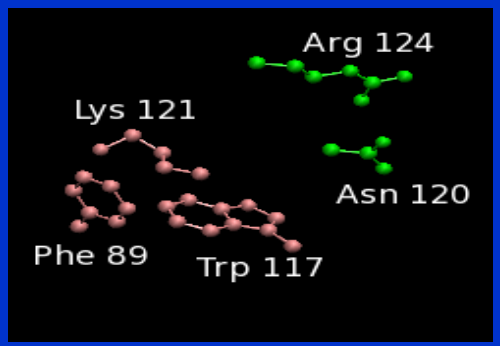
Interacting biomolecular residues

Adjacency Matrix

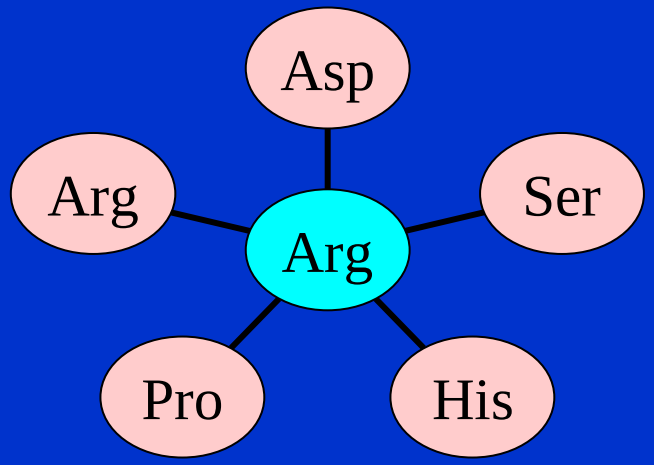
PSN
Size of largest cluster,
Degree distribution, Hubs

PSG
Side chain Clusters
(DFS)





Hubs in Protein Structure Networks



- Hubs - highly connected amino acids in the protein structure.
- Identified as amino acids having a contact number of >= **4** at a given I_{min}.

Applications of Graph Spectral Analysis to Protein structures

A) Clusters of Importance:

a) Active/Binding site

b) Domain identification

c) Determinant of thermal stability-Aromatic clusters

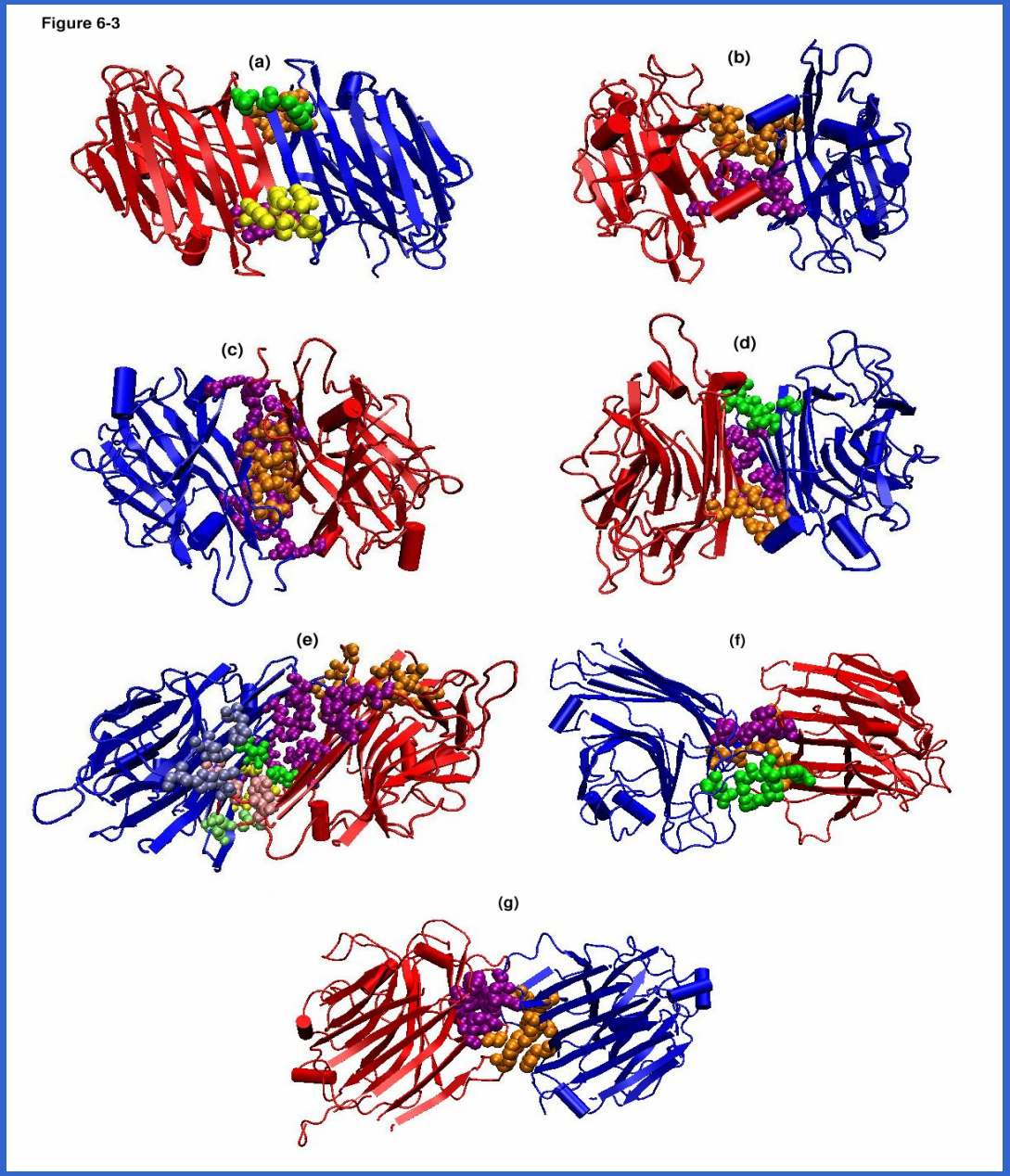
d) Protein-Protein interaction surfaces

B) Cluster centers:

Identification of hot spots, Signature motifs of oligomerization

A case study

Oligomerization in Legume lectins



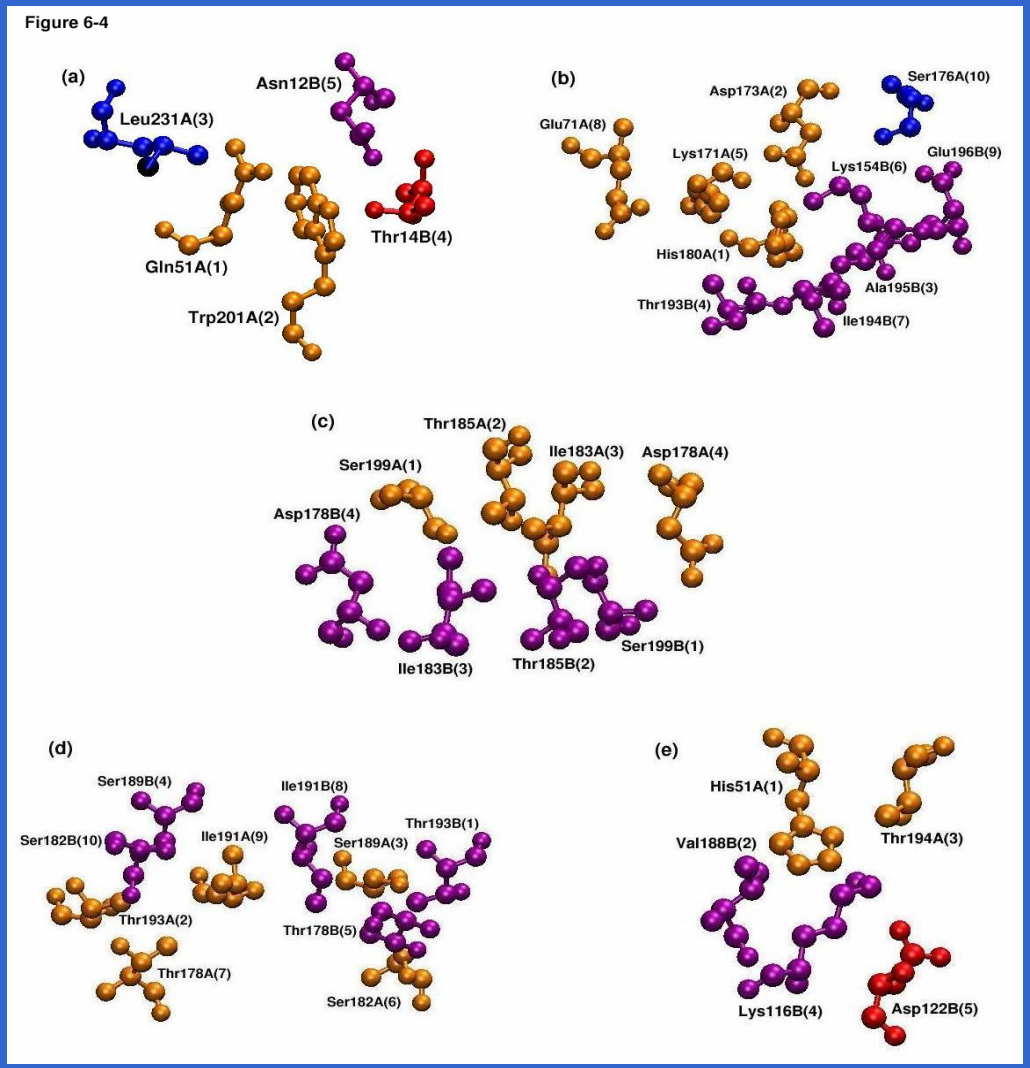
Legume Lectins:

