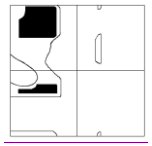
Multiscale Exploration of Conformational Space Using the MOLS Technique

N. Gautham

Department of Crystallography and Biophysics
University of Madras, Guindy Campus
Chennai 600 025

In this talk

- Introduction – the problem
- Mean Field Theory
- Mutually Orthogonal Latin Squares
- Applications to –
  - Peptide structure
  - Energy landscapes
  - Protein structure
  - Multiscale modelling

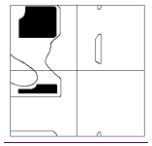


The Problem

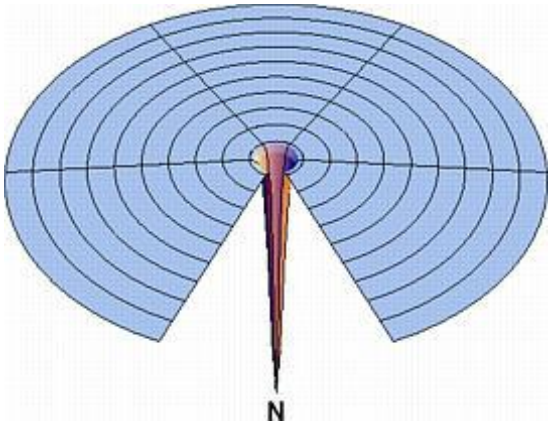
- Exploring and understanding the complex conformational space of molecules (proteins, peptides…)
- Large number of dimensions
- Energy landscape is rugged
- Multiple minima

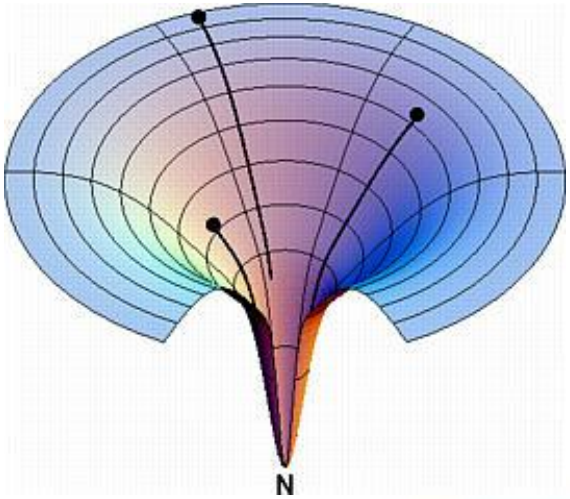
Questions

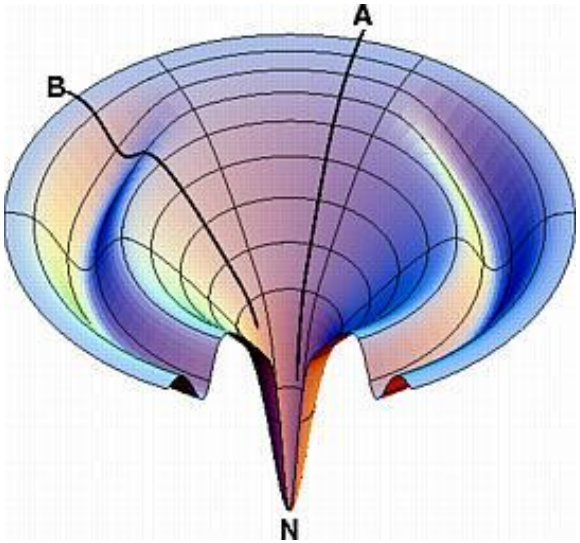
- How rugged?
- How many minima?
- How deep are they?

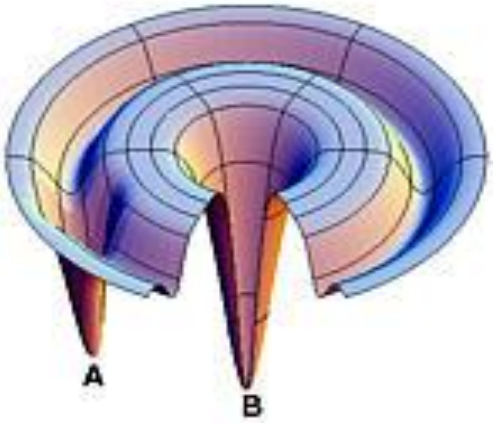


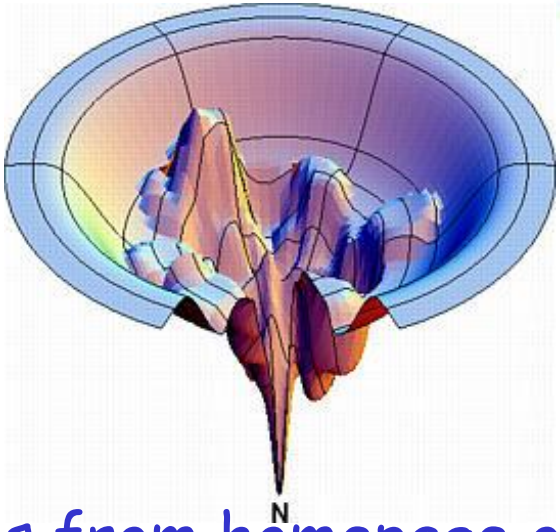
Introduction











These images from homepage of K.A. Dill

Introduction

- Conformational space can be explored by MD simulation at high temperature + quenching
- Also by the MOLS technique
- MOLS is essentially an optimization technique
- MOLS may be explained as a variant of Mean Field Technique

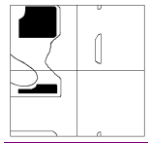
Mean Field Technique

- Mean field technique has been applied, e.g. to predict protein side chain structure

- Φ is the conformational search space

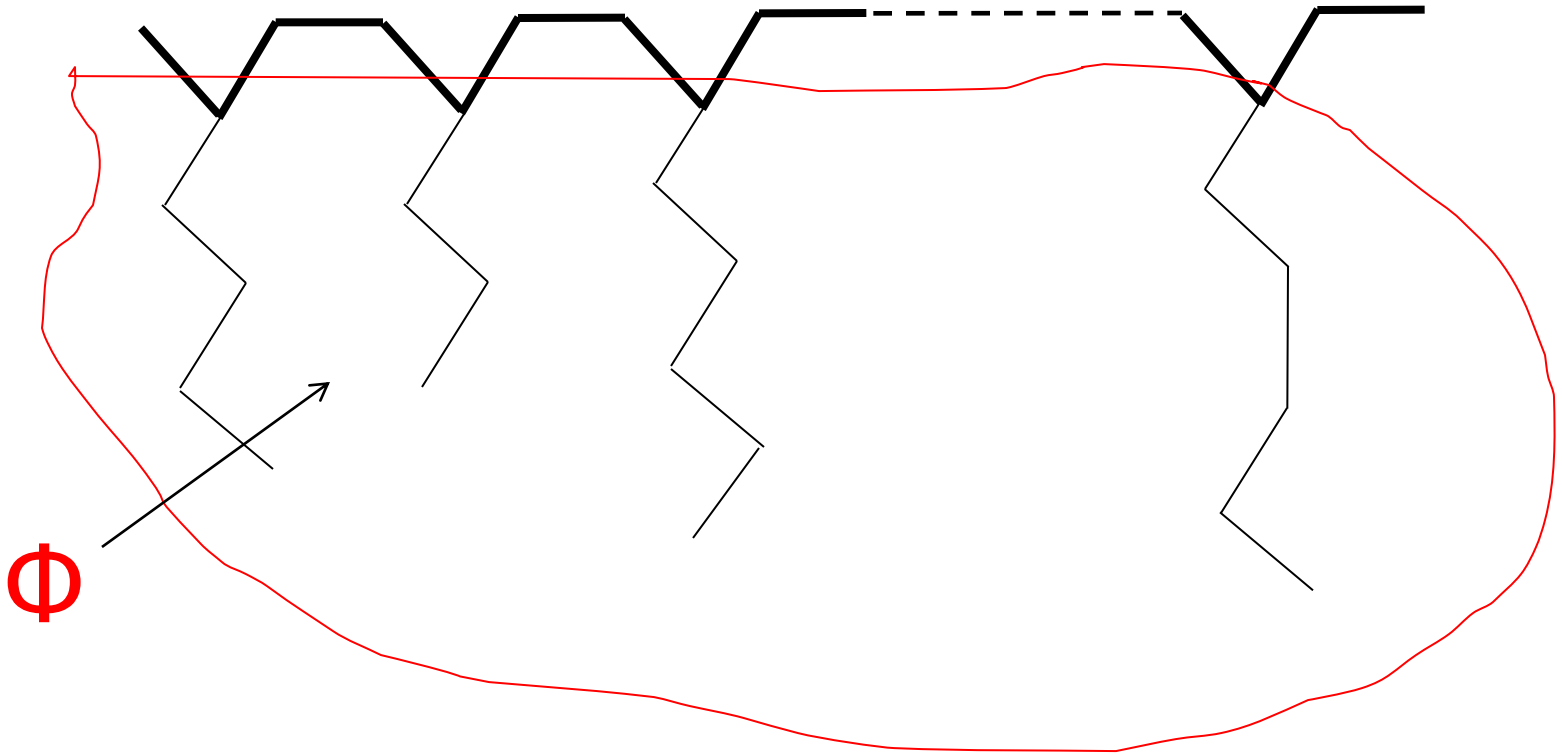
- This is divided into a number of subspaces φ_i

