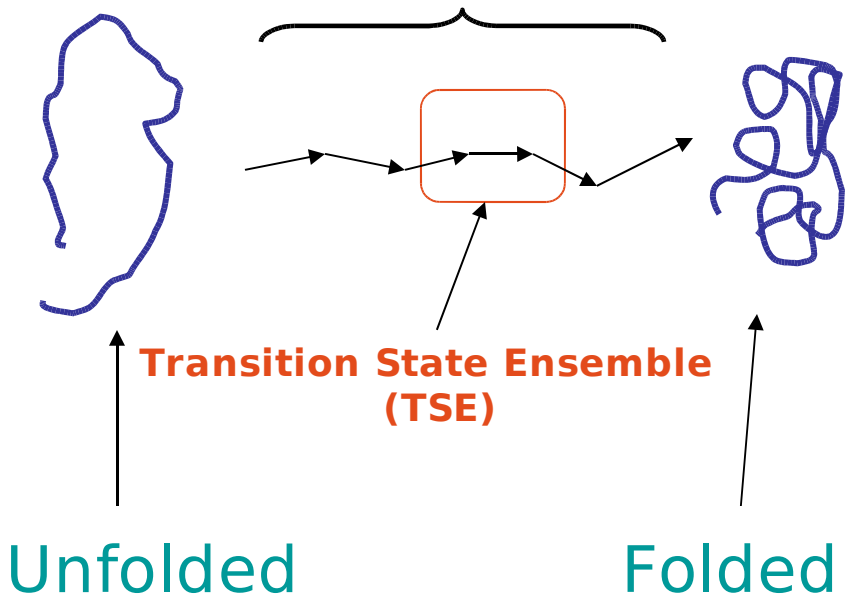
Statistical Physics of the Transition State Ensemble in Protein Folding

Alfonso Ramon Lam Ng,
Jose M. Borreguero,
Feng Ding,
Sergey V. Buldyrev,
Eugene Shakhnovich
and H. Eugene Stanley

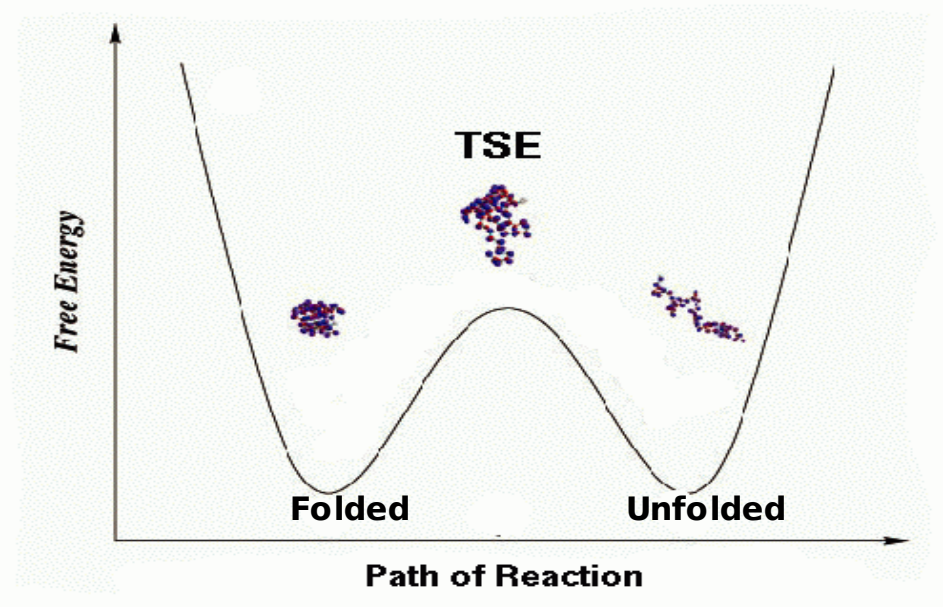
2004

The Protein Folding Problem

Protein Folding Process



At the Protein Folding Temperature T_F:



What is TSE?

Set of Protein Conformations located at the top of the free energy barrier that has the ***same probability*** *to fold or unfold*. (Du et al.,1998)

Purpose: *Characterize the structure of the TSE at the amino acid level.*

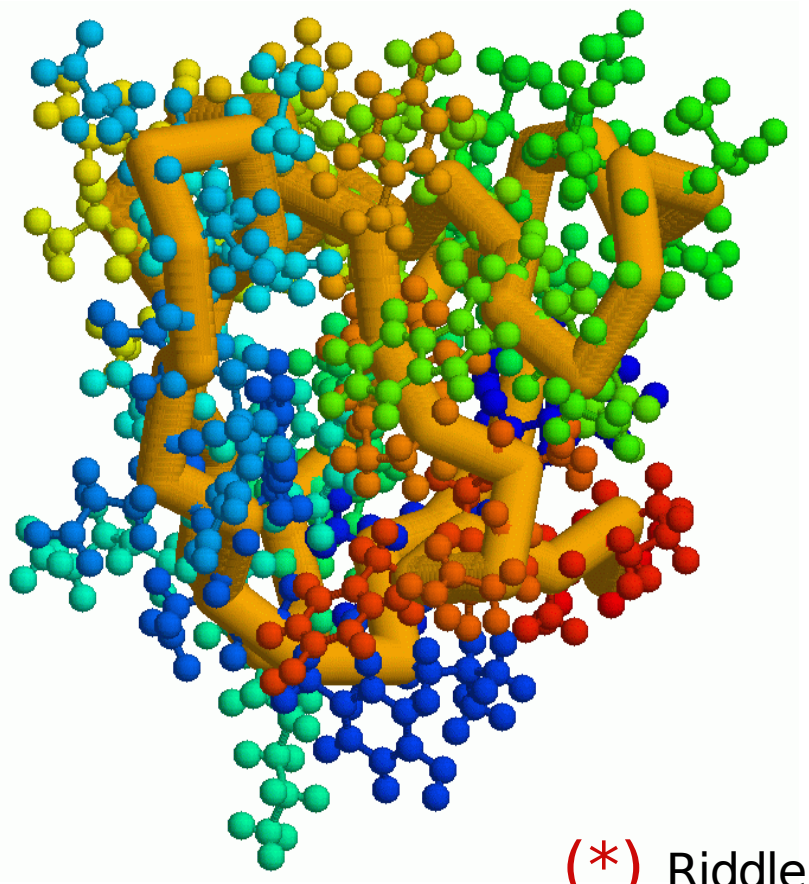
Why?: *Provide a guidance for amino acid substitution in experiments.(*)*

Our Results: *1.- The calculated TSE agrees with experimental data.*

2.- The TSE can be divided into a discre set of folding pathways.