High sequence similarity

Similar 3-D structures

Diverse Quaternary Associations

Interface clusters at different types of dimeric interfaces

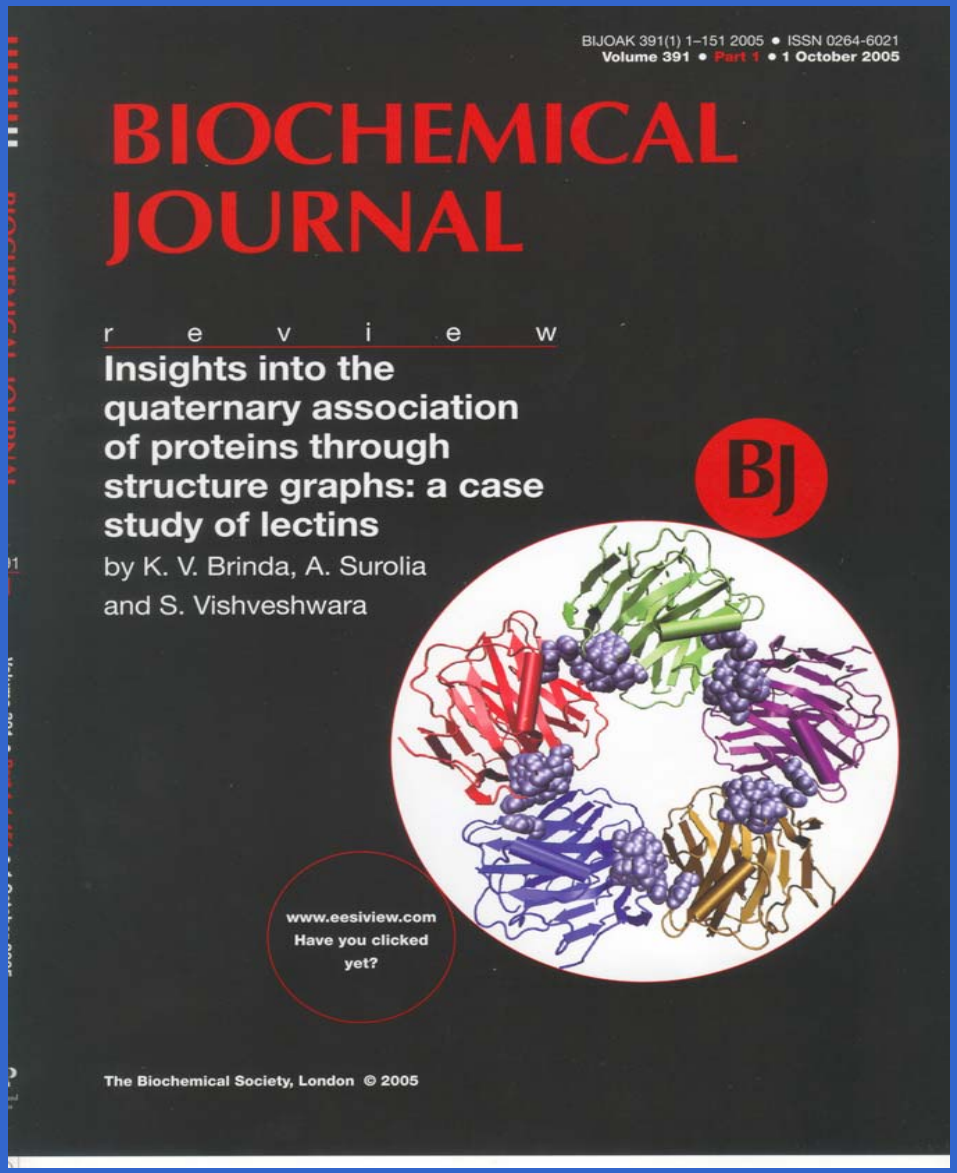


Interface clusters

Different colors represent residues from different chains (note: The residues are not sequential and hence the signature of association type can not be obtained from sequence analysis)

Cluster Center

The residue position in the cluster (interface graph) is obtained from graph spectral analysis. The residues with highest rank are the cluster centers (hot spots). They may be the targets for mutational experiments

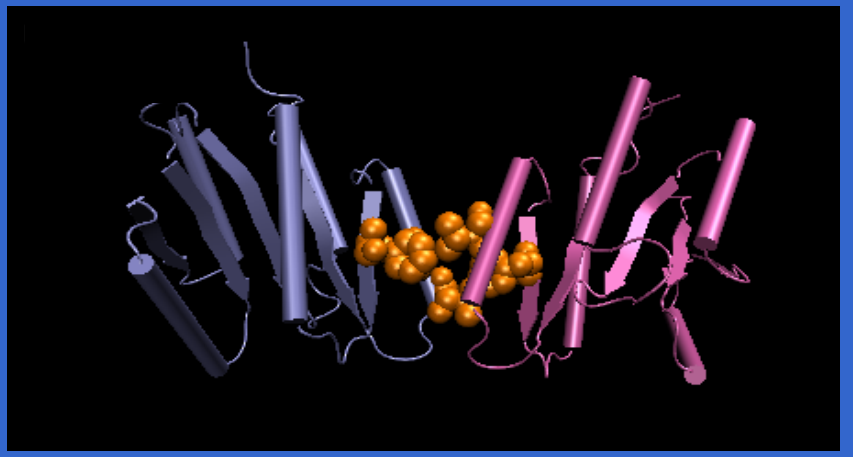


Pentraxin is yet another quaternary structure of a lectin

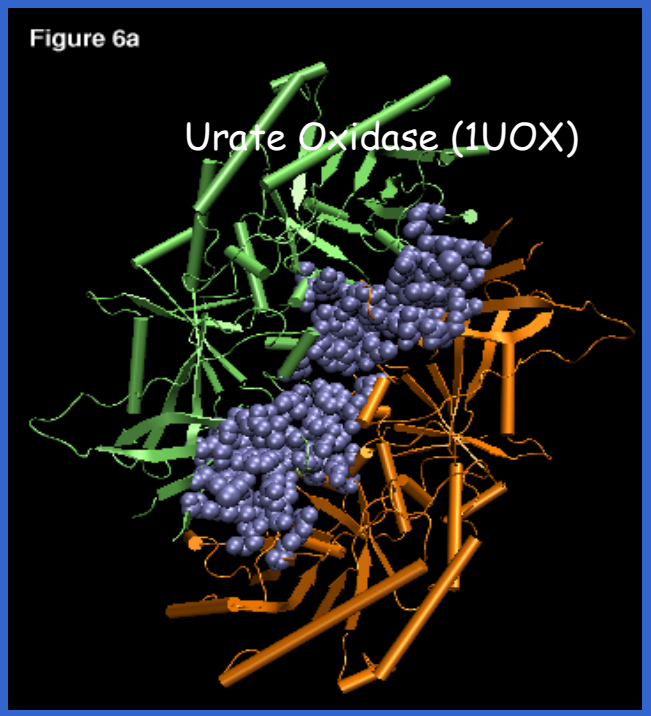
Protein-Protein Interaction Networks

(Analysis of a large dataset)

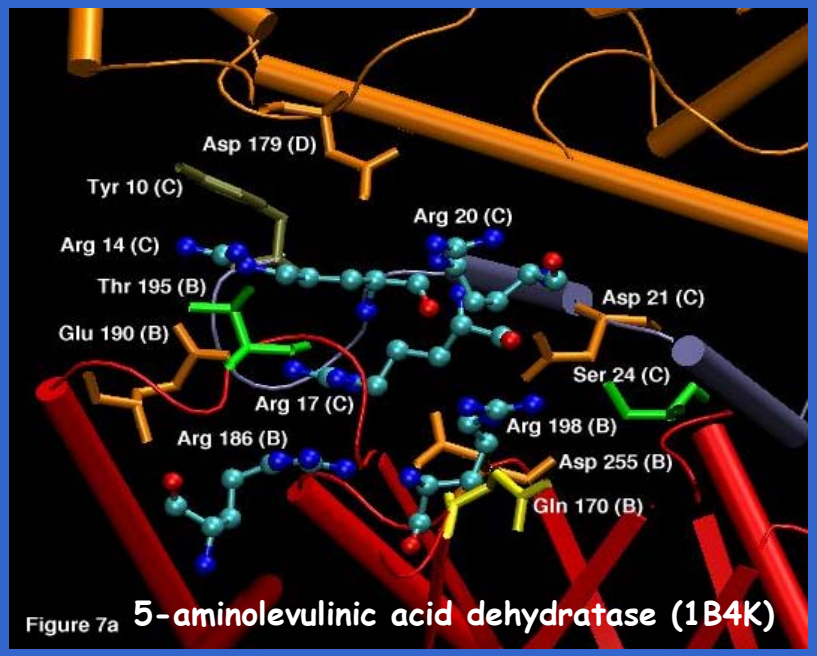
Weak Interface



Strong Interface



Centre of interface cluster



Brinda, Vishveshwara, *BMC Bioinformatics*, 2005

References on Protein structure networks

Three key residues form a critical contact network in a protein folding transition state.
Nature. 409:641-645, Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2001.

Small-world view of the amino acids that play a key role in protein folding.
Phys. Rev. E. 65:061910 Vendruscolo, M., N. V. Dokholyan, E. Paci, and M. Karplus. 2002.

Topological determinants of protein folding.
Proc. Natl. Acad. Sci. USA. 99:8637-8641,Dokholyan, N. V., L. Li, F. Ding, and E. I. Shakhnovich. 2002.

Uncovering network systems within protein structures.
J. Mol. Biol. 334:781-791. Greene, L. H., and V. A. Higman. 2003.

Network analysis of protein structures identifies functional residues.
J. Mol. Biol. 344:1135-1146 Amitai, G., A. Shemesh, E. Sitbon, M. Shklar, D. Netanely, I. Venger, and S. Pietrokovski. 2004.

Small-world communication of residues and significance for protein dynamics.
Biophys. J.86:85-91.Atilgan, A. R., P. Akan, and C. Baysal. 2004.