- Each such subspace has a number of states φ_{ij} each with a probability of occurrence ρ_{ij}



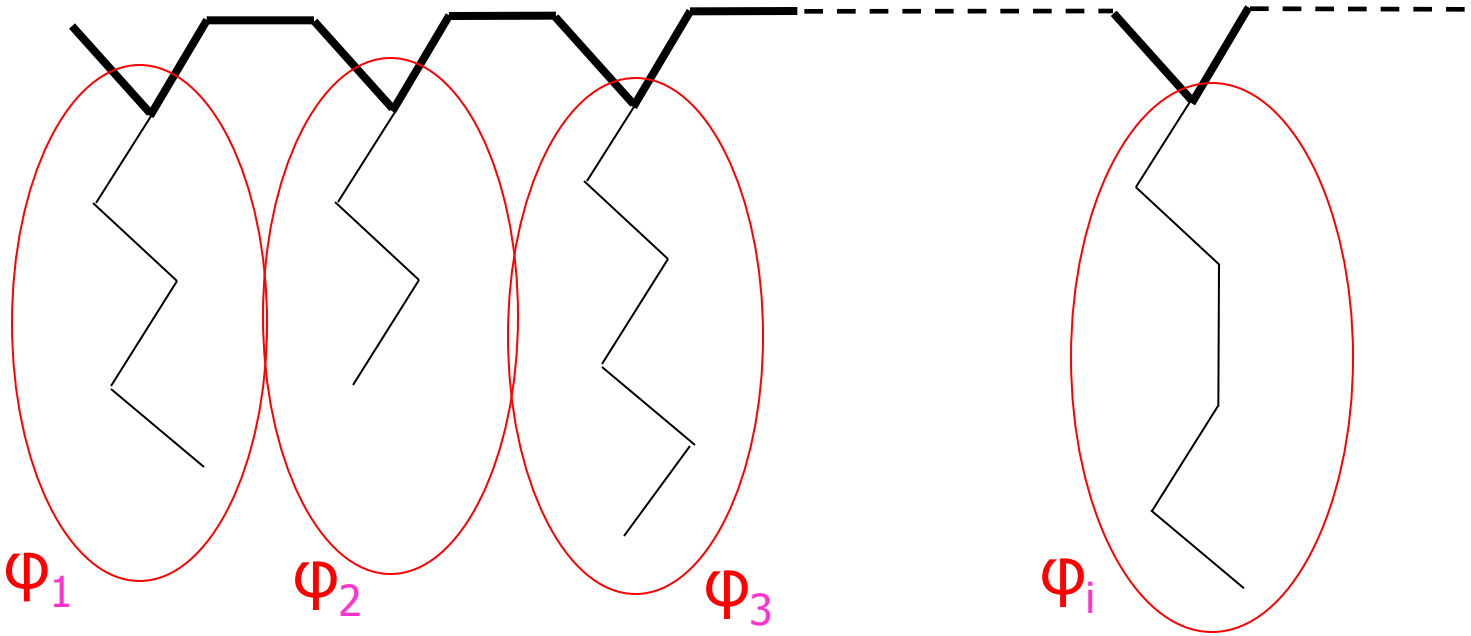
Mean Field Technique

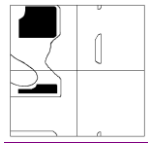
Φ is the conformational search space



Mean Field Technique

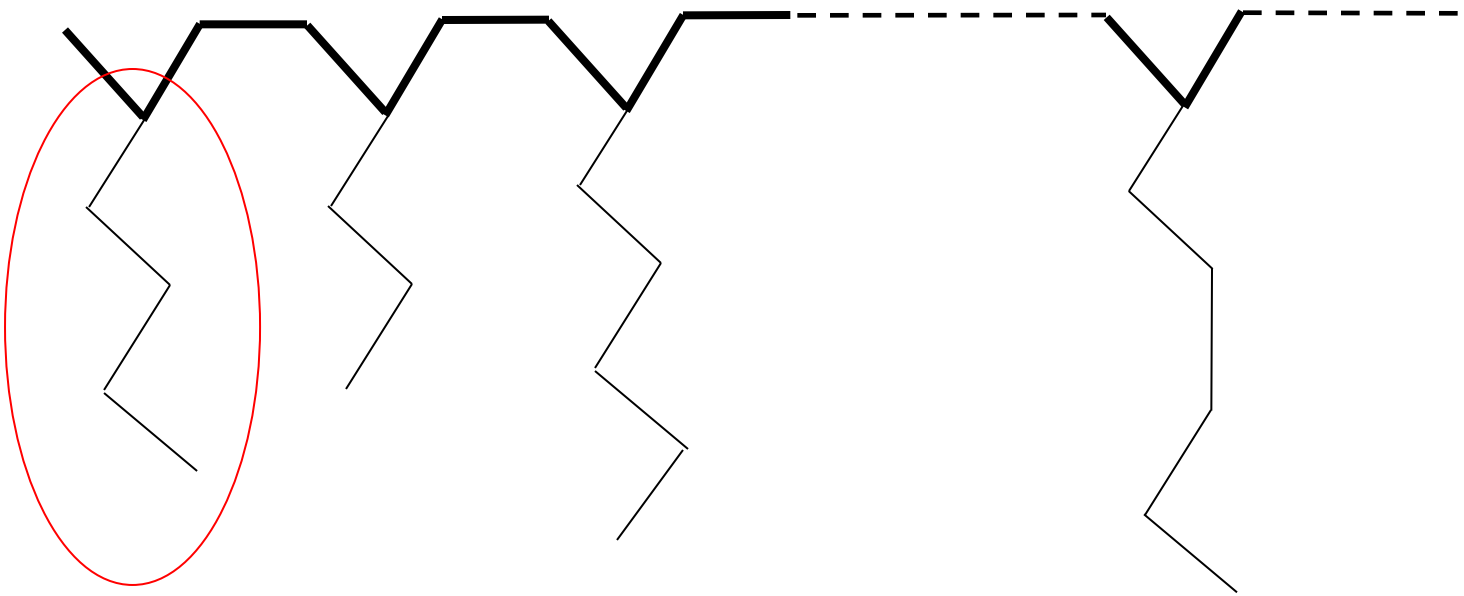
Φ is divided into a number of subspaces φ_i





Mean Field Technique

- Each such subspace has a number of states φ_{ij} each with a probability of occurrence ρ_{ij}

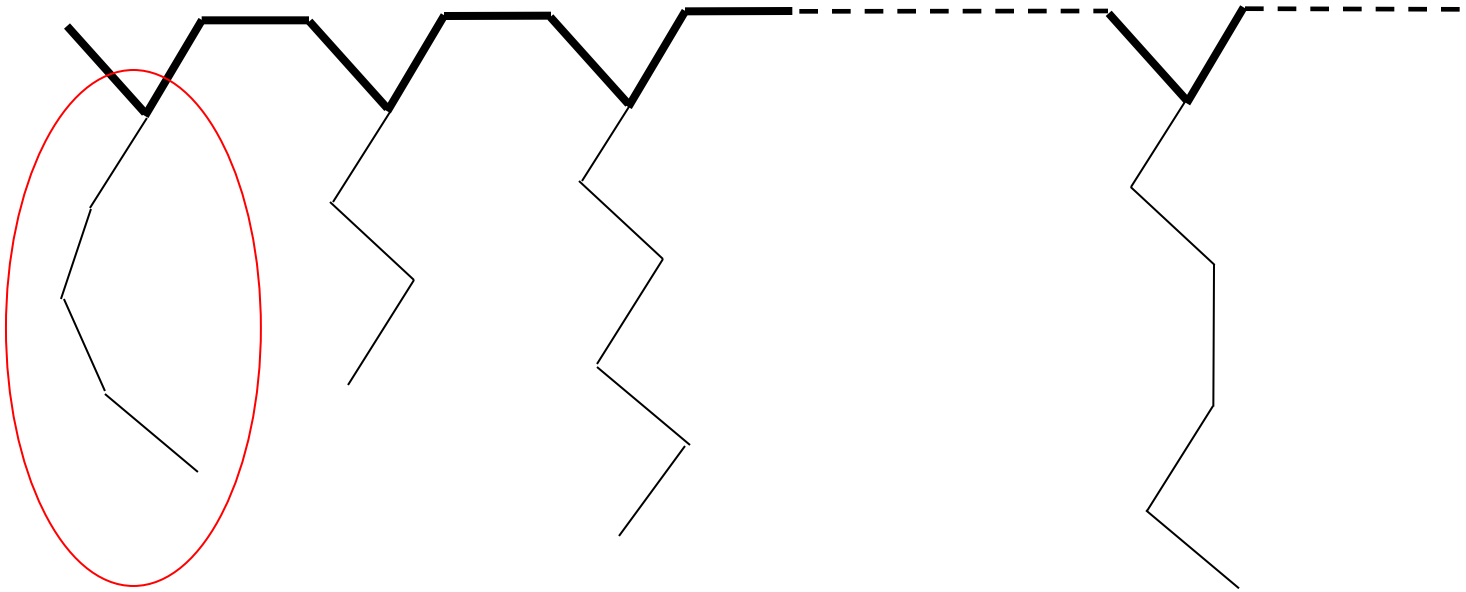


Φ_{11}

Probability = ρ_{11}

Mean Field Technique

- Each such subspace has a number of states φ_{ij} each with a probability of occurrence ρ_{ij}