(*): Martinez and Serrano, Nature Structural Biology,1999.
Riddle et al, Nature ,1999,
Gsponer et al, PNAS,2002,
Vendruscolo et al, Biophysical Journal.,2003

Our Protein: SH3



1.- Small Protein: 56 amino acids

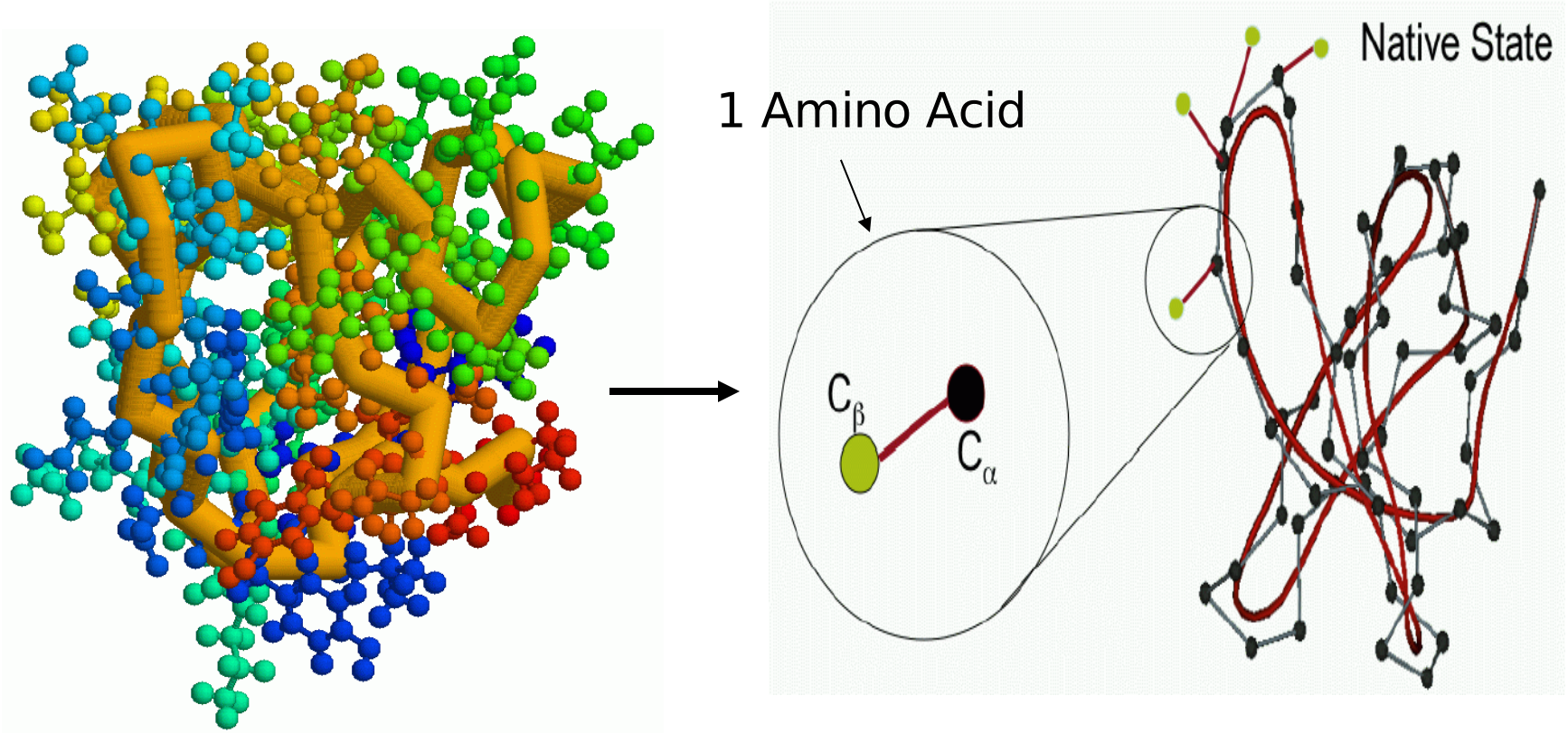
2.- Folds quickly.

3.- Well studied by experiments.

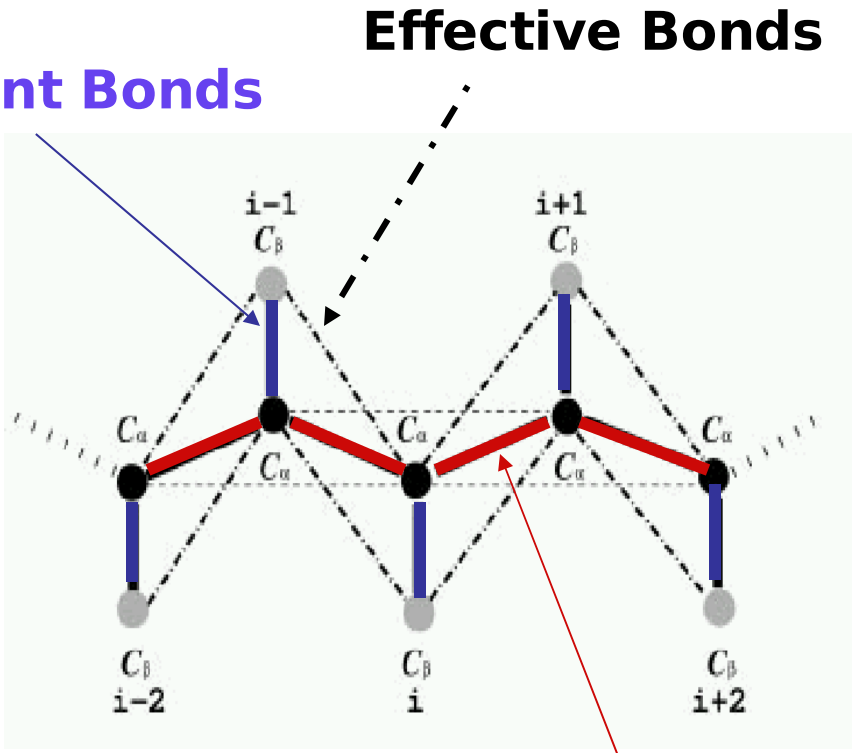
(*) Riddle et al, Nature ,1999.
Martinez and Serrano, Nature Struct. Bio.,1999.

The Two-Bead Representation

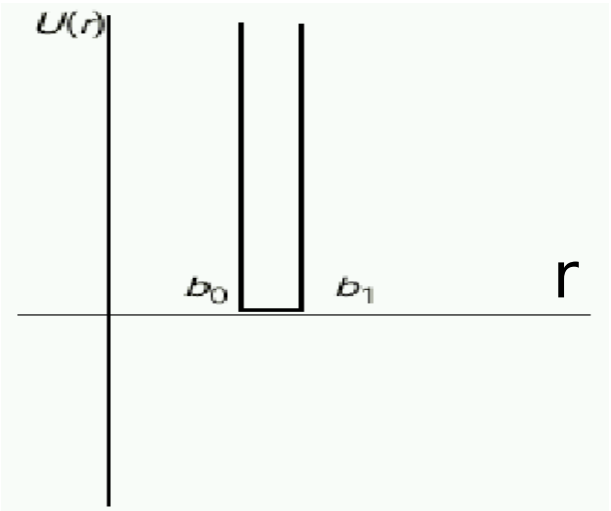
From its atomic structure to a coarse grained model



The Two-Bead Model



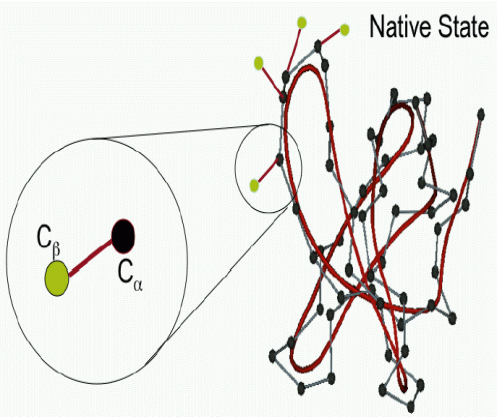
a) C_α - C_α and C_α - C_β Interactions:

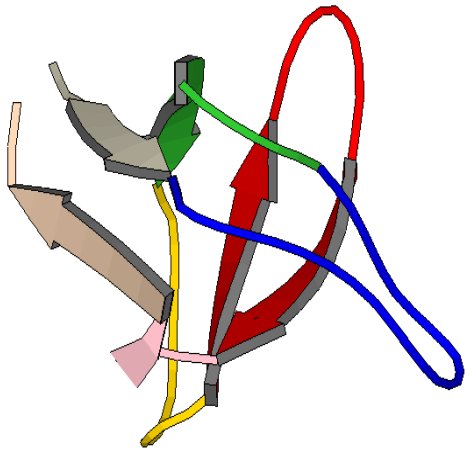


Bond	Bond length (Å)	Deviation (%)
$C_{\alpha i}$-$C_{\beta i}$	1.55	2.4
$C_{\alpha i}$-$C_{\alpha(i\pm1)}$	3.82	3.1

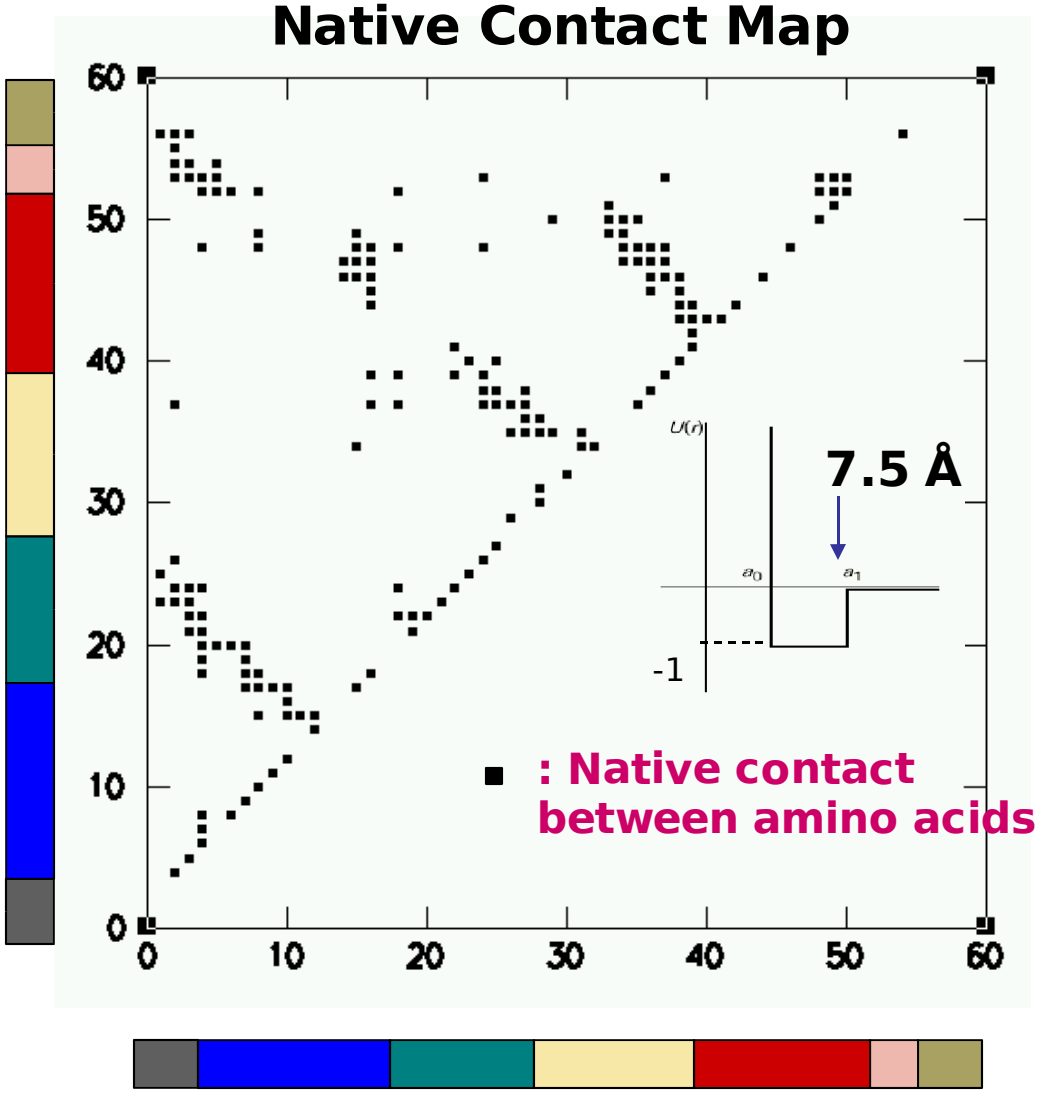
Feng Ding et al., Biophysical Journal,2001

b) C_β - C_β Interactions:





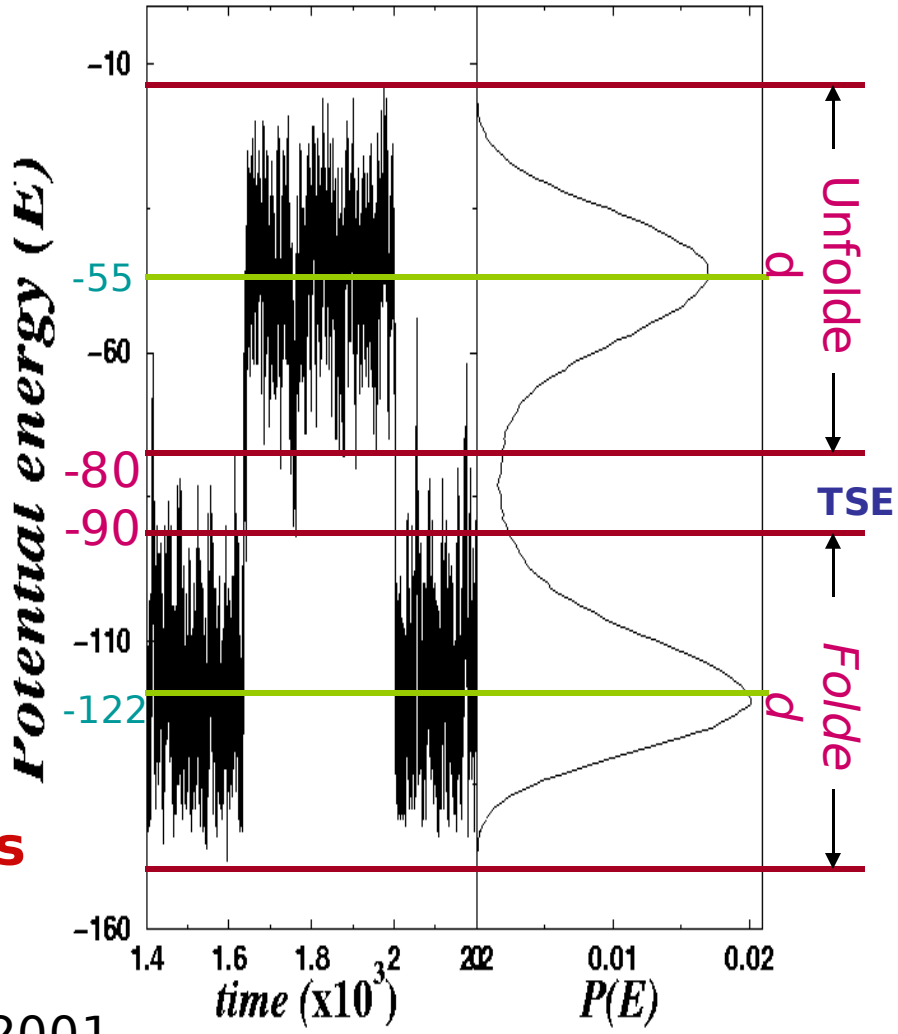
Cutoff = 7.5 Å



Obtaining the TSE

- *Run Long Time DMD Simulation at T_F=0.91.*
- *Select the Potential Energy as the approximate Path of Reaction.*
- *Energy window between -90 and -80 to locate the TSE (*).*
- *Record conformations within the energy window.*