Network properties of protein structures.
Physica A. 346:27-33. Bagler, G., and S. Sinha. 2005.

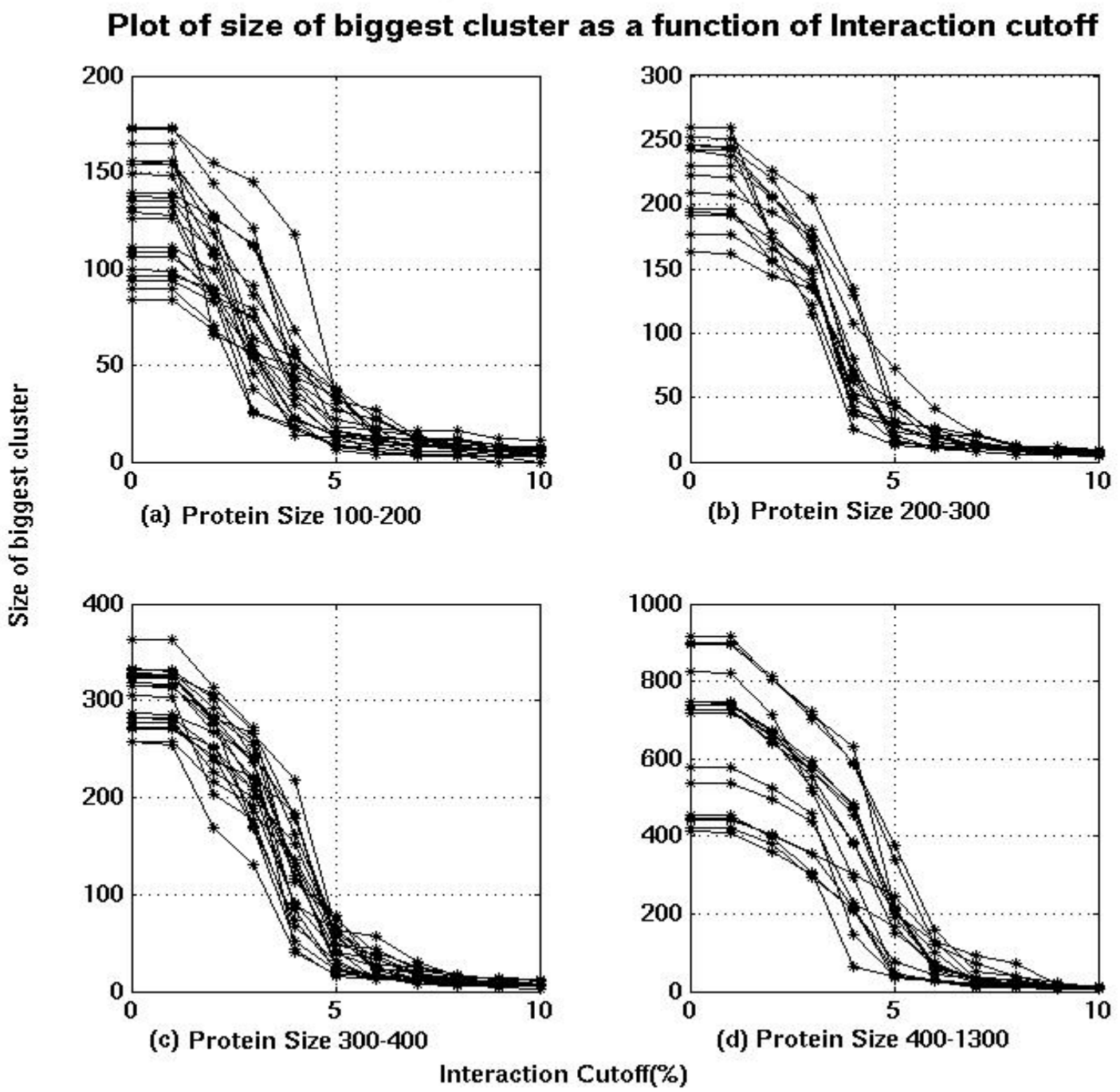
A Network Represenataion of Protein Structures: Implications for Protein Stability,
Biophysical Journal, 6, 296, Brinda K.V. and Vishveshwara.S, 2005

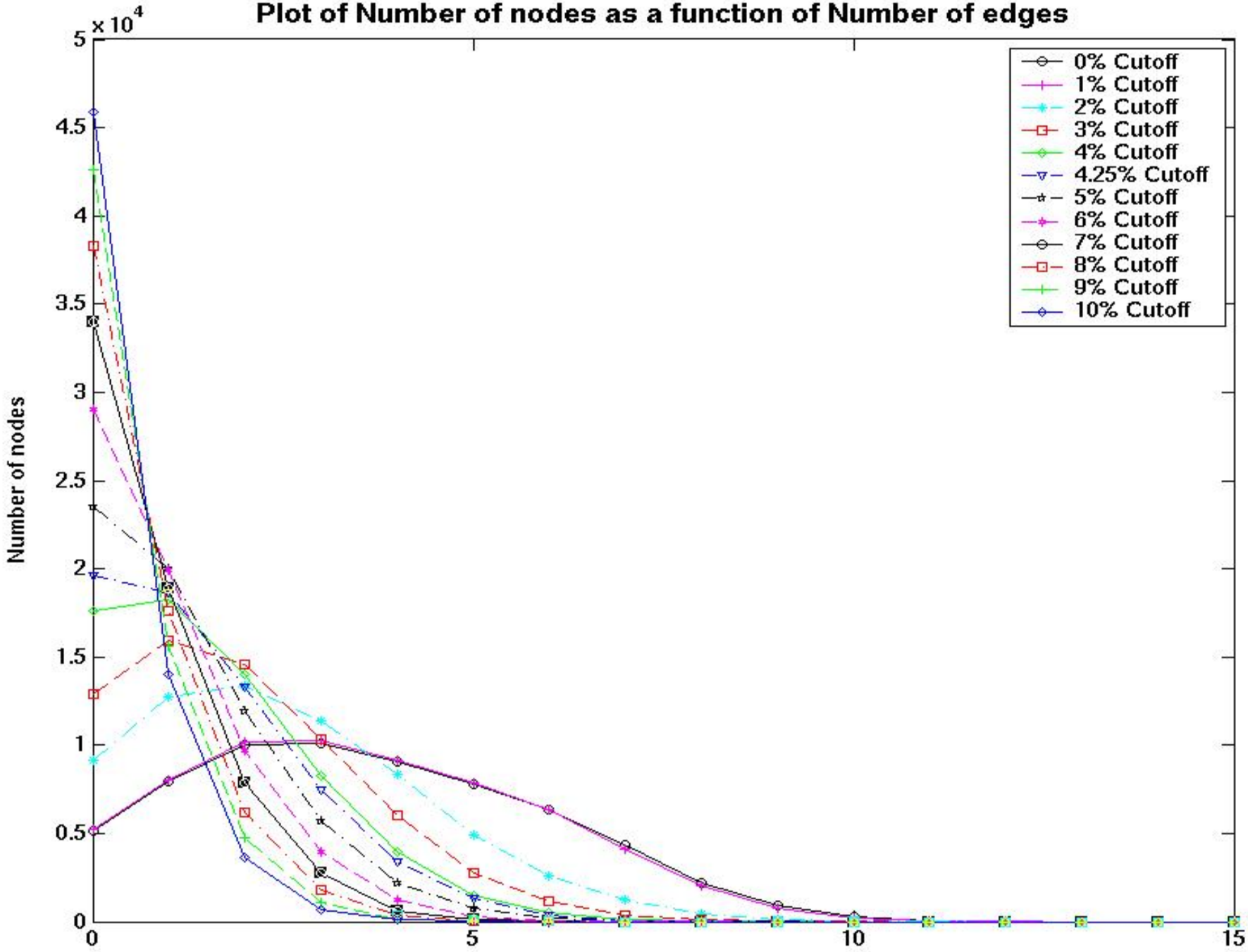
Dynamics of Lysozyme Structure Network: probing the Processes of Unfolding
Biophysical Journal 92, Ghosh Amit, Brinda.K.V., Saraswathi Vishveshwara. , April 1, 2007

Protein Structure Graph and Network Topology

Investigation of a large number of non-redundant structures

Brinda and S. Vishveshwara, *Biophysical Journal*, Dec 2005


Plot of size of biggest cluster as a function of Interaction cutoff
Size of biggest cluster
Interaction Cutoff(%)
(a) Protein Size 100-200
(b) Protein Size 200-300
(c) Protein Size 300-400
(d) Protein Size 400-1300


Plot of Number of nodes as a function of Number of edges
x 10^4
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Number of nodes
0
5
10
15
Number of nodes with 'k' links
0% Cutoff
1% Cutoff
2% Cutoff
3% Cutoff
4% Cutoff
4.25% Cutoff
5% Cutoff
6% Cutoff
7% Cutoff
8% Cutoff
9% Cutoff
10% Cutoff