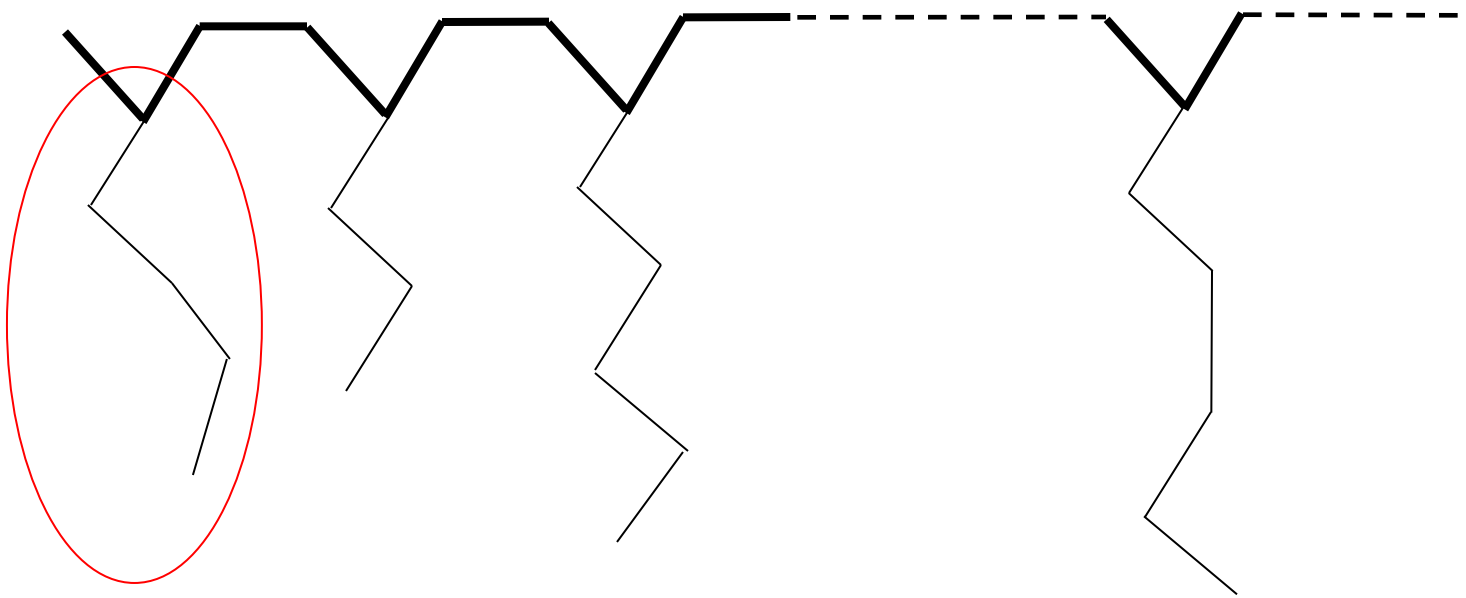


Φ_{12}

Probability = ρ_{12}

Mean Field Technique

- Each such subspace has a number of states φ_{ij} each with a probability of occurrence ρ_{ij}



Φ_{13}

Probability = ρ_{13}

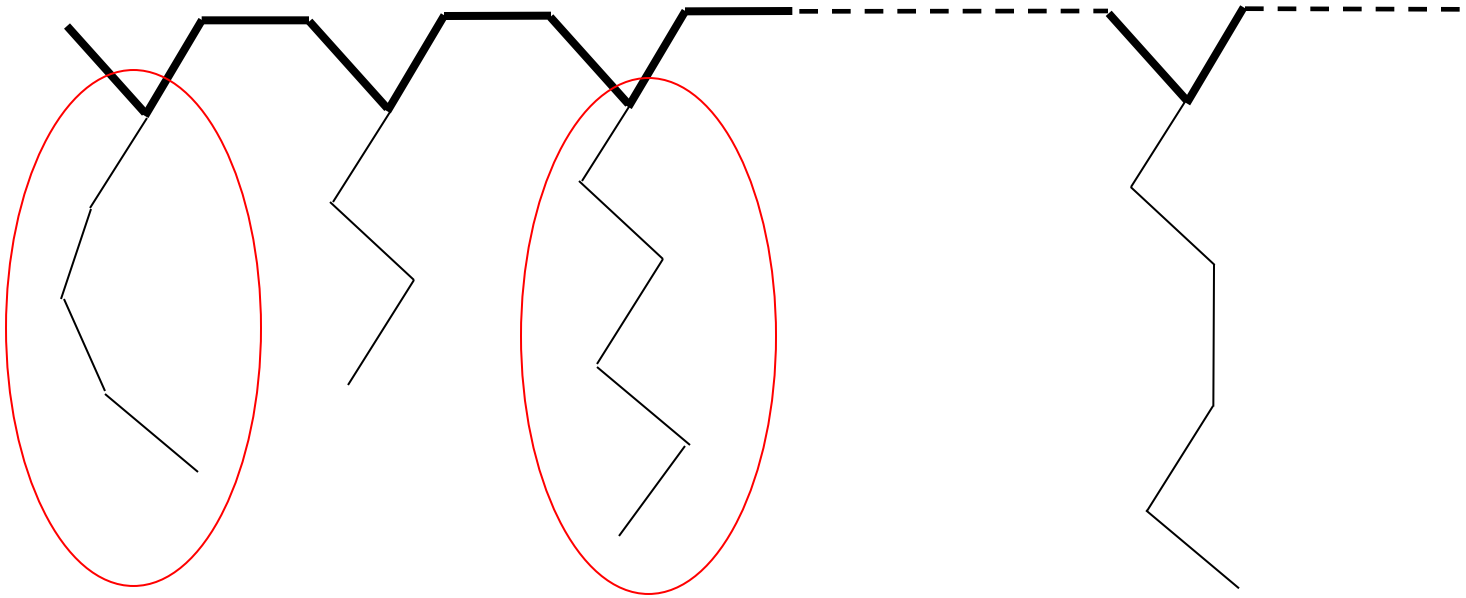
Mean Field Technique

- First $\rho_{ij} = 1/n_i$
- n_i is number of states of subspace i

Mean Field Technique

Next, the effective potential due to a state φ_{rs} of a subspace φ_r is evaluated as

$$V^{eff}(\varphi_{12}) = \Sigma_{i,j}\, \rho_{ij} V(\varphi_{12}, \varphi_{ij})$$



$r = 1, s = 2$

Φ_{12} , ρ_{12}

$i = 3, j = 1$

Φ_{31} , ρ_{31}

$\rho_{31} V(\varphi_{12}, \varphi_{31})$

Mean Field Technique

- Next, all the probabilities ρ_{ij} re-evaluated from the respective effective potential as

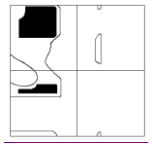
$$\rho_{ij} = \exp\{-V^{eff}(\varphi_{ij})/RT\} / \Sigma_q \exp\{-V^{eff}(\varphi_{ij})/RT\}$$

- The cycle is repeated -------
  - Evaluate effective potential based on probability
  - Evaluate probability based on effective potential

----- until convergence

Mean Field Technique

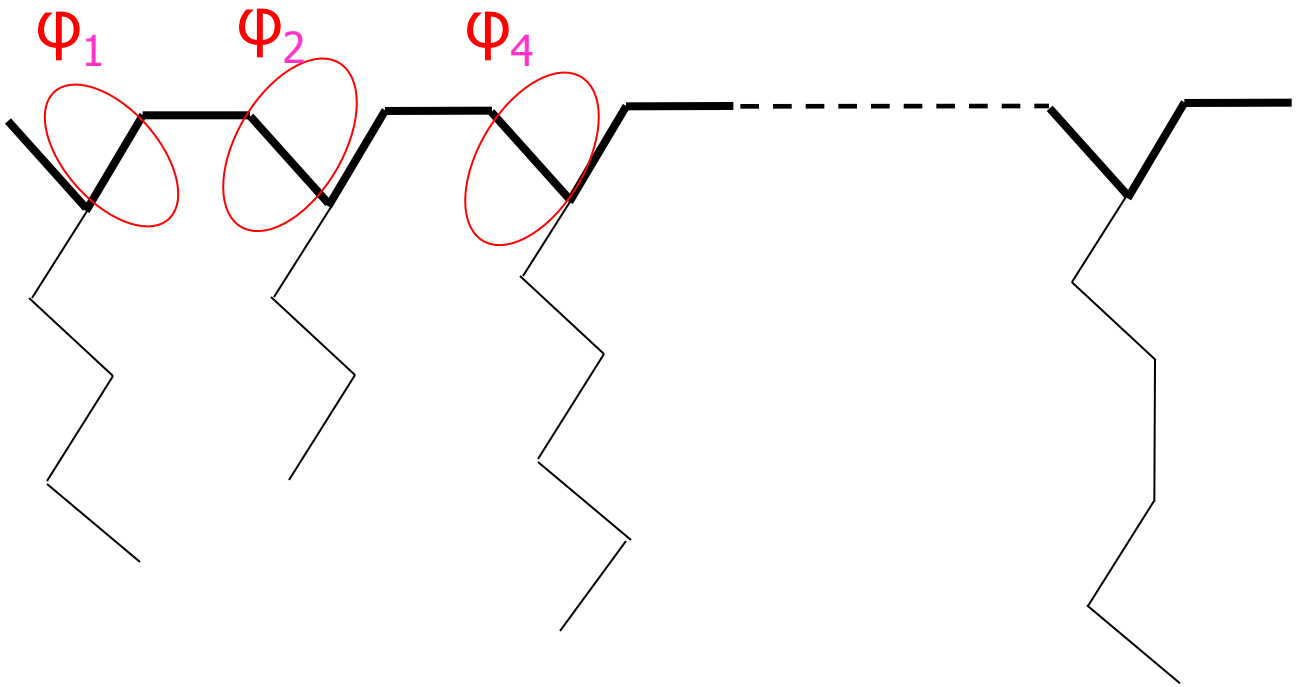
- At convergence we have determined all the probabilities ρ_{ij} of all the states of every subspace
- Finally, side-chain conformations (subspace states) with highest probability represent the 'true state' of the side-chain conformational space Φ



Mean Field Technique

Application to peptide/protein (backbone) structure

------ Backbone torsion angles as subspaces?