5200 Candidate Conformations obtained.



(*) Feng Ding et al., Biophysical Journal,2001

Filtering the TSE Candidates

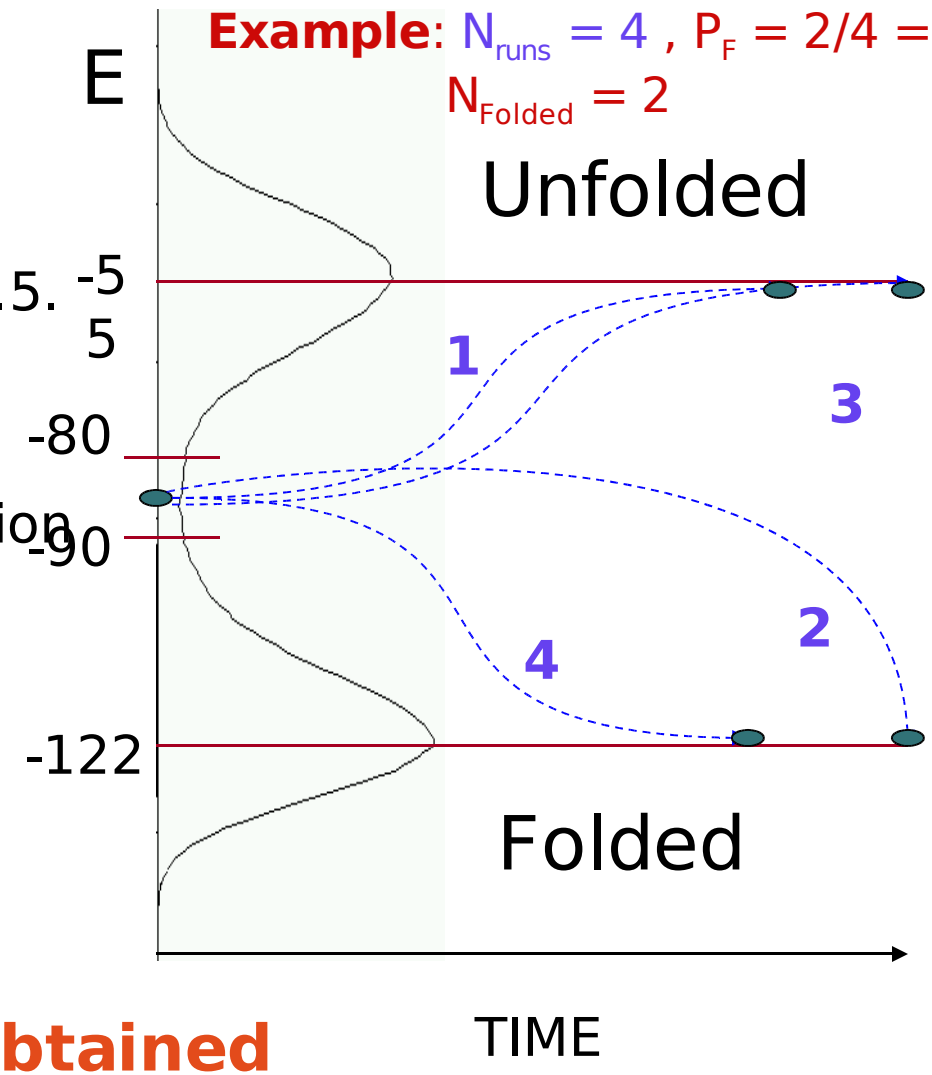
Calculating Folding Probability P_F

1.- TSE conformation has $P_F \sim 0.5$.

2.- Run N_{runs} of DMD simulations for each candidate conformation at T = 0.91.

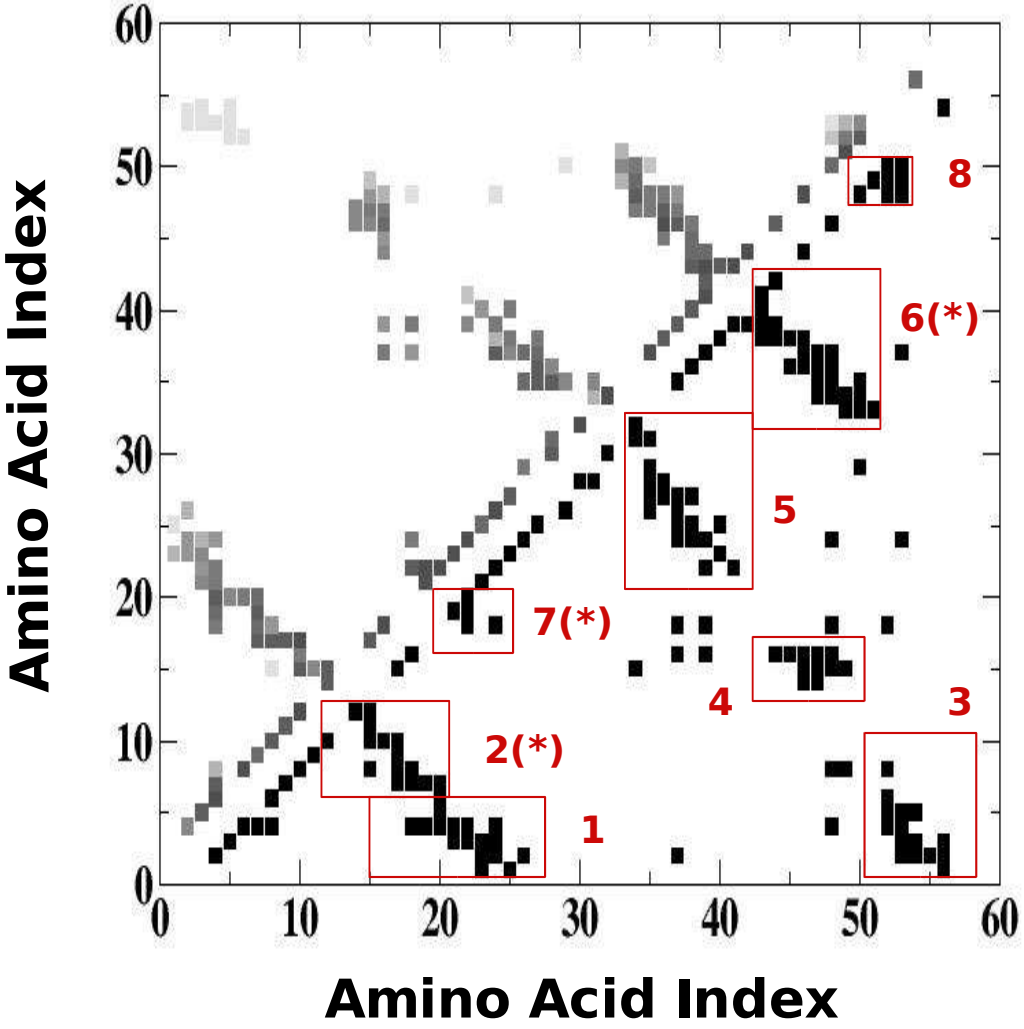
3.- Calculate P_F with:

$$P_F = N_{Folded}/N_{runs}$$

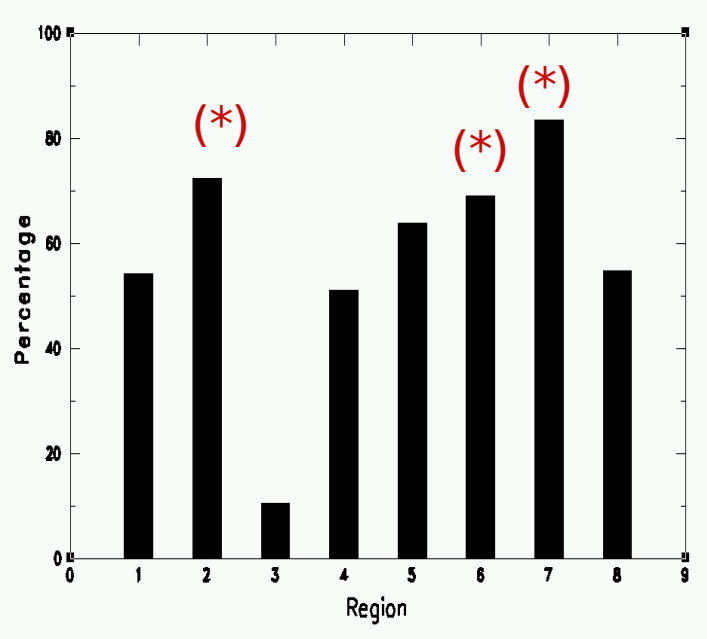


1525 Members of TSE obtained

The TSE Average Contact Map



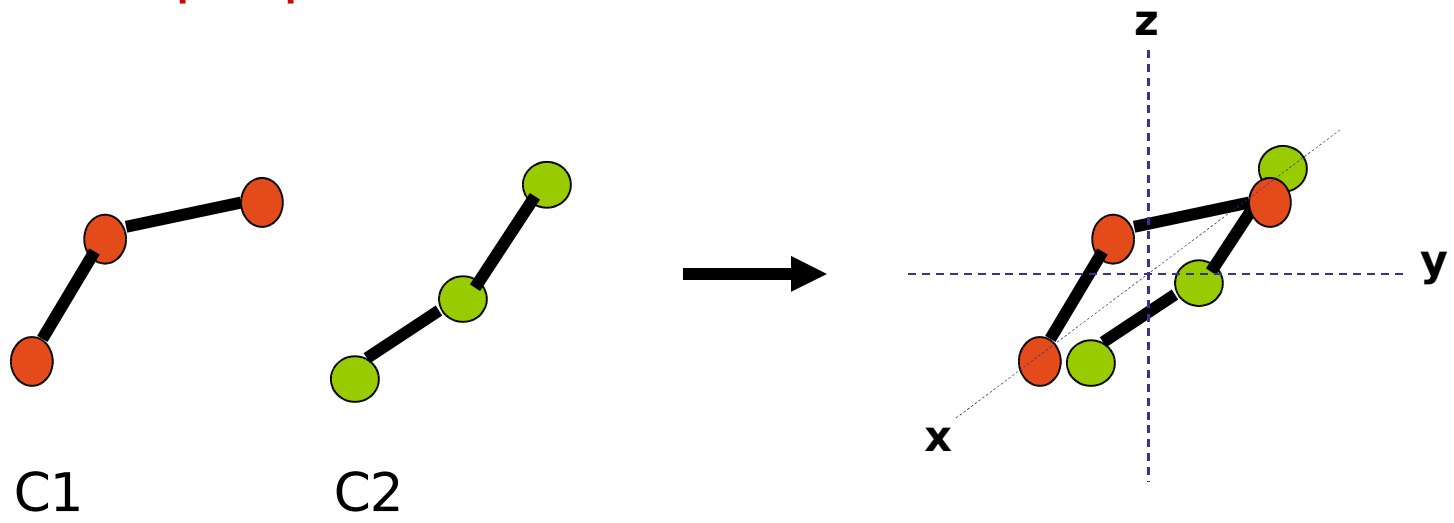
Frequency



(*): Martinez and Serrano, Nature Structural Biology,1999

Root Mean Square Distance (rmsd)

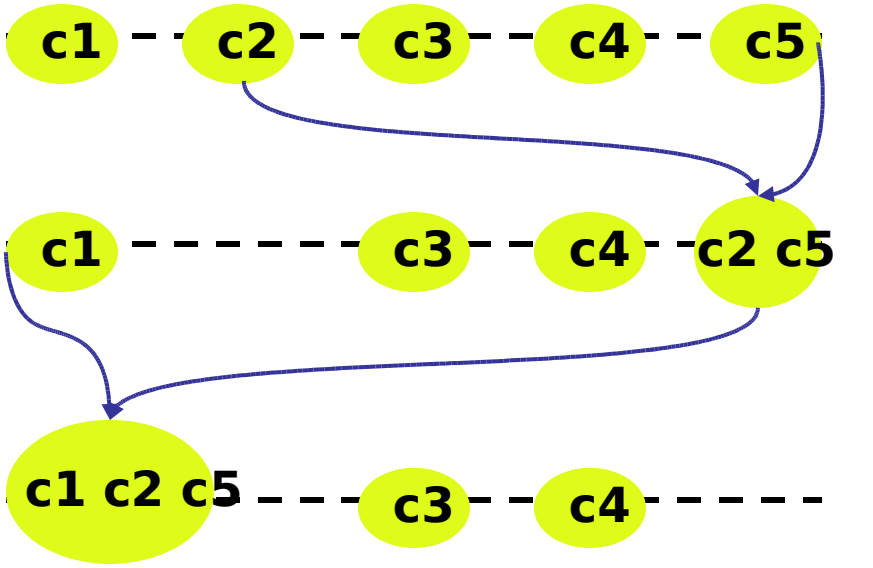
1.- Take two protein conformations C1 and C2 and superpose the centers of mass.



2.- Find the rotation $\Re$ that minimizes the function:

$$\mathbf{rmsd}(C_1, C_2) = \sqrt{\left(\sum_{k=1}^{N} (\mathbf{r}_k^{(C1)} - \Re \mathbf{r}_k^{(C2)})^2 \right)}$$

The Clustering Algorithm(*)



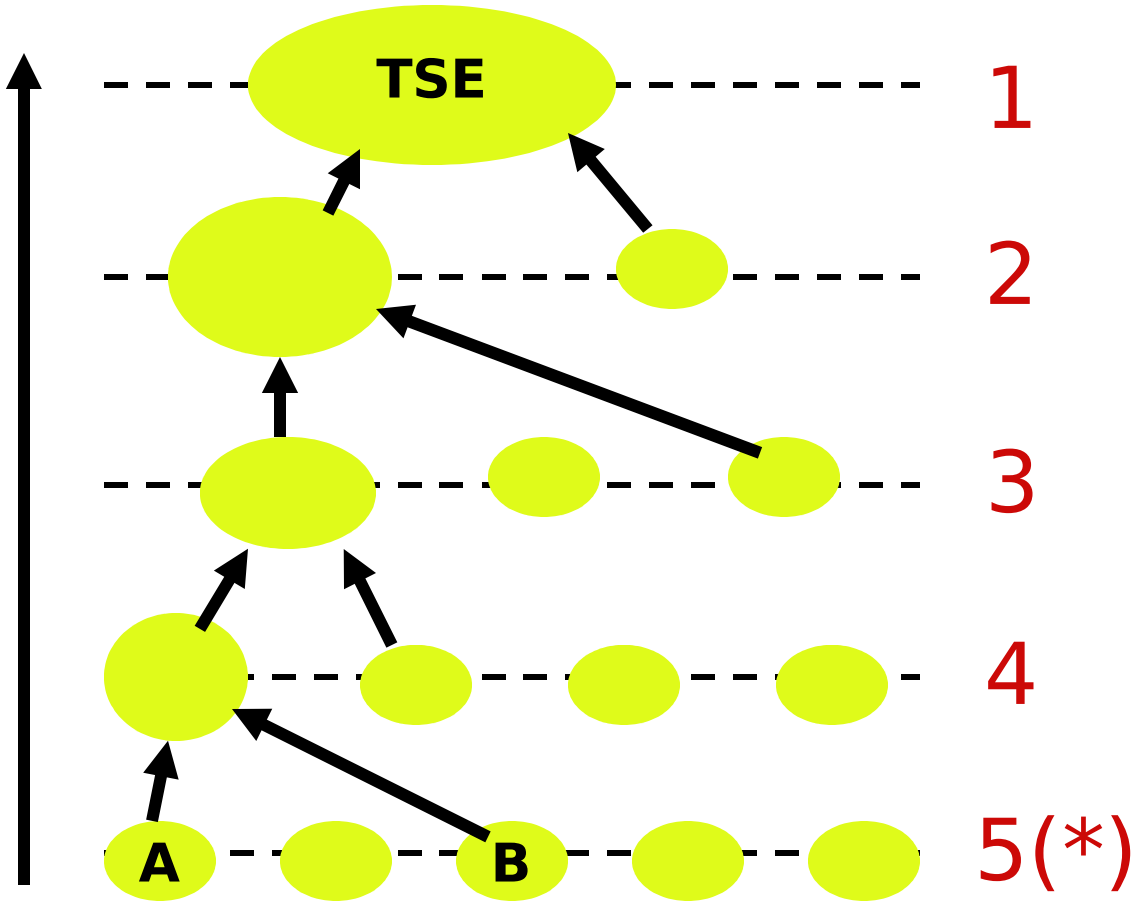
1.- Find pair of clusters with minimal root mean square distance (rmsd).