The Protein network topology exhibits a complex behaviour

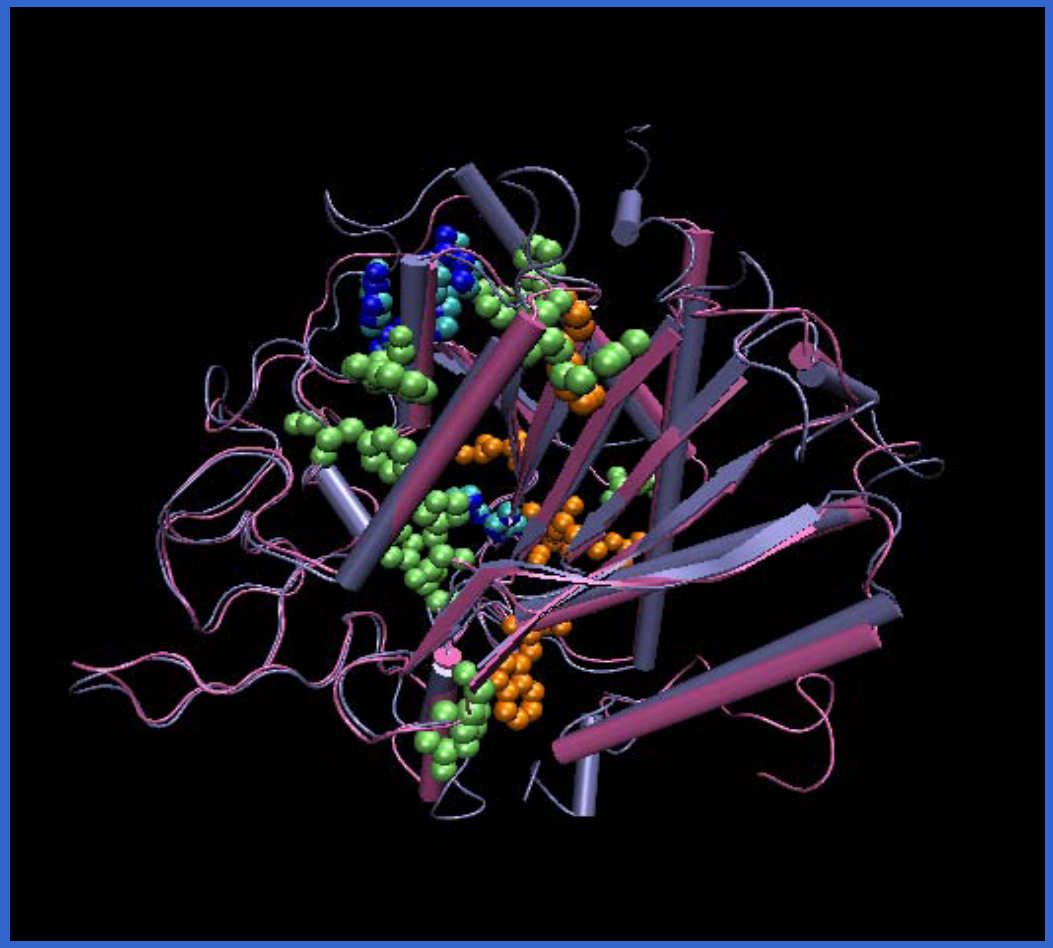
Poisson distribution?

Scale free- power law?

The topology depends on the Interaction strength (Imin)

- Poisson Distribution when all the weak interactions are considered
- Exponential-like behaviour at Interaction strengths greater than Icritical

Hubs and Thermal Stability of Proteins



Hubs in carboxy peptidase of a thermophilic and a mesophilic organism

Blue: hubs common to both proteins

Green: Hubs exclusive to the thermophilic

Orange: Hubs exclusive to the mesophelic

Protein Structure Networks: Two proteins performing the same function at different temperatures. Although the overall structures are very much similar, the additional stability of the protein functioning at high temperature is due to increased number of hubs (shown in green). The concept is useful in protein design

Brinda and S.Vishveshwara, *Biophysical journal* , Dec 2005

Dynamics of structure networks

Concept of Structure Network Integrated with Dynamics

Equilibrium and unfolding simulations of T4-Lysozyme

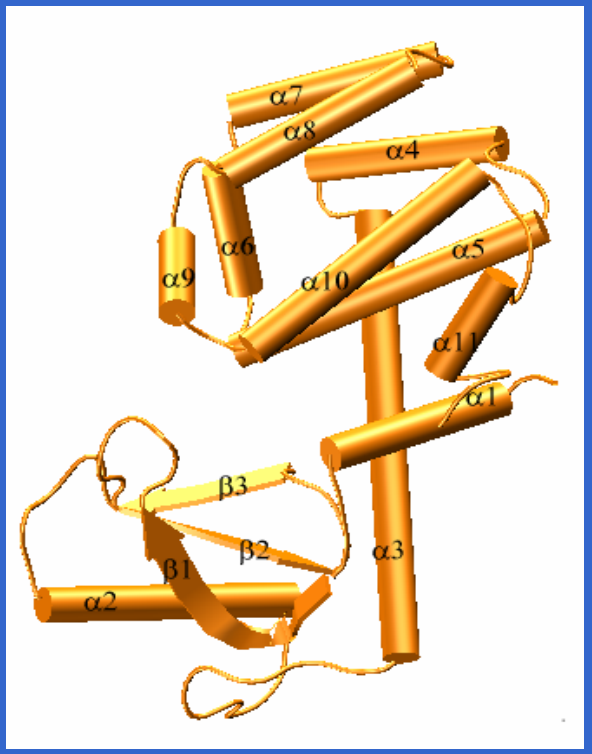
probed from Network perspective

Protein Folding / unfolding

Analysis of the dynamics trajectories

Amit Ghosh, Brinda, Vishveshwara,

Biophysical J **April 2007**

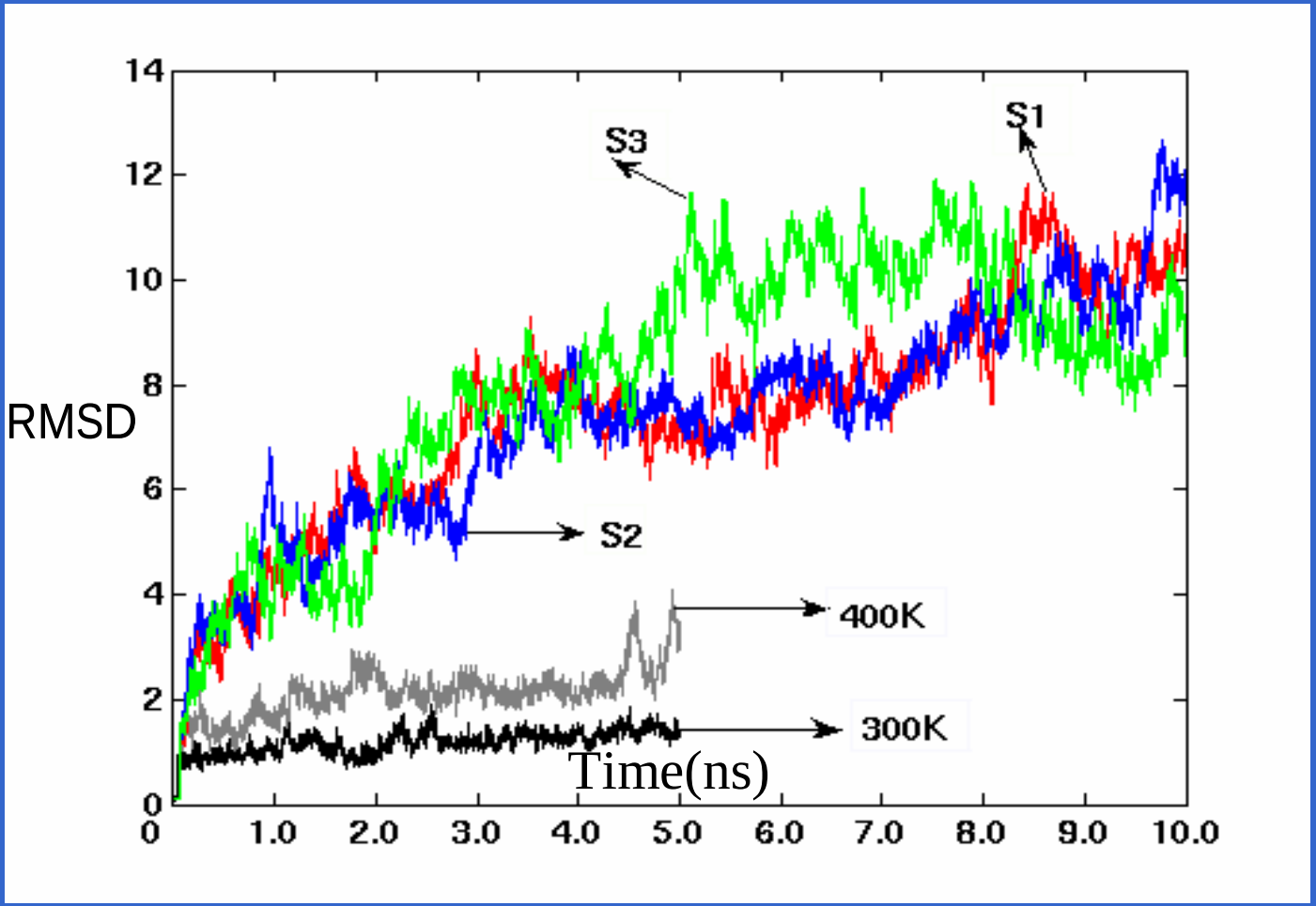


T4 Lysozyme

Simulation details

Trajectory no.	Temperature	Trajectory length (ns)	No of water's added	Density of water (gm/cc)
I	300K	5.0	7854	1.010
II	400K	5.0	7854	0.908
S1	500K	10.0	7854	0.730
S2	500K	10.0	7854	0.728
S3	500K	10.0	9593	0.717

Simulation Profiles: Root Mean Square Deviation(RMSD)