Interaction between a pair of subspaces is not independent of other subspaces

Mean Field Technique

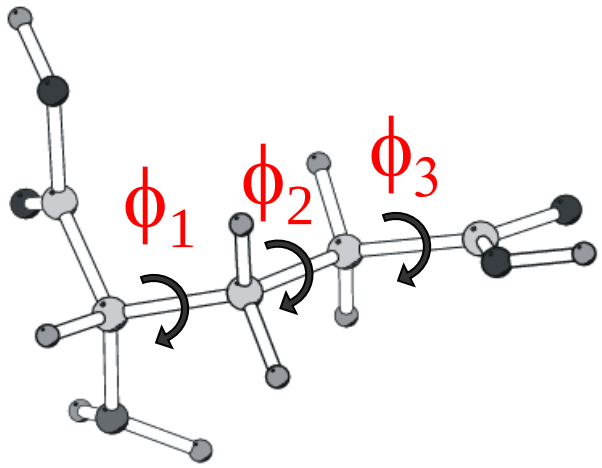
- Extension to torsion angle space is not straight-forward
- The interaction between a pair of subspaces (when the subspaces are the torsion angles) does not depend only on their respective states, but is a function of the states of all other subspaces.
- i.e. $V(\varphi_{rs}, \varphi_{ij})$ is not meaningful – we need

$$V(\varphi_{rs}, \varphi_{ij}, \ldots\ldots)$$

- Combinatorial explosion ! Complexity is NP

Mutually Orthogonal Latin Squares

- To avoid combinatorial explosion we use a small sample (~ n^2) of the possible (m^n) combinations
- We use mutually orthogonal Latin squares (MOLS) to identify the sample
- e.g. 3 torsion angles, 5 values each -



φ_{ij} , i = 1, 3; j = 1, 5

Totally 5^3 = 125 conformations

Mutually Orthogonal Latin Squares

3 MOLS of order 5

$\varphi_{11}, \varphi_{21}$ φ_{31}	$\varphi_{12}, \varphi_{22}$ φ_{32}	$\varphi_{13}, \varphi_{23}$ φ_{33}	$\varphi_{14}, \varphi_{24}$ φ_{34}	$\varphi_{15}, \varphi_{25}$ φ_{35}
$\varphi_{15}, \varphi_{24}$ φ_{33}	$\varphi_{11}, \varphi_{25}$ φ_{34}	$\varphi_{12}, \varphi_{21}$ φ_{35}	$\varphi_{13}, \varphi_{22}$ φ_{31}	$\varphi_{14}, \varphi_{23}$ φ_{32}
$\varphi_{14}, \varphi_{22}$ φ_{35}	$\varphi_{15}, \varphi_{23}$ φ_{31}	$\varphi_{11}, \varphi_{24}$ φ_{32}	$\varphi_{12}, \varphi_{25}$ φ_{33}	$\varphi_{13}, \varphi_{21}$ φ_{34}
$\varphi_{13}, \varphi_{25}$	$\varphi_{14}, \varphi_{21}$	$\varphi_{15}, \varphi_{22}$	$\varphi_{11}, \varphi_{23}$	$\varphi_{12}, \varphi_{24}$

- **Each sub square corresponds to one conformation of the molecule**
- **The energy V is calculated for each of the n^2 conformations**

Mutually Orthogonal Latin Squares

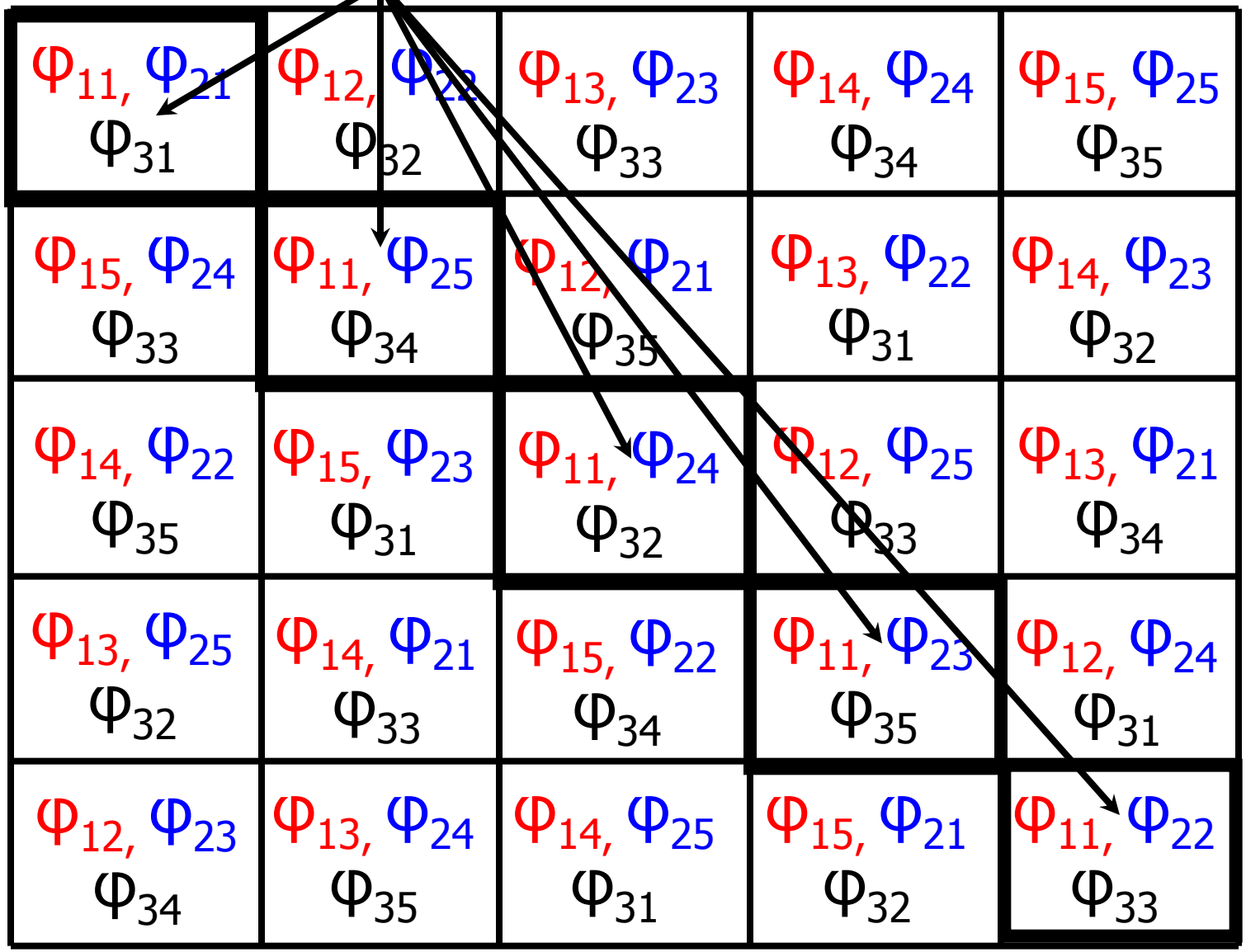
- The effective energy is now

$$V^{eff}(\varphi_{rs}) = \Sigma_q w_q V_q(\varphi_{rs}\ldots)$$

- The summation is over all the points in the MOLS grid in which φ_{rs} occurs

Mutually Orthogonal Latin Squares

e.g. for φ_{11}



Mutually Orthogonal Latin Squares

- The effective energy is now

$$V^{eff}(\varphi_{rs}) = \Sigma_q \, w_q V_q(\varphi_{rs}\ldots)$$

$$w_q = \exp\{-V_q(\varphi_{rs}\ldots)/RT\} \,/\, \Sigma_q \exp\{-V_q(\varphi_{rs}\ldots)/RT\}$$