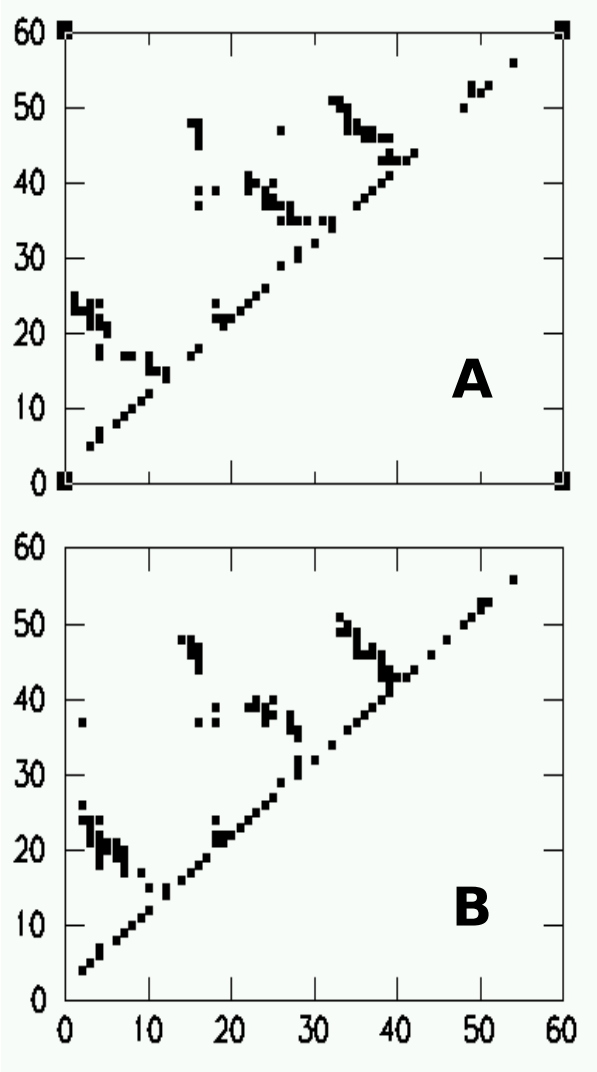
$$\mathrm{rmsd}_{ij} = \frac{1}{\mathrm{N}_i.\mathrm{N}_j} \sum_{C_i} \sum_{C_j} \mathrm{rmsd}(C_i, C_j)$$

2.- Join them. Go to next stage and repeat procedure.

(*): Sneath,P.H.A., Numerical Taxonomy,1973

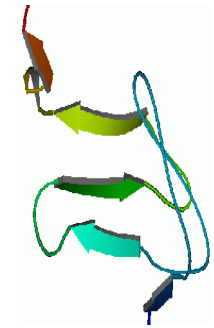
Clustering Tree

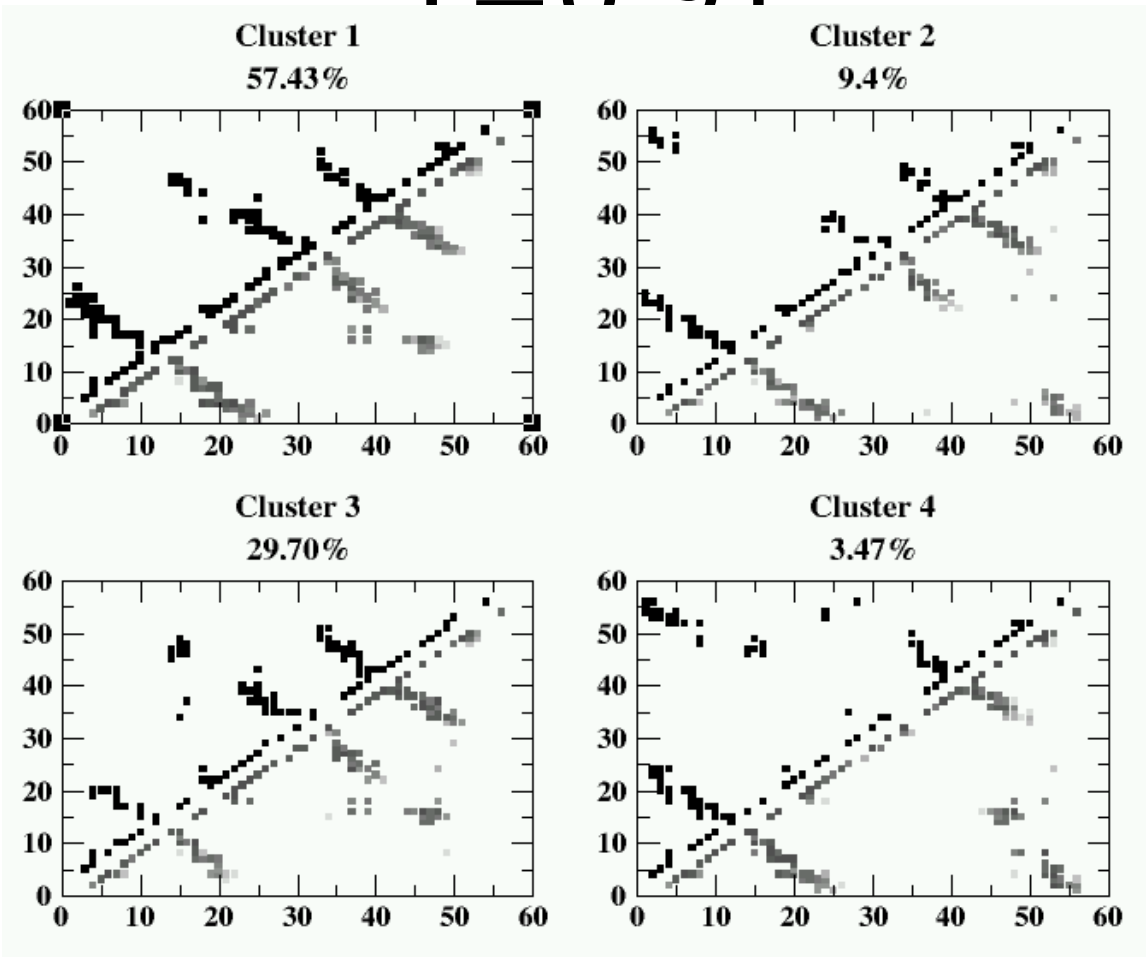


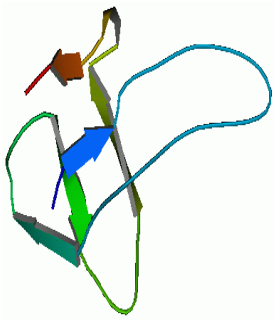


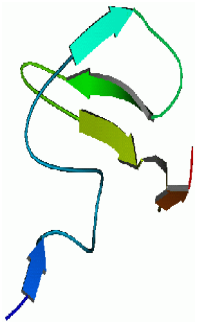
(*)Two similar Folding Pathways when we have 5 Clusters.