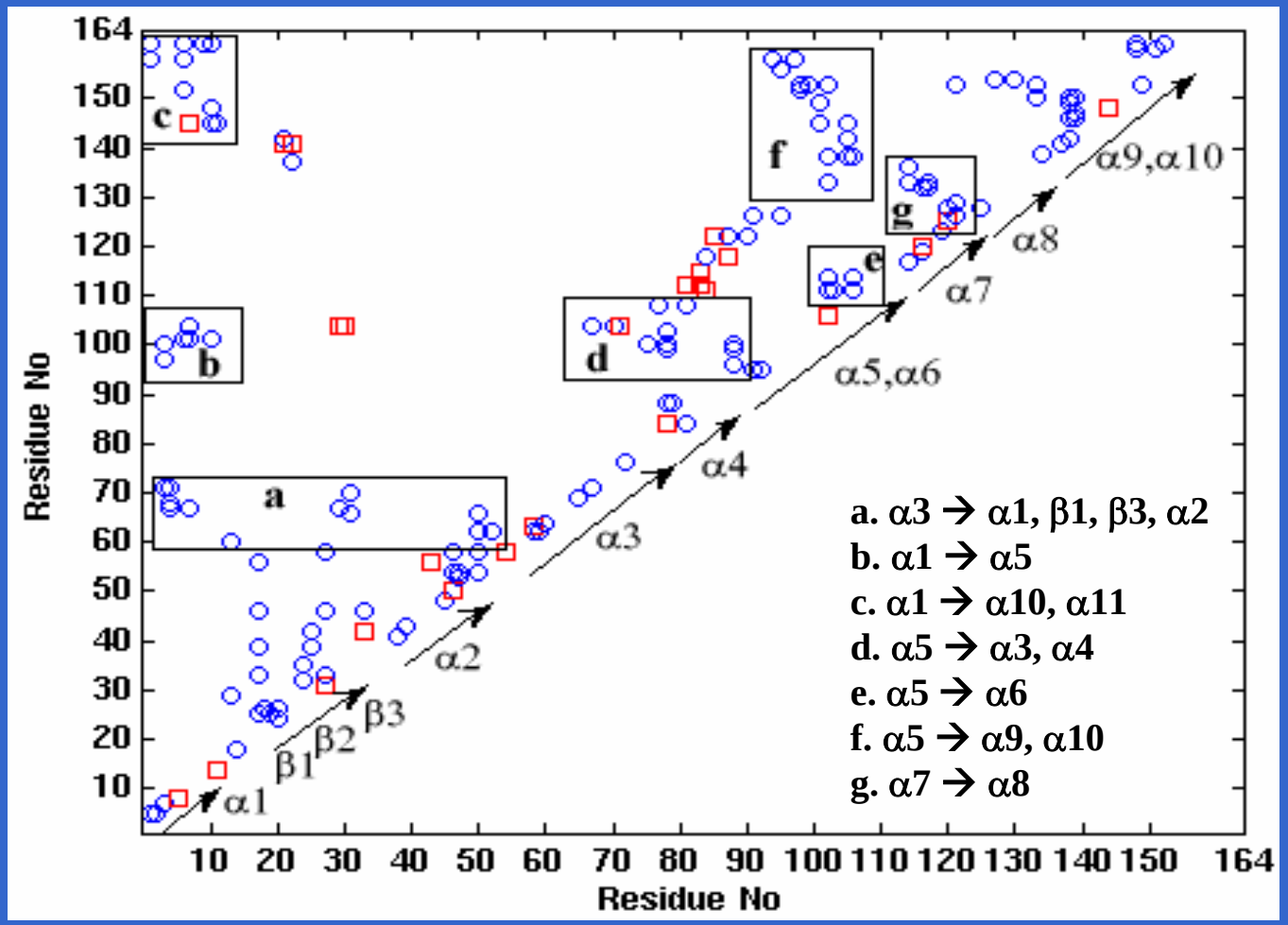


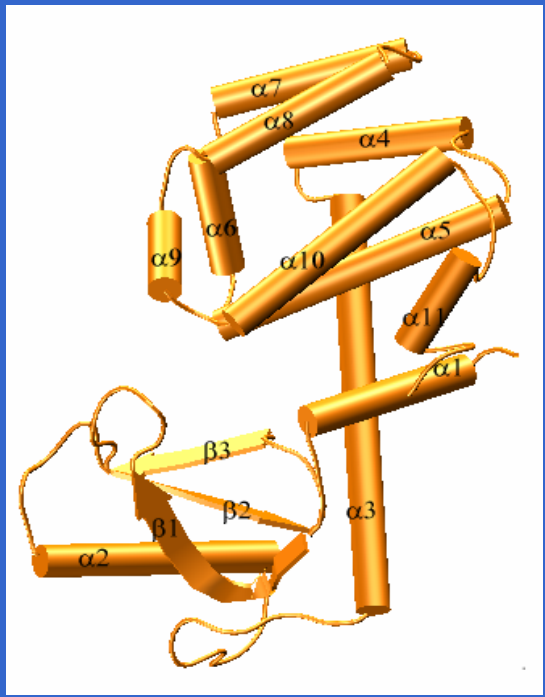
Interactions across secondary structures (300K)

Non-native contact compared crystal structure

o Native contact retained

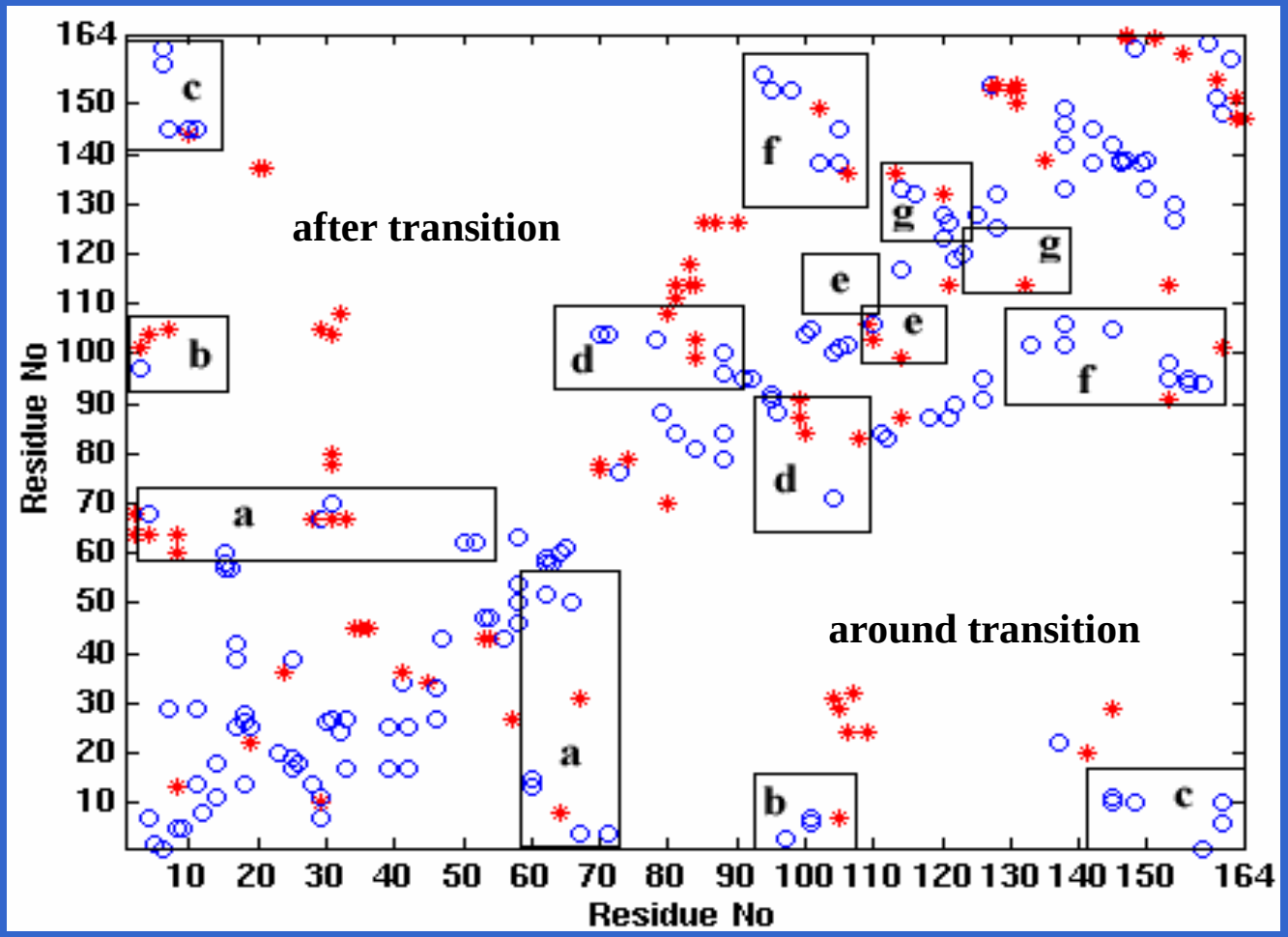
300K Imin=3.4%

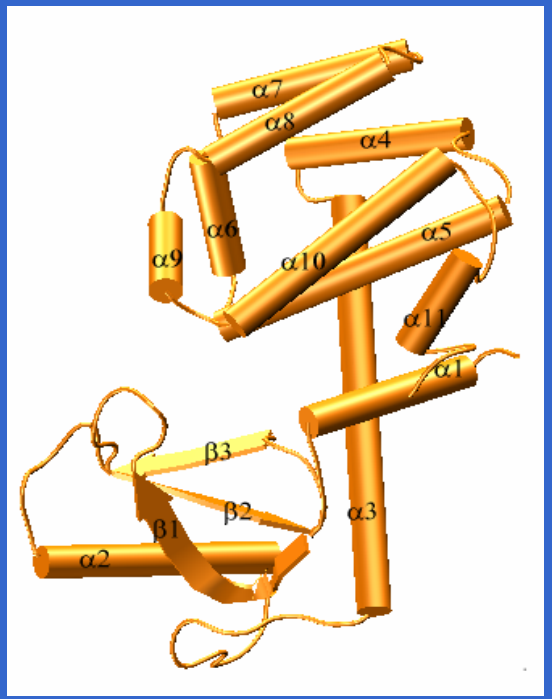




Interactions across secondary structures (500K)