- Note: w_q is calculated from V_q, not from V^{eff}

Mutually Orthogonal Latin Squares

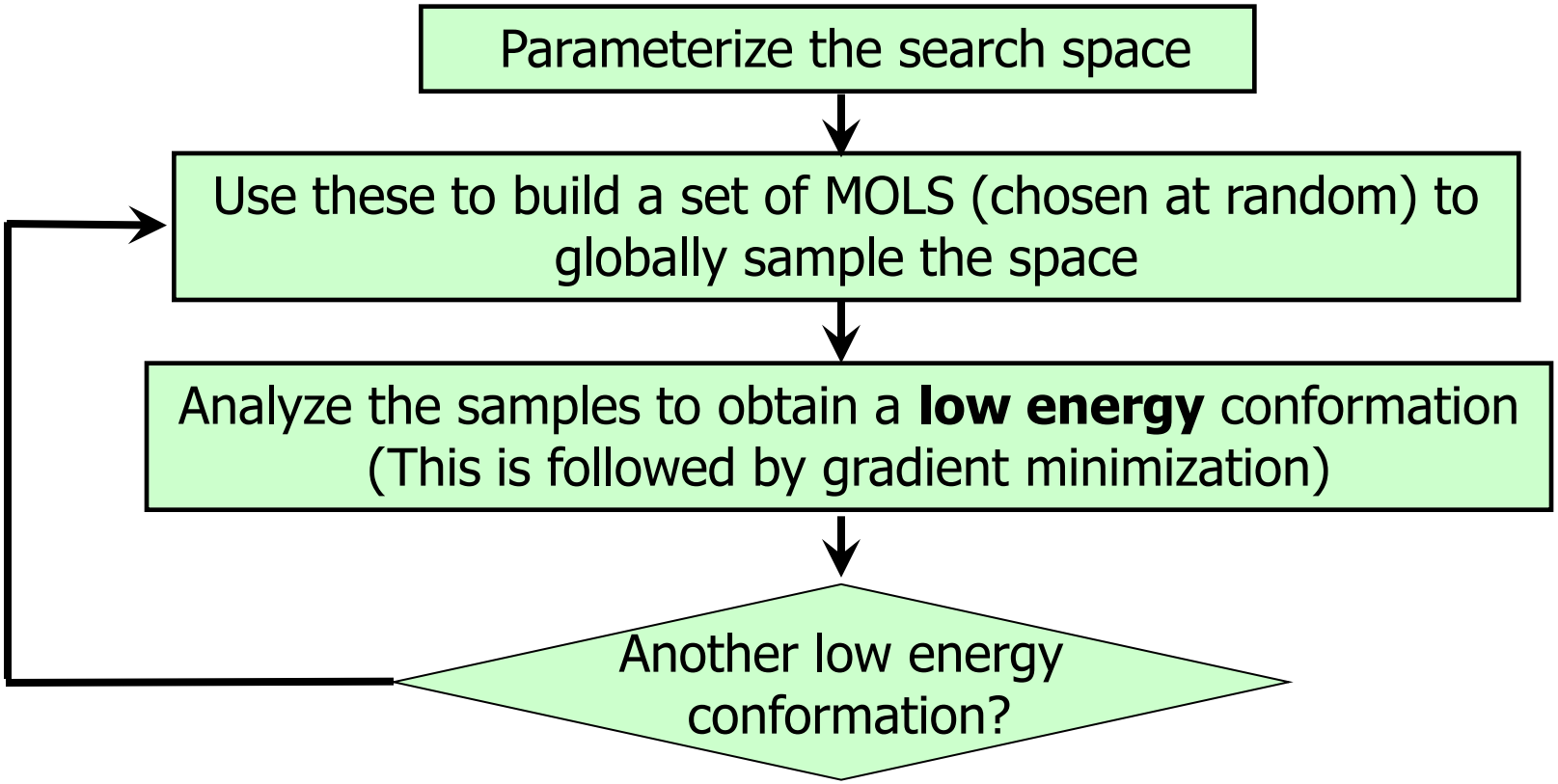
- Since w_q is calculated from V_q, not from V^{eff}
  therefore w_q is not ρ_{ij}
  and the procedure is not iterative

- For each torsion find value that gives Min(V^{eff})

- The set of Min(V^{eff}) values is Minimum (Low) energy conformation

- Procedure repeated for another low energy structure

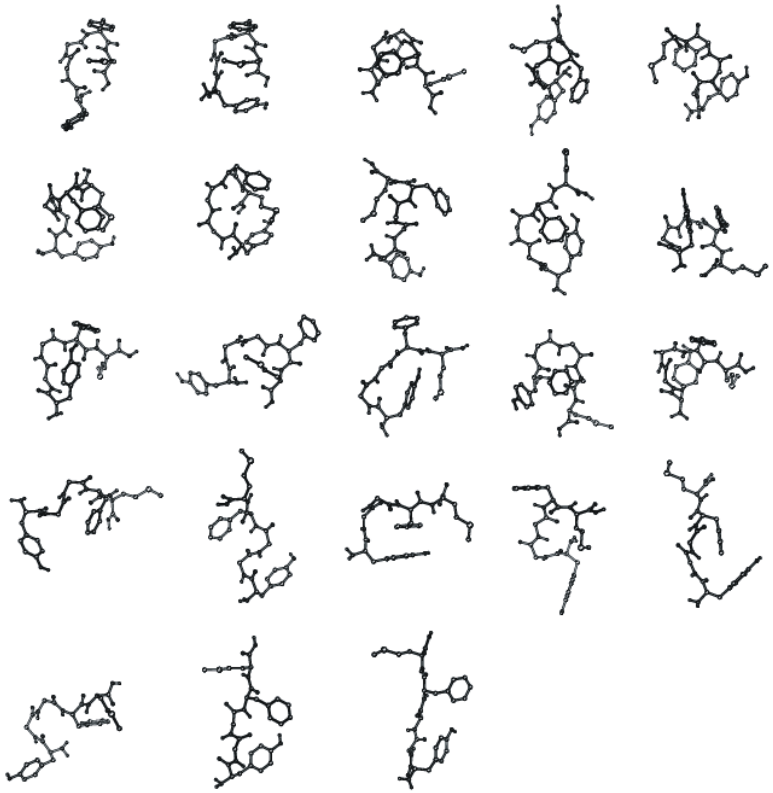
Mutually Orthogonal Latin Squares



Biophysical Jl., 84, 2897, 2003

Mutually Orthogonal Latin Squares

- We obtain ~1500 low energy structures
- By clustering, we show these may be reduced to ~ 50 mutually dissimilar structures

