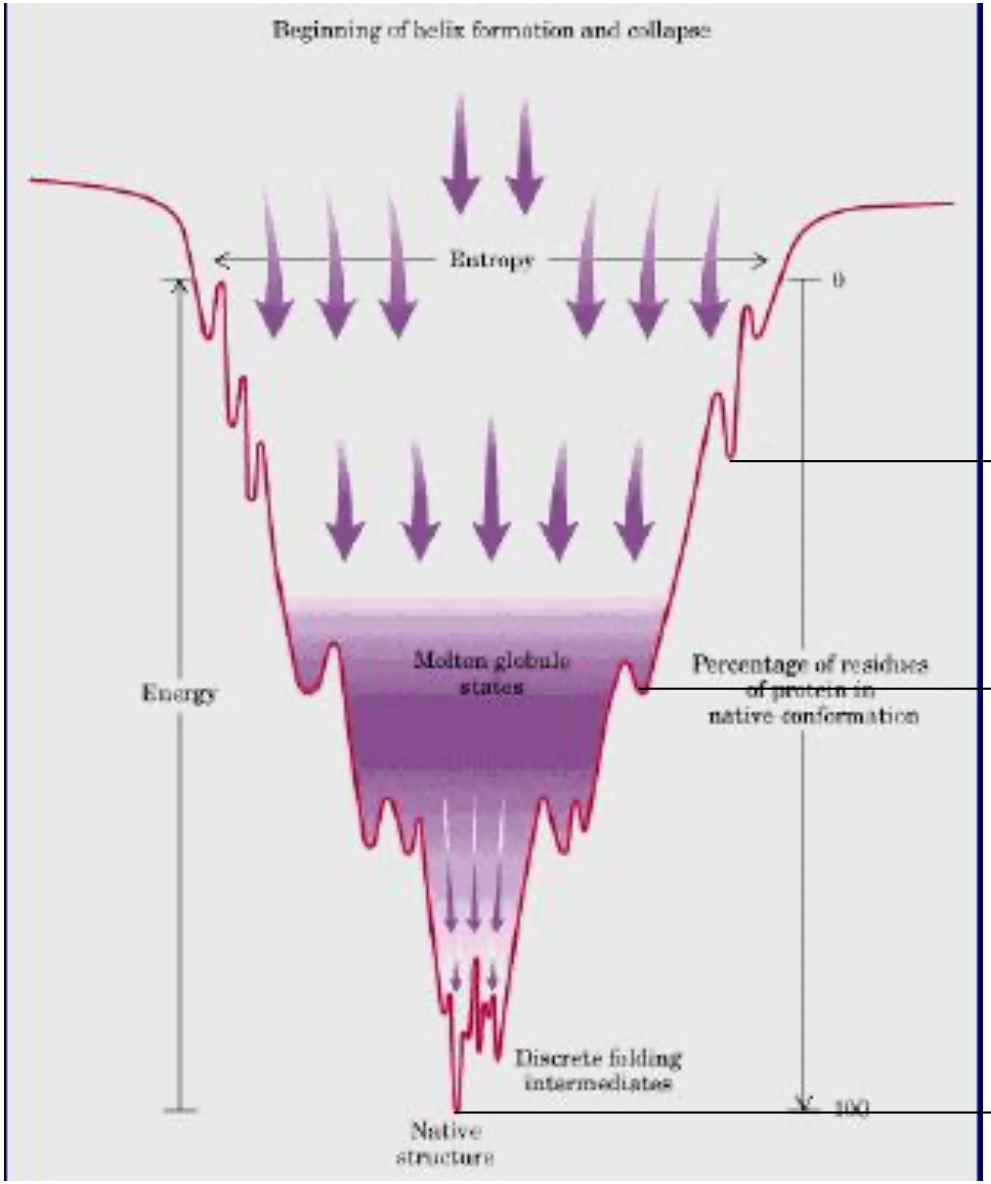
Applications of MM in *Ab Initio* PSP

- **Basic idea**

  Anfinsen's theory: Protein native structure corresponds to the state with the lowest free energy of the protein-solvent system.

- **General procedures**
  - **Potential function**
    - Evaluate the energy of protein conformation
    - Select native structure
  - **Conformational search algorithm**
    - To produce new conformations
    - Search the potential energy surface and locate the global minimum (native conformation)