500K Imin=2.5%

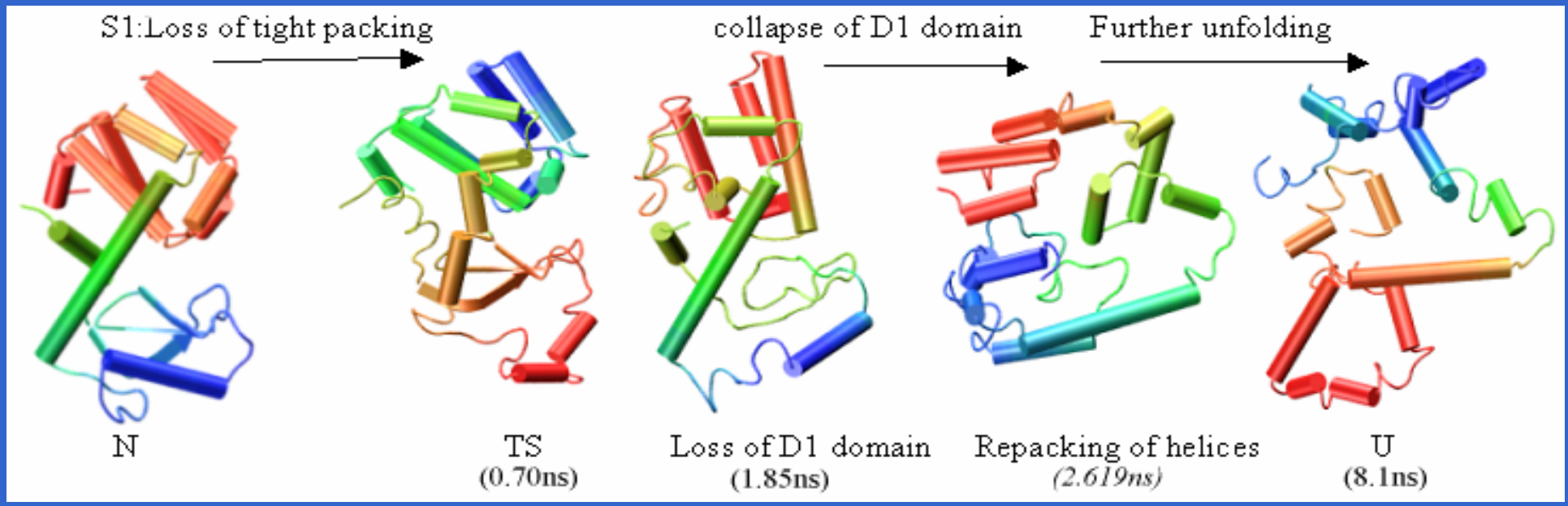




* Non-native contact compared with 300K simulation

o Native contact retained

Unfolding events

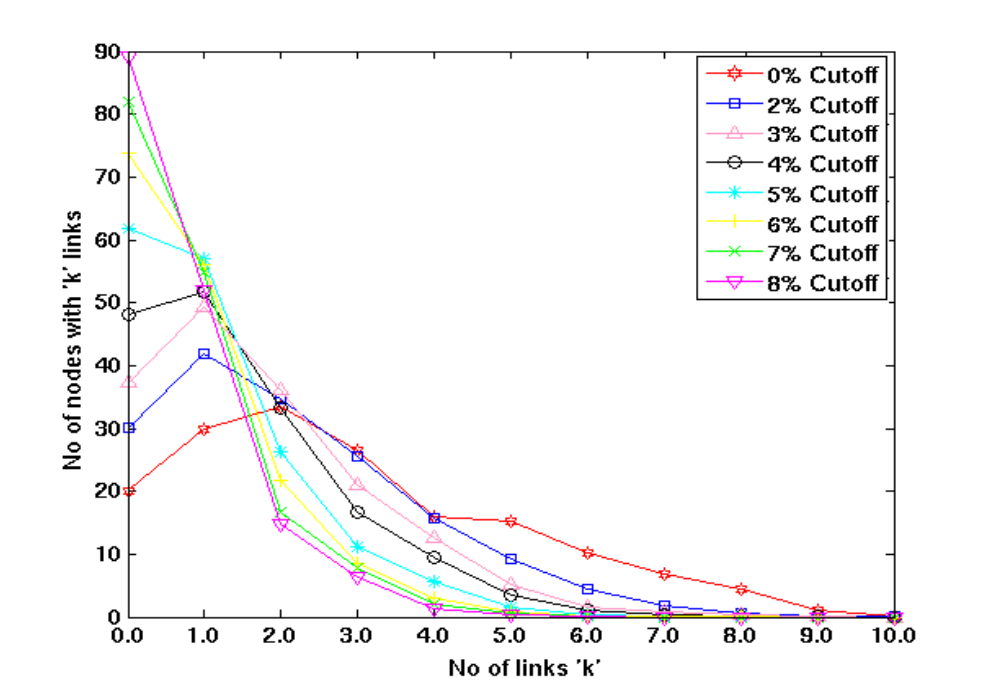


Analysis of Network Parameters

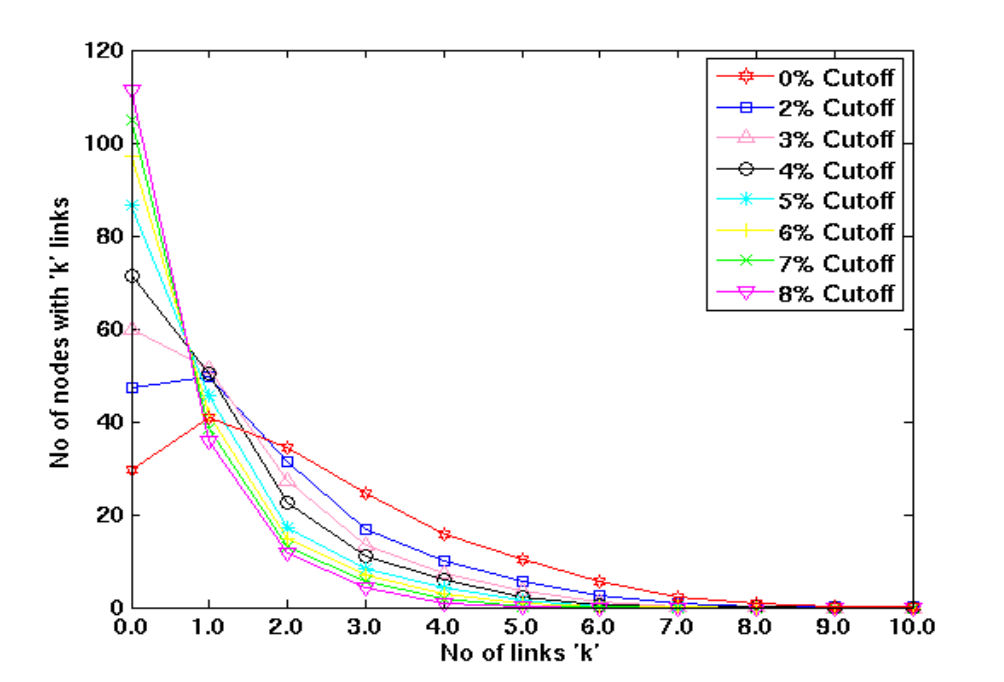
Side-Chain Interaction Strength Dependent Analysis

- Degree distribution profiles
- Largest cluster profiles
- Native/Non-native contacts as a function of simulation time

Distribution of nodes with k links (**degree distribution**)



300K



500K

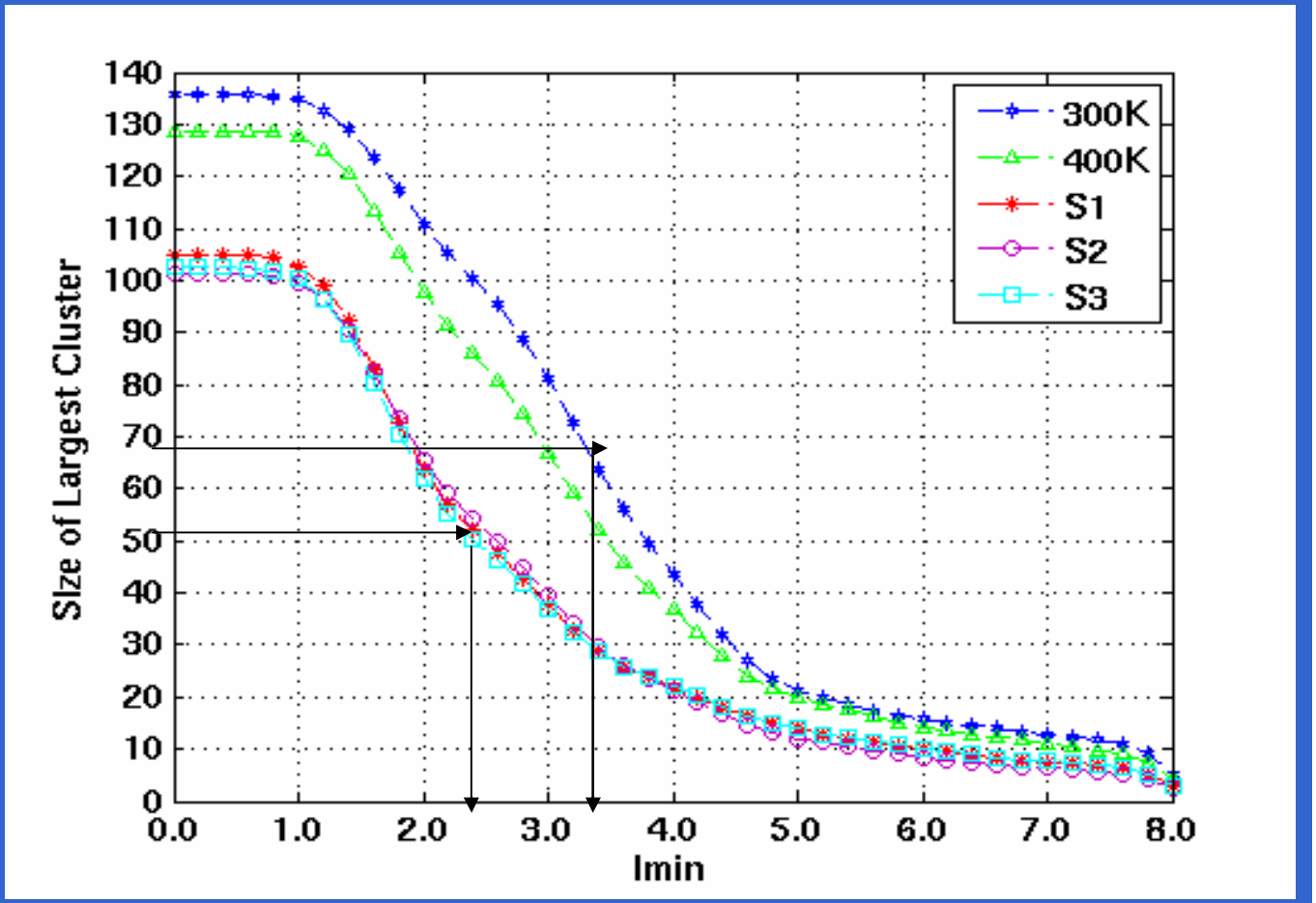
At 300K the number of nodes with one or two links is higher than the number of nodes with zero link(orphans), for I_{min} values less than 5%.