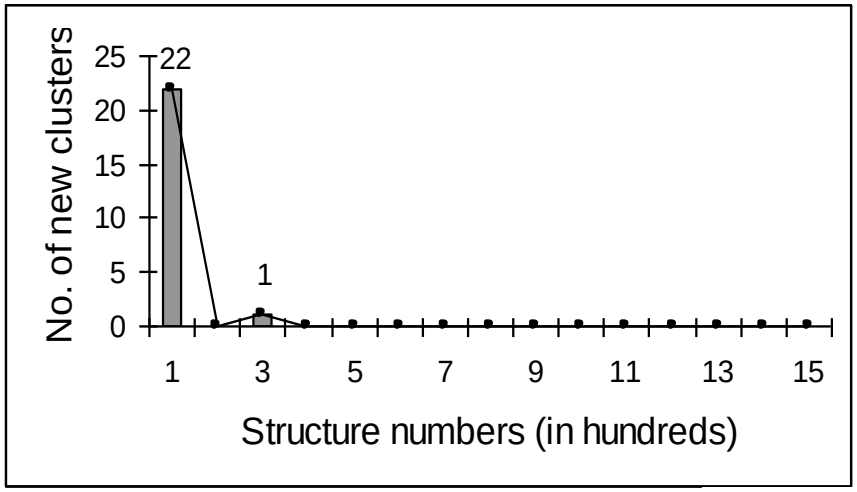


e.g. 23 structures for Met-enkephalin

Biopolymers, 74, 476 2004

Mutually Orthogonal Latin Squares

- The search is exhaustive

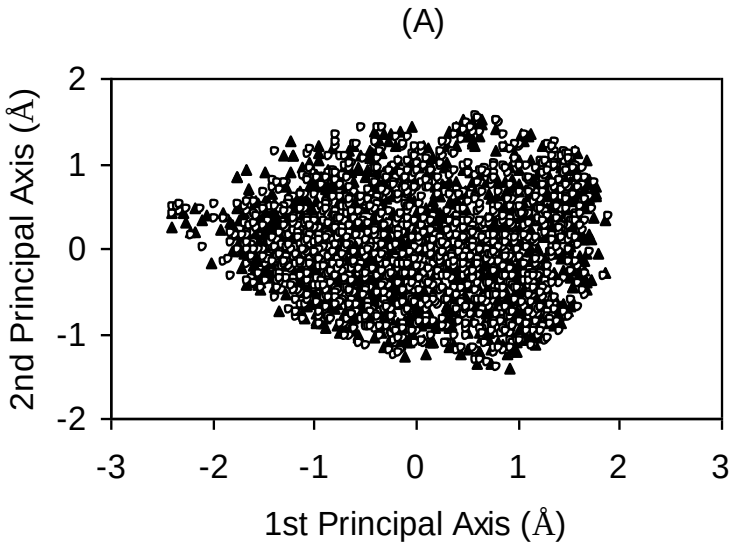


Plot of 'new' structure versus structure number

Sample Overlap

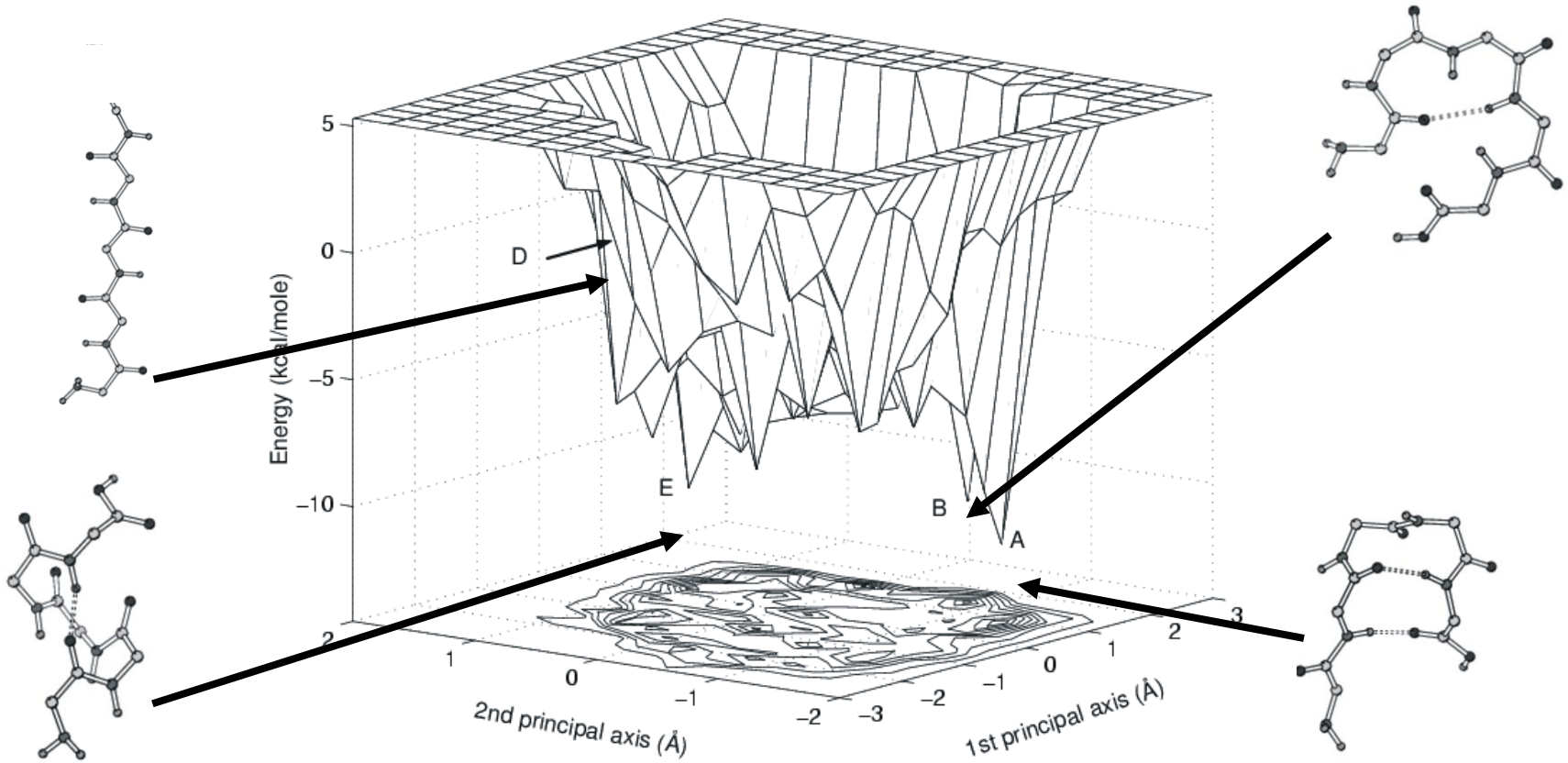
▲ First sample

○ Second sample



Applications – Peptide Structure

- Energy landscape, ECEPP/3 force field



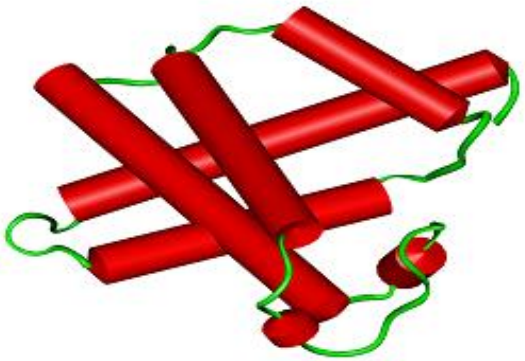
Minimal energy envelope for Met-enkephalin

Jl. Phys. Chem. B, 108, 11196, 2004

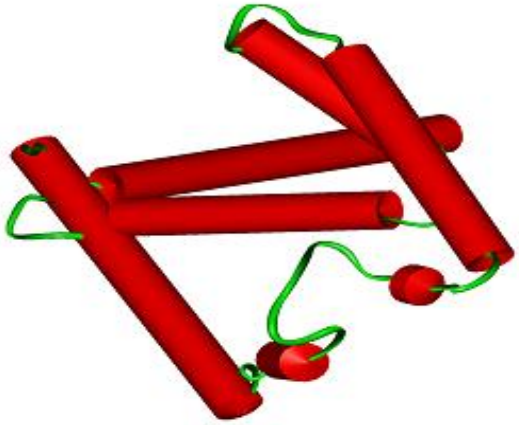
Applications – Protein Structure

- Multiscale approach to protein structure prediction
  MOLS libraries + MOLS assembly
  (ECEPP/3) (AMBER + 'hydrophobic')

Myoglobin



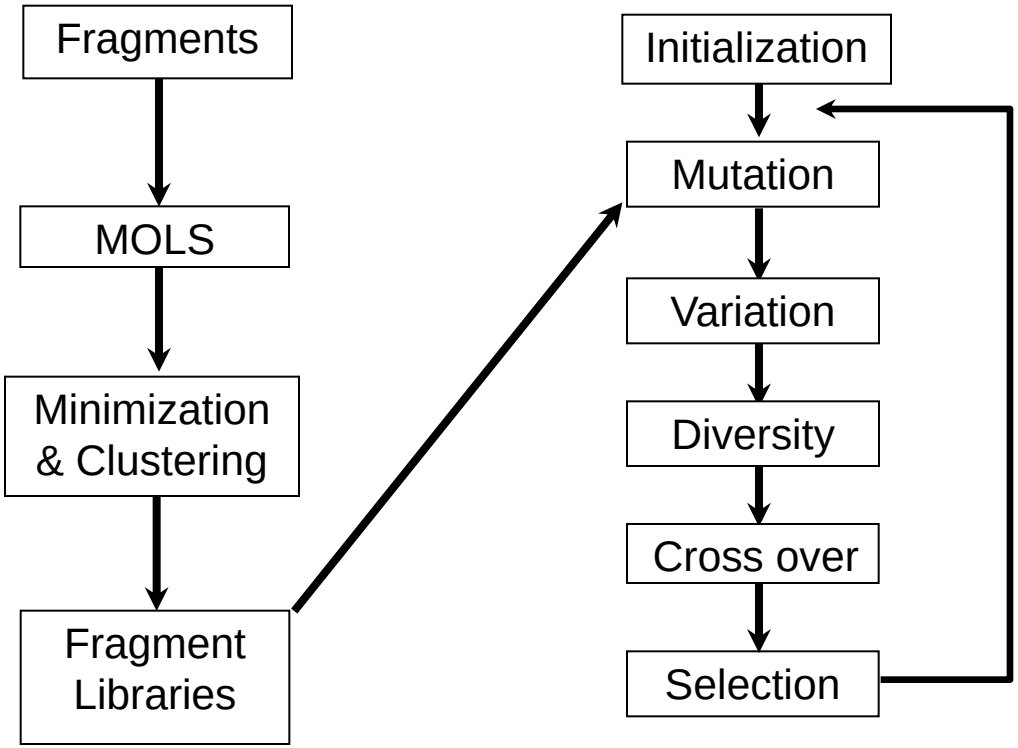
Crystal structure



Prediction
rmsd : 14.1 Å

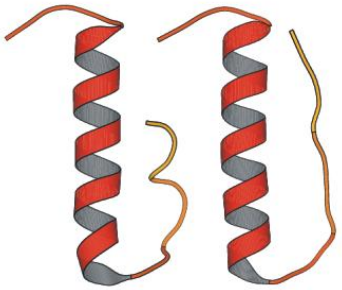
Applications – Protein Structure

- Multiscale approach to protein structure prediction
- MOLS libraries + genetic algorithms
  (ECEPP/3) (AMBER + 'hydrophobic')