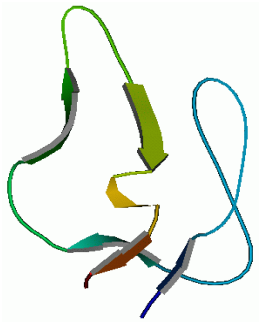
The Clusters Set of TSE for T=0.91











The TSE can be divided into a discrete set of Folding Pathways

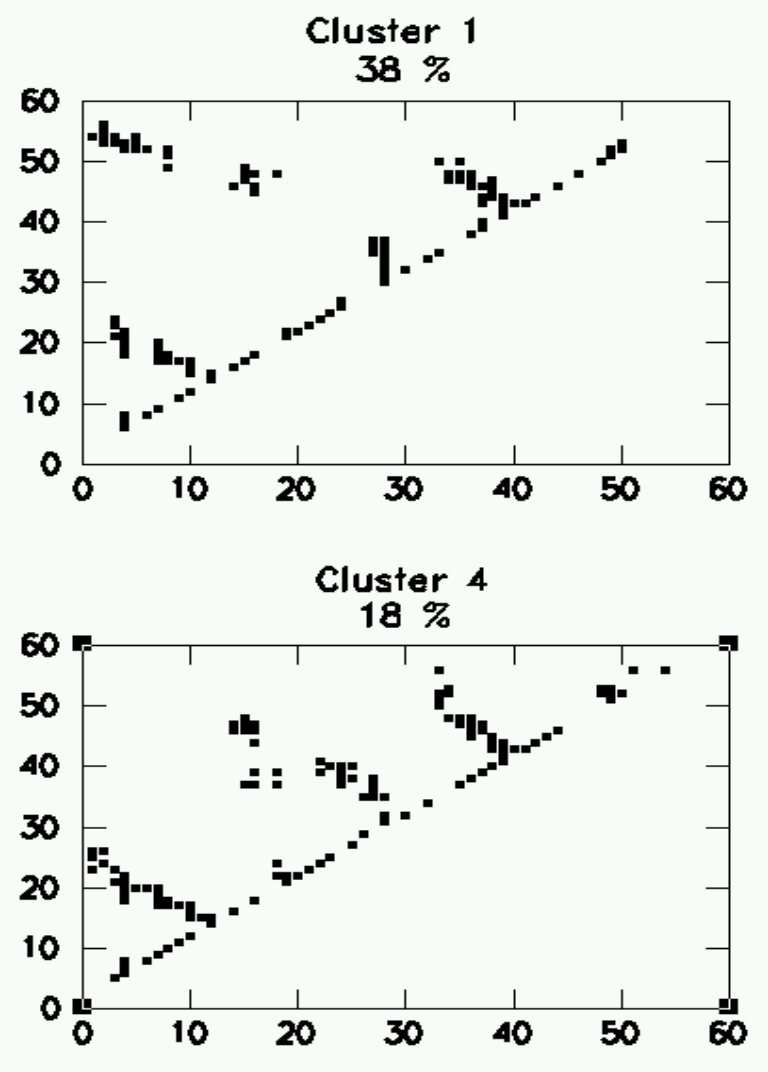
Changing Pathways Preferences

1.- Increase the intensity of contacts for cluster 4 and reduce the intensi of contacts for cluster 1 by using **E’ = E(1-δ).**

2.- The new conformations are used as starting point in our analysis.

3.- Repeat all steps to get the TSE members and make clustering.

Results: The majority of the conformations changes their pathway.

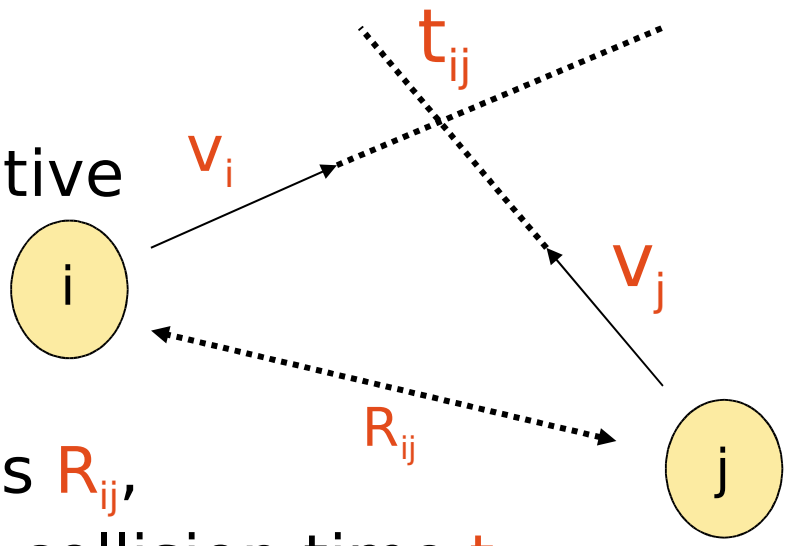


Summary

- TSE structure can be described through a finite number of folding pathways.
- The preference of a conformation to follow a folding pathway can be manipulated by changing the intensity of the contacts present in a pathway.

Discrete Molecular Dynamics (DMD)

Two particles moves with velocities v_i and v_j with relative position r_{ij}.