The transition from bell shaped curve to a decay-like curve takes place at a lower Imin of 3% in 500K

Size of the Largest Cluster Profiles

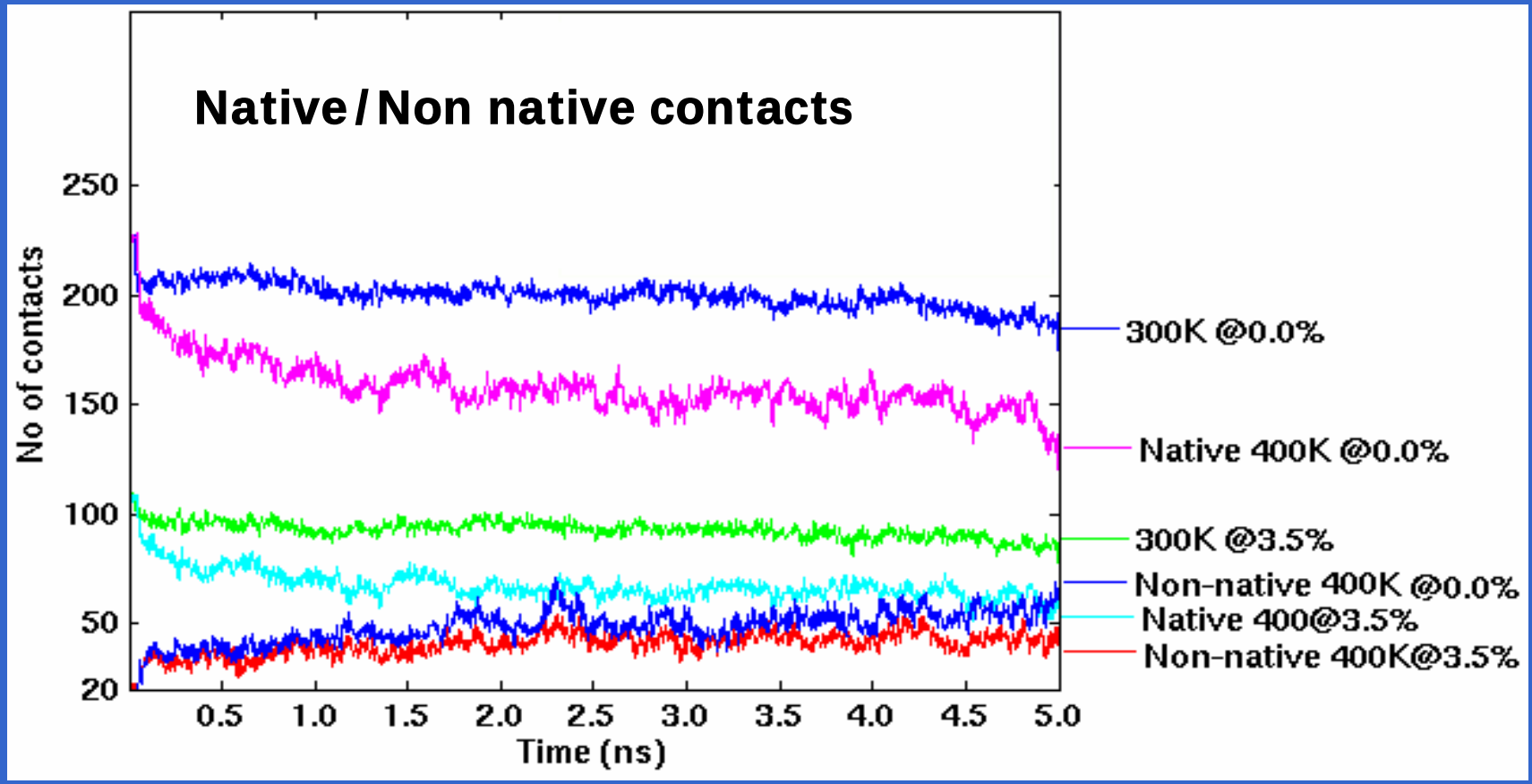
Size of the largest cluster at Imin=0% (averaged over the simulation)

Large (135) at 300K
Small (104) at 500K



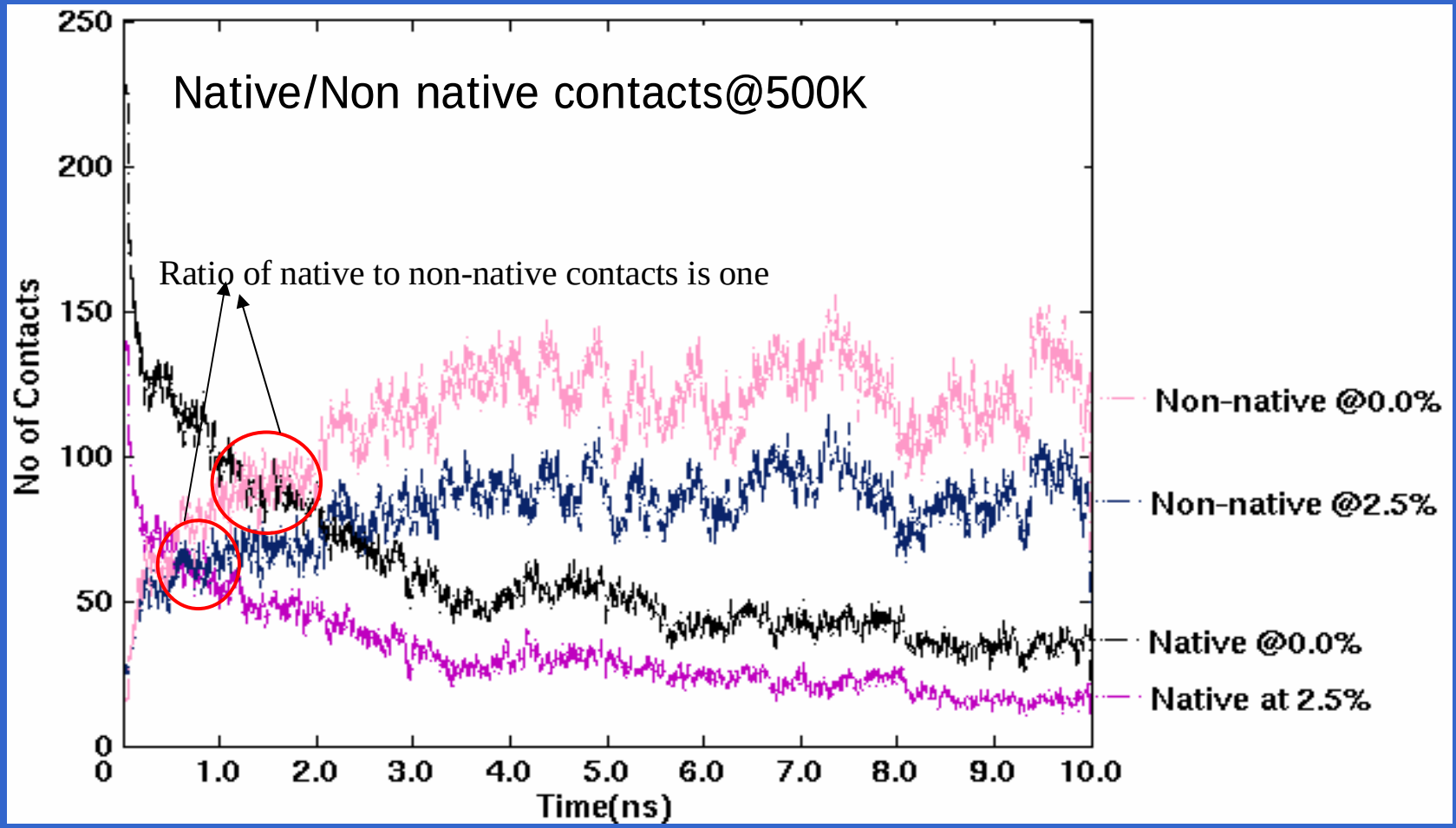
The size of the largest cluster in each of the simulations undergoes a transition at Imin = $I_{critical}$. (the size of the largest cluster is half of the maximum)

$I_{critical}$ shifts from 3.4% at 300K to 2.5% at 500K.



The average number of contacts

	300K	400K
native (0%)	206	160
native (3.5%)	98	72
non-native(0%)		50
non-native(3.5%)		45



native/non-native contacts = 1 at approximate time points 2 ns and 0.7 ns respectively for Imin values 0% and 2.5%.