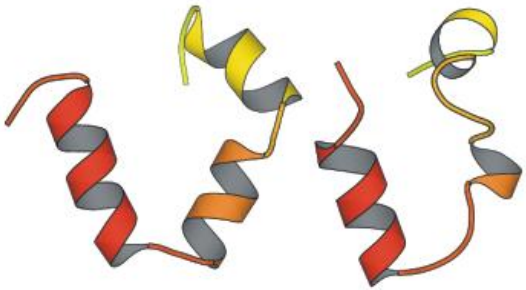


Applications – Protein Structure

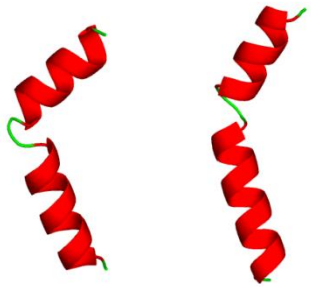
Left → Predicted; Right → Experimental



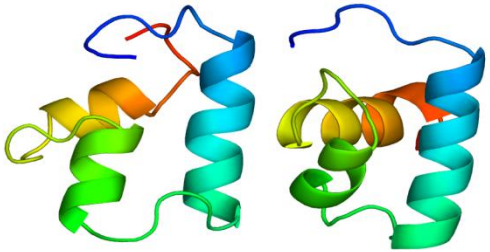
Avian Pancreatic Polypeptide 4.0 A



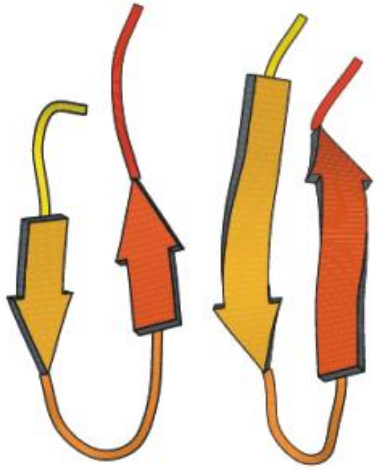
Villin Head Piece 5.2 A



Mellitin 4.3



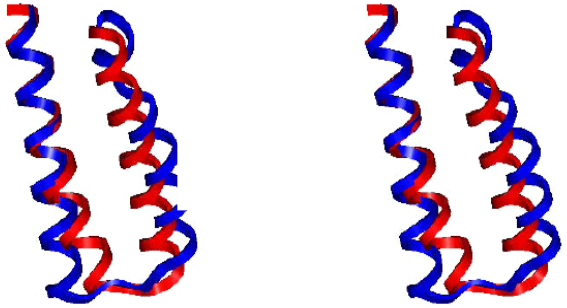
c-MYB 6.1 A



Tryptophan zipper 1.8 A

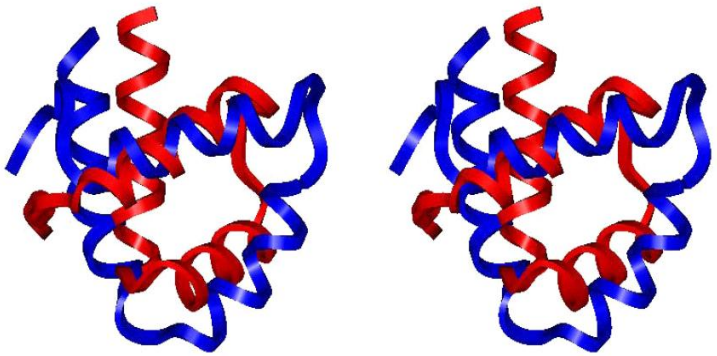
BBRC, 342,424, 2006

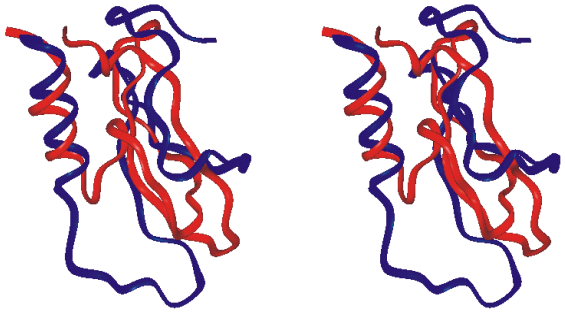
Applications – Protein Structure



Rab4 binding domain of Rabenosyn 5
46 residues
Backbone rmsd 3.6 Å

Engrailed Homeodomain
56 residues
Backbone rmsd 6.5 Å



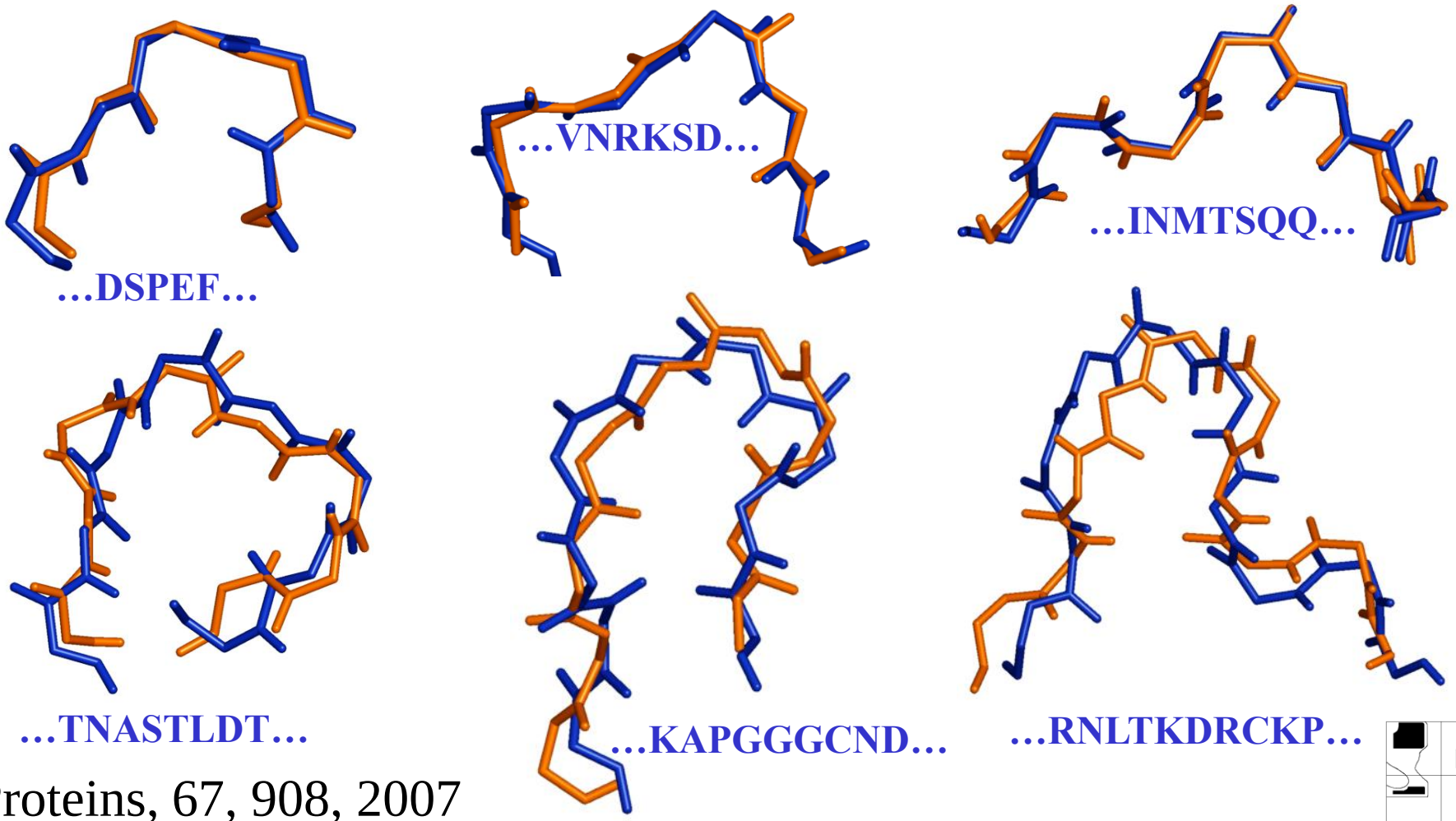


Bovine Pancreatic Trypsin Inhibitor
58 residues
Backbone rmsd 10.2 Å

Applications – Protein loops

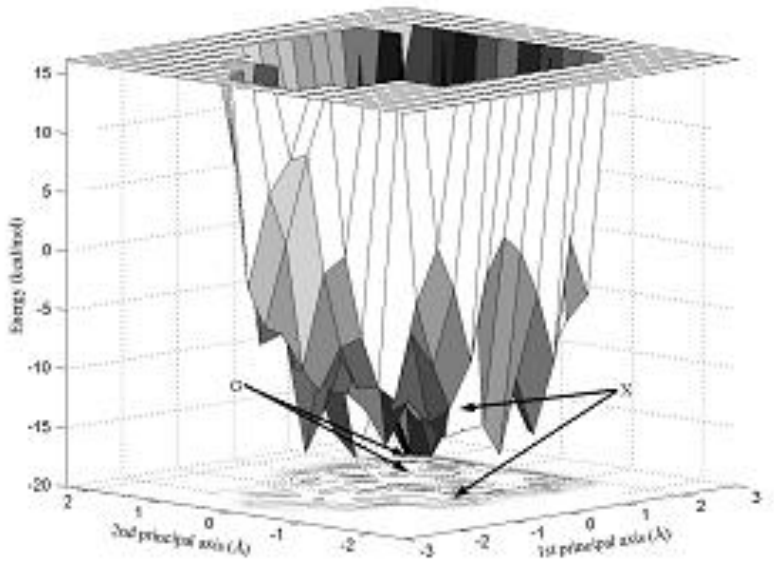
- Loops in protein crystal structures classified by size. Structure predicted using ECEPP/3

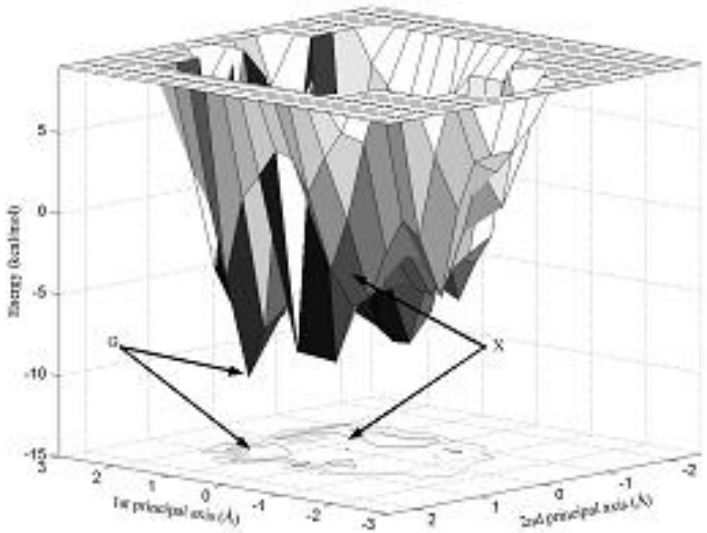
Blue – Predicted structure Orange – Crystal structure