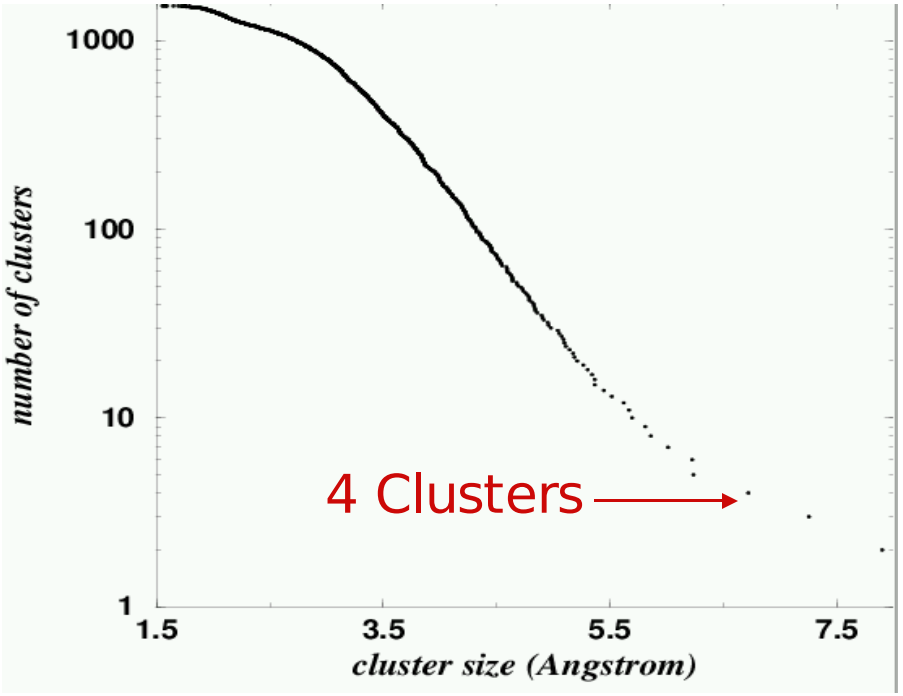


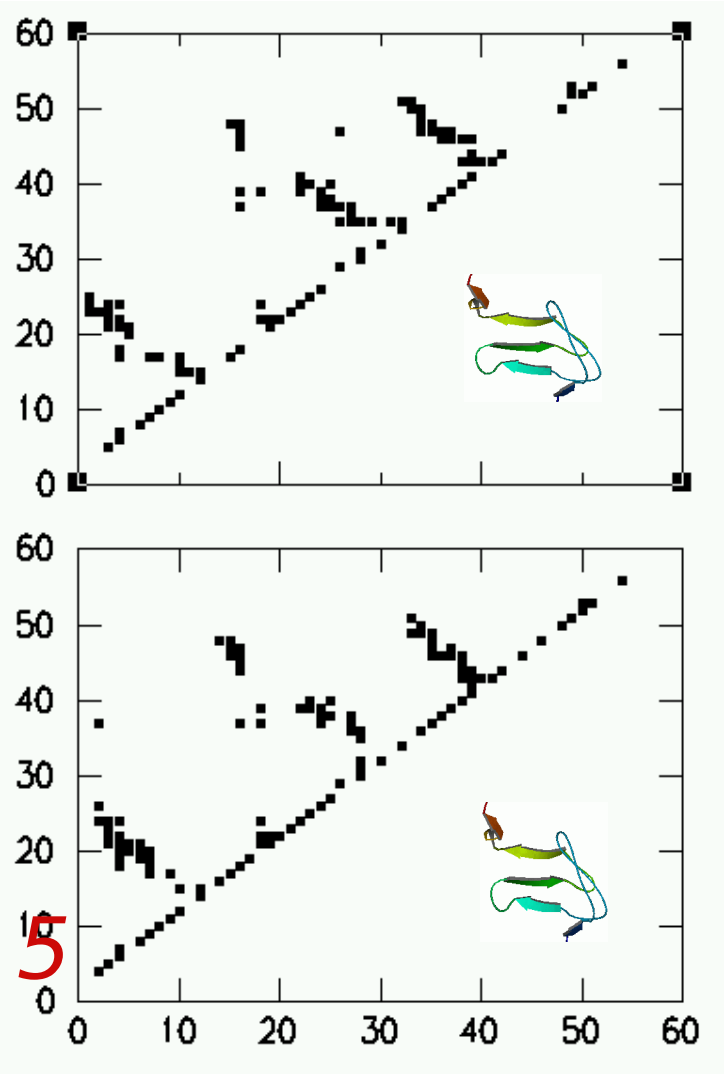
When distance between these two particles becomes R_{ij}, a collision happens and the collision time t_{ij} satisfies: $(r_{ij} + t_{ij}v_{ij})^2 = R_{ij}^2$

Select the pair with minimum t_{ij} and repeat.

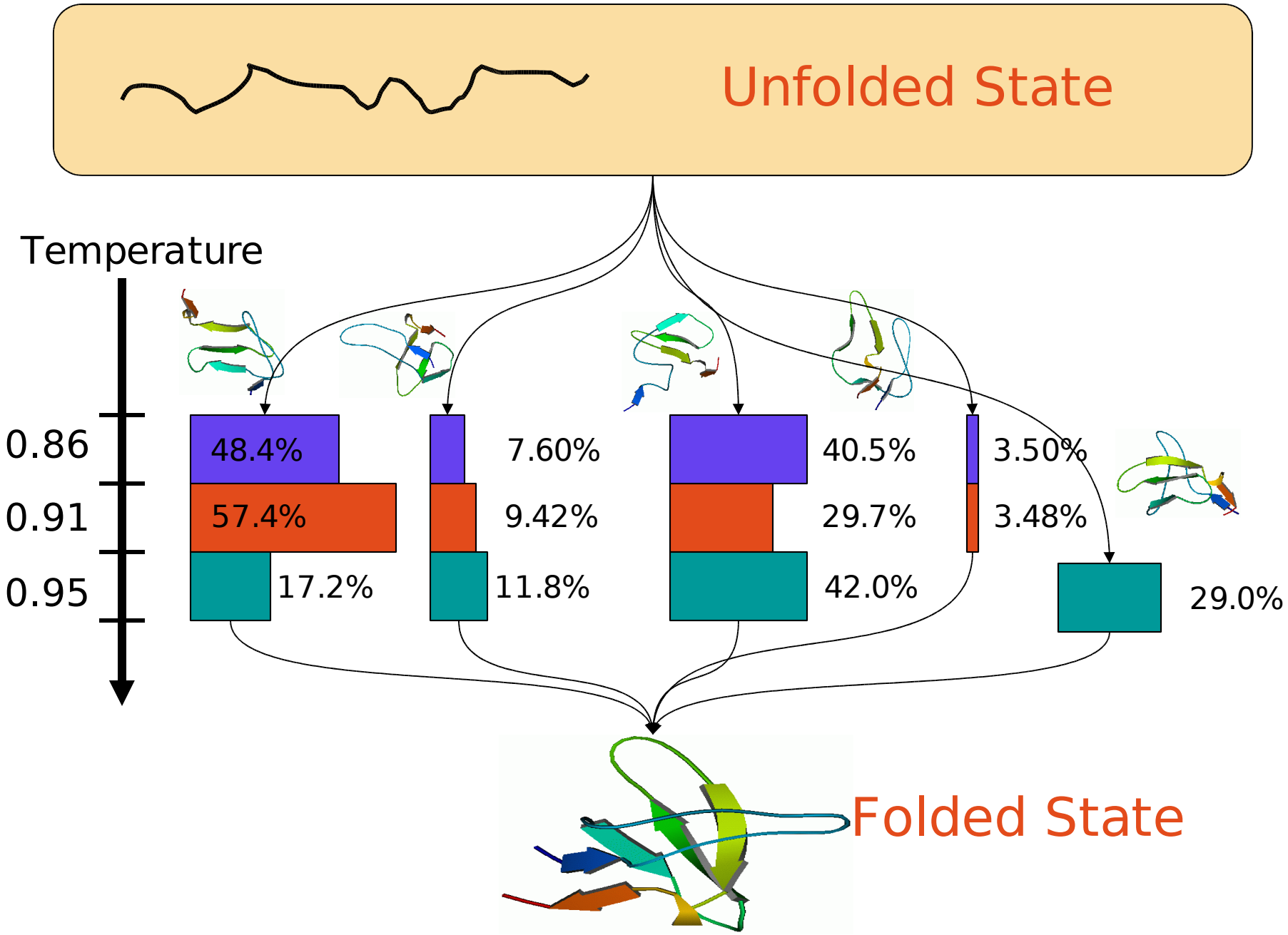
These quantities are conserved: $E_i + E_j$, $P_i + P_j$ and

Why 4 Clusters?





Two similar Folding Pathways when we have 5 Clusters.


Unfolded State
Temperature
0.86
0.91
0.95
48.4%
57.4%
17.2%
7.60%
9.42%
11.8%
40.5%
29.7%
42.0%
3.50%
3.48%
29.0%
Folded State