The non-native contacts increase after these time points

Size and the composition of the large clusters

- **300K simulation**:

At **Imin = $I_{critical}$**, the largest cluster encompasses the residues of both the domains

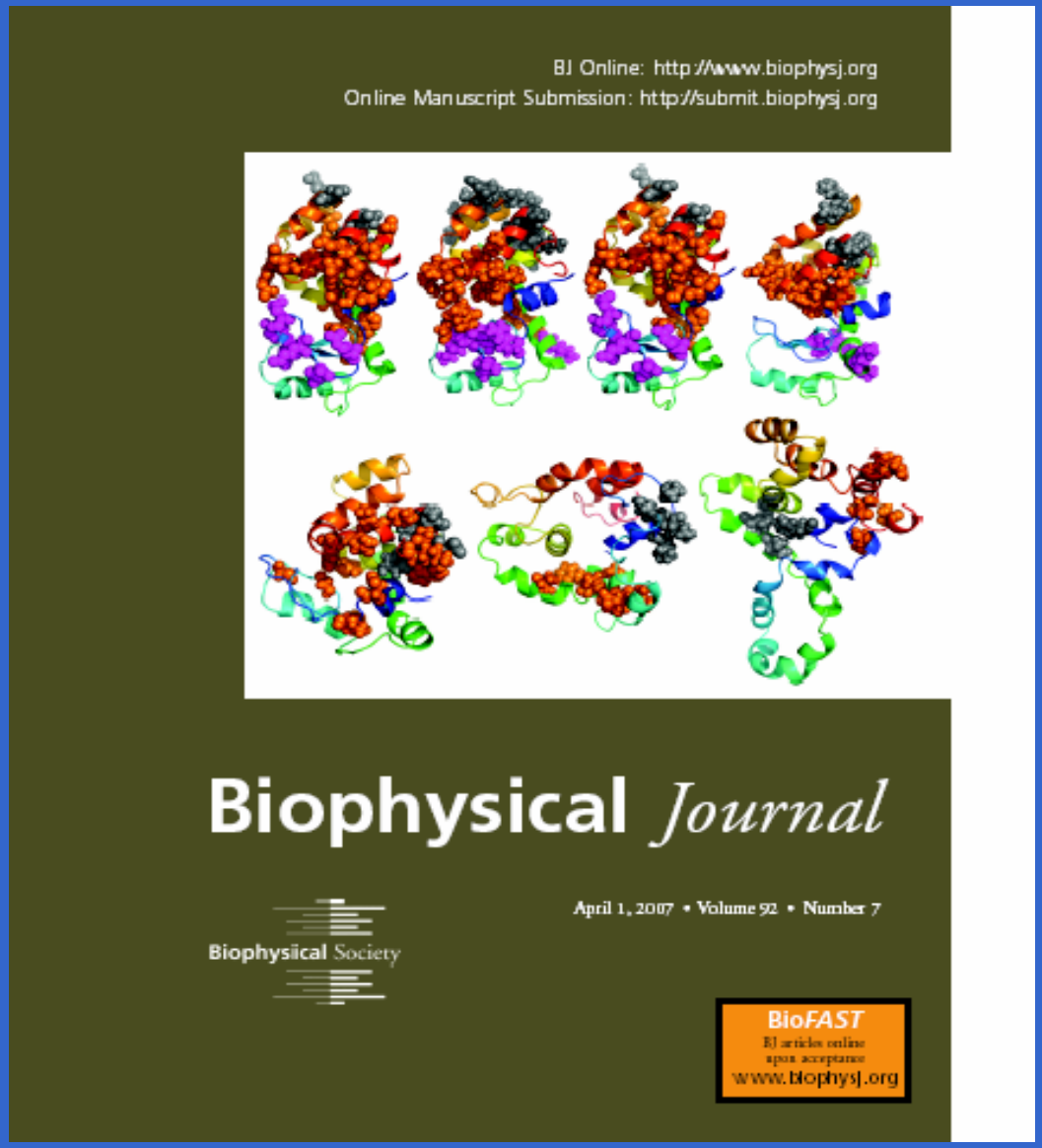
Average cluster composition (residues present in the largest cluster in >50% snapshots) is fairly constant. The largest cluster includes a considerable amount of hydrophobic residues

At **Imin > $I_{critical}$** , the top 2 largest clusters belong to the larger domain and the 3rd largest cluster belongs to the smaller domain. The participation of hydrophobic residues reduce considerably and aromatic residues dominate

- **500K simulation**

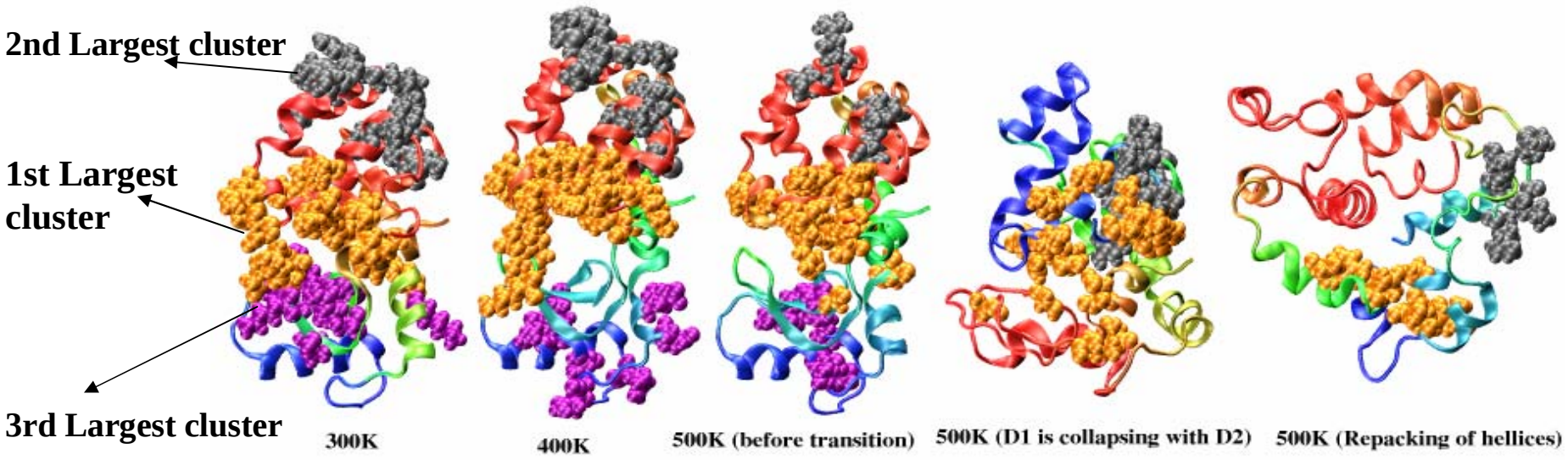
Large fluctuations in the size and the composition of the clusters.

Different unfolding states such as the transition state, collapse state, unfolded state can be recognized from the size and the composition of the top large cluster



Lysozyme unfolding monitored through the changes in the size and the composition of large clusters formed by non-covalent interactions of amono acid side-chains

Composition of the top three large clusters (Imin=5%)



300K 4929ps 1st	300K 4929ps 2nd	300K 4929ps 3rd	400K 4177ps 1st	400K 4177ps 2nd	400K 4177ps 3rd	500K 495ps 1st	500K 495ps 2nd	500K 495ps 3rd	500K 1876ps 1st	500K 1876ps 2nd	500K 3000ps 1st	500K 3000ps 2nd
10D	88Y	14R	10D	88Y	33L	10D	88Y	27I	10D	88Y	45E	11E
20D	90S	17I	19K	91L	42A	21T	91L	42A	13L	90S	47D	13L
21T	91L	19K	21T	95R	45E	28G	95R	45E	23G	91L	51G	15L
23G	94V	24Y	24Y	97C	47D	70D	125R	47D	30G	94V	67F	17I
69Q	95R	26T	101N	119R	52R	104F	127D	53N	49A	95R	68N	26T
70D	115T	31H	104F	120M	55N	105Q	152T	57V	105Q	152T	74A	
101N	119R	32L	105Q	124K	61D	136	155T		138W	155T	46L	
104F	120M	65K	137R	125R	64E	137R			144N	159D		
105Q	124K		138W	127D		138W			147K	161Y		
137R	125R		141Q	153F		140N			151T			
138W	127D		143P	155 T		141Q			158W			
140N	153F		144N			144N			160A			
141Q	155T		145R			145R			162 K			
144N			147K			147K						
145R			148R			148R						
147K			162K			163N						

Correlations with Experiments

Mutation studies:

A large number of mutations and their effects on stability have been experimentally carried out.

The destabilizing mutations (Val111,Trp138, Phe153) have been identified as hubs from our study. We predict that the mutation of other residues (Phe104,Arg145,Arg148) will also destabilize the protein, since they are hubs at high Imin and also remain as hubs in 400K simulation

Stages of domain formation during folding:

Our high temperature simulation points to the fact that the domain D2 is formed at an early stage, reinforcing the experimental findings

Summary

Structure Networks

Protein structures are represented as graphs of non-covalent interactions. The network behaviour is dependent on the strength of interaction.

An optimal strength of interactions is present in all proteins which leads to a transition in the size of the largest cluster.

Structurally and functionally important clusters and hubs can be identified in the structure by choosing an appropriate strength of interaction

Dynamics of Structure networks

The network changes from simulation trajectories have been monitored.

The transition from the folded to unfolded state has been elucidated from the changes in the large clusters, chosen at suitable interaction strengths using the example of Lysozyme

The mutation and the domain formation experimental results have been correlated with network parameters

The communication pathways between the anticodon region and the amino acylation site have been deduced from the dynamic cross correlations and the network analysis of the MD trajectories of Methionyl tRNA Synthetase

Acknowledgements

Amit Ghosh: Lysozyme and MetRS

Dr. K.V. Brinda: structure network

Financial support: Department of Biotechnology,
Government of India