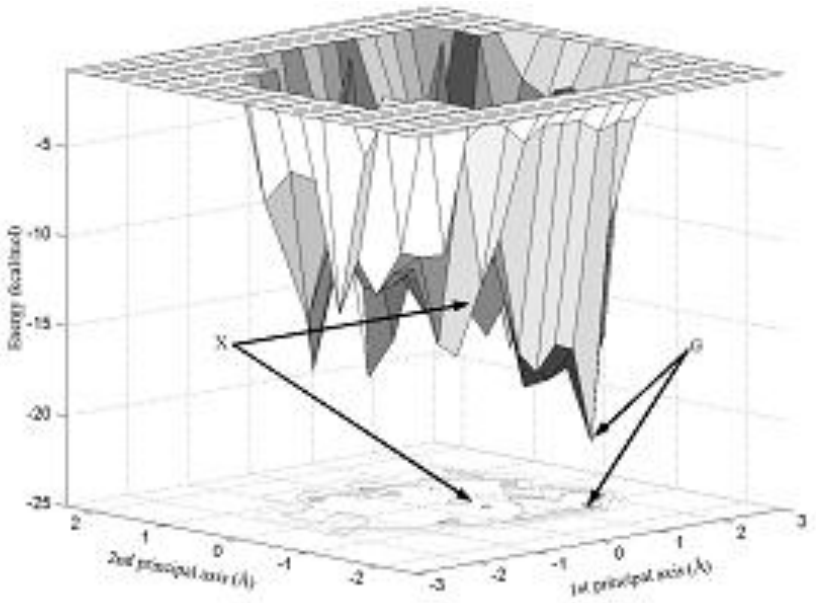


Proteins, 67, 908, 2007

Applications – Protein loops







Applications – Docking

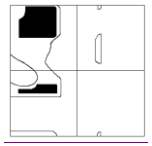
- Ligand (drug) docking to proteins

- Redefine the search space as the conformational space of peptide ligand (i.e. n torsion angles φ_1 to φ_n), plus the 'docking' space (i.e. the rotation and translation parameters of the peptide in receptor site, r_1 to r_6).

- Composite scoring function is now

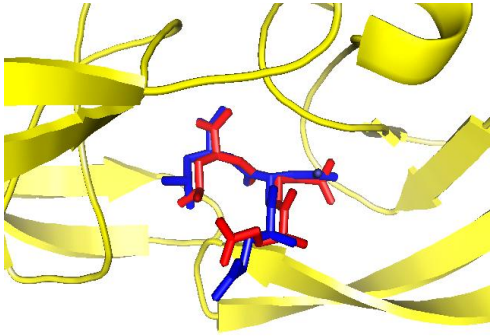
$$f_1\{\varphi_1 \text{ to } \varphi_n\} \quad + \quad f_2\{r_1 \ldots r_6\}$$

conformational energy + 'docking' energy

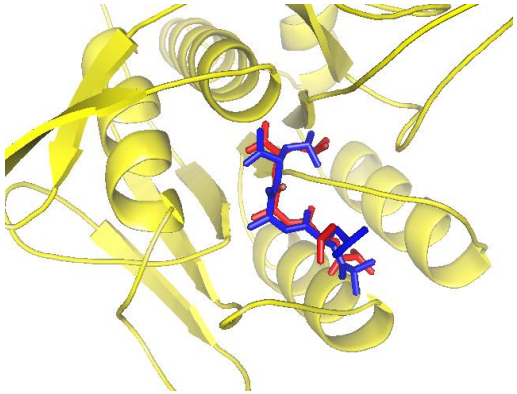


Applications – Docking

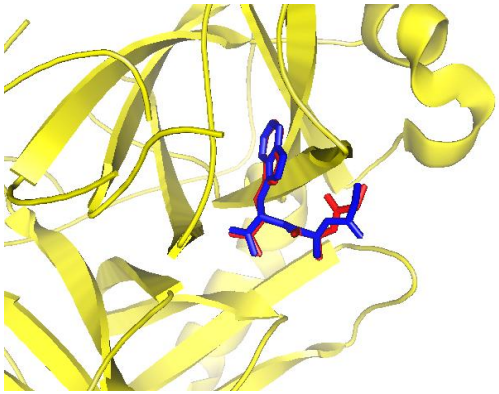
PDB ID: 1a30, RMSD = 0.66Å
Sequence: EDL



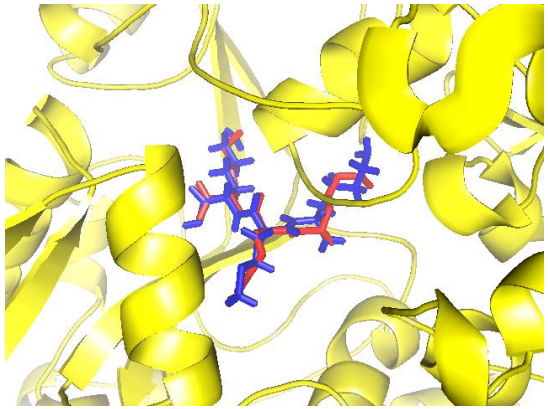
PDB ID: 1sua, RMSD = 1.50Å
Sequence: ALAL



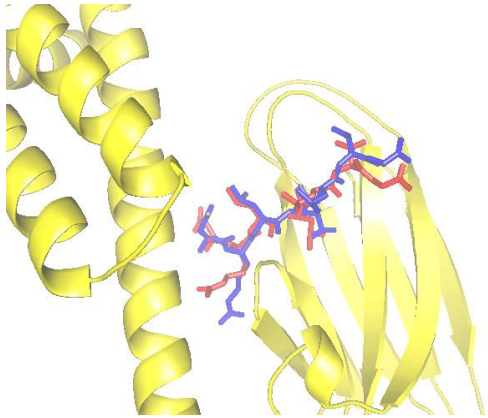
PDB ID: 8gch, RMSD = 1.04Å
Sequence: GAW



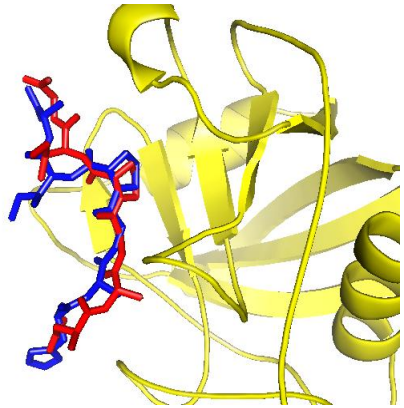
PDB ID: 1b32, RMSD = 0.51Å
Sequence: KMK



PDB ID: 1dkx, RMSD = 1.35Å
Sequence: NRLLLTG



PDB ID: 1awq, RMSD = 1.50Å
Sequence: HAGPIA

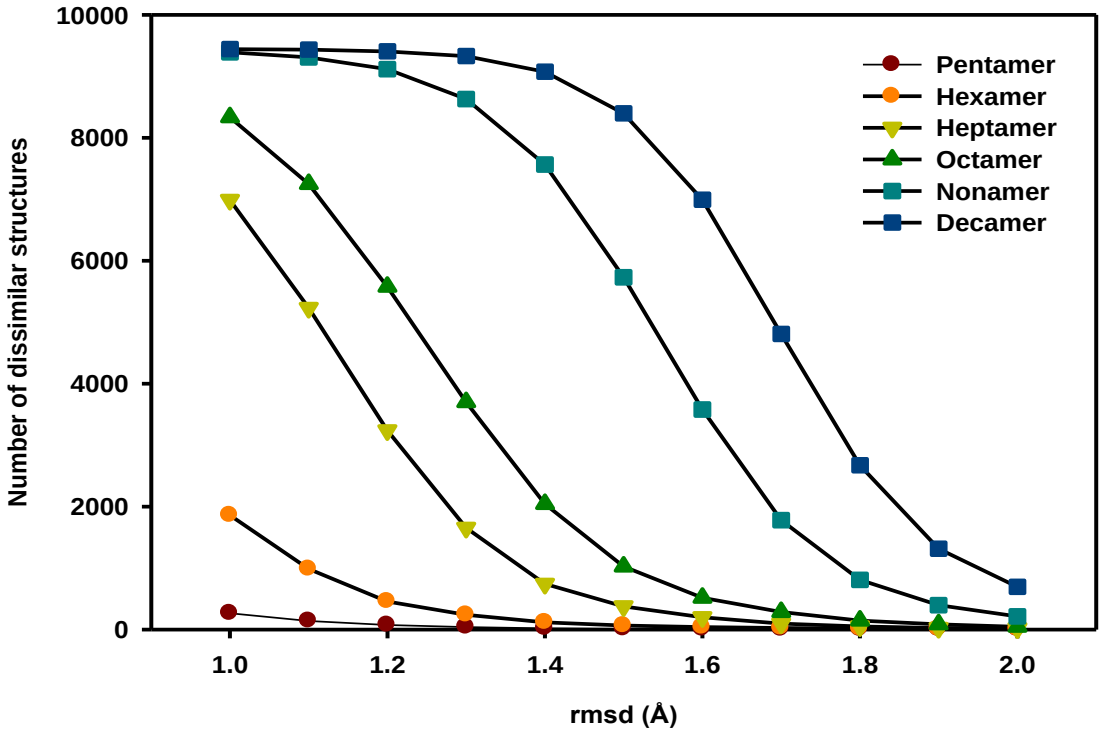


Blue – predicted structure
Red – Crystal structure

JCAMD, 22, 815, 2008

Applications – Structure density

- **Estimation of the numbers and density of low-energy structures in the conformational landscape of proteins.**



The number of mutually dissimilar structures found at different rmsd cut-offs for each peptide in 10,000 MOLS structures using the ECEPP/3 potential.

$$m = a \exp(bn)$$

PLoS One, 2009

Acknowledgements

Z A Rafi, K Vengadesan, J Arunachalam,
V Kanakasabai, P Arun Prasad

CSIR; DST; DBT; UGC

www.unom.ac.in/Gautham_mols.pdf

Thank You