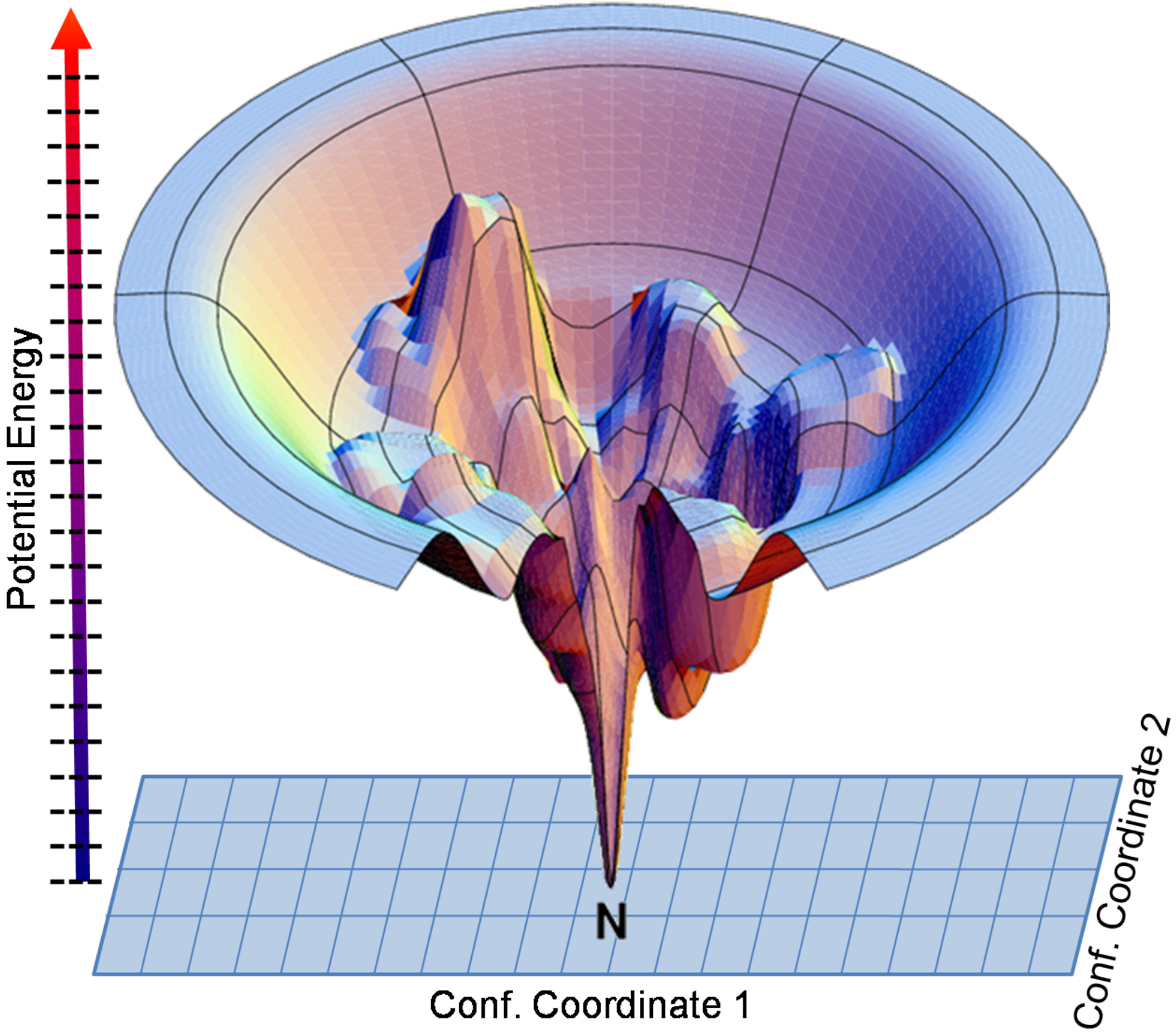
Protein Folding Funnel



Local mimina

Global minimum

Native Structure


Potential Energy
N
Conf. Coordinate 1
Conf. Coordinate 2

Potential Functions for PSP

- **Potential function**
  - **Physical based energy function**

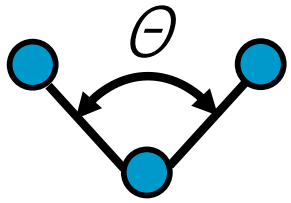
    Empirical *all-atom* forcefields: CHARMM, AMBER, ECEPP-3, GROMOS, OPLS

    Parameterization: Quantum mechanical calculations, experimental data

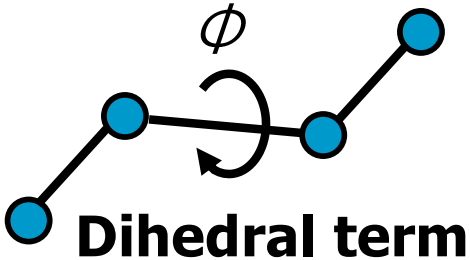
    Simplified potential: UNRES (*united residue*)
  - **Solvation energy**
    - Implicit solvation model: Generalized Born (GB) model, surface area based model
    - Explicit solvation model: TIP3P (computationally expensive)

General Form of All-atom Forcefields

Bond stretching term



Angle bending term



Dihedral term

$$V_{\text{total}} = \sum_{\text{bonds}} K_b\left(r - r_0\right)^2 + \sum_{\text{angles}} K_\theta\left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} K_\phi\left[1 + \cos\left(n\phi - \gamma\right)\right]$$

$$+ \sum_{\text{Hbonds}}\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + \sum_{\substack{\text{van der Waals} \\ i,j \text{ pairs}}}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{\substack{\text{electrostatic} \\ i,j \text{ pairs}}}\frac{q_i q_j}{\varepsilon r_{ij}}$$

The most time demanding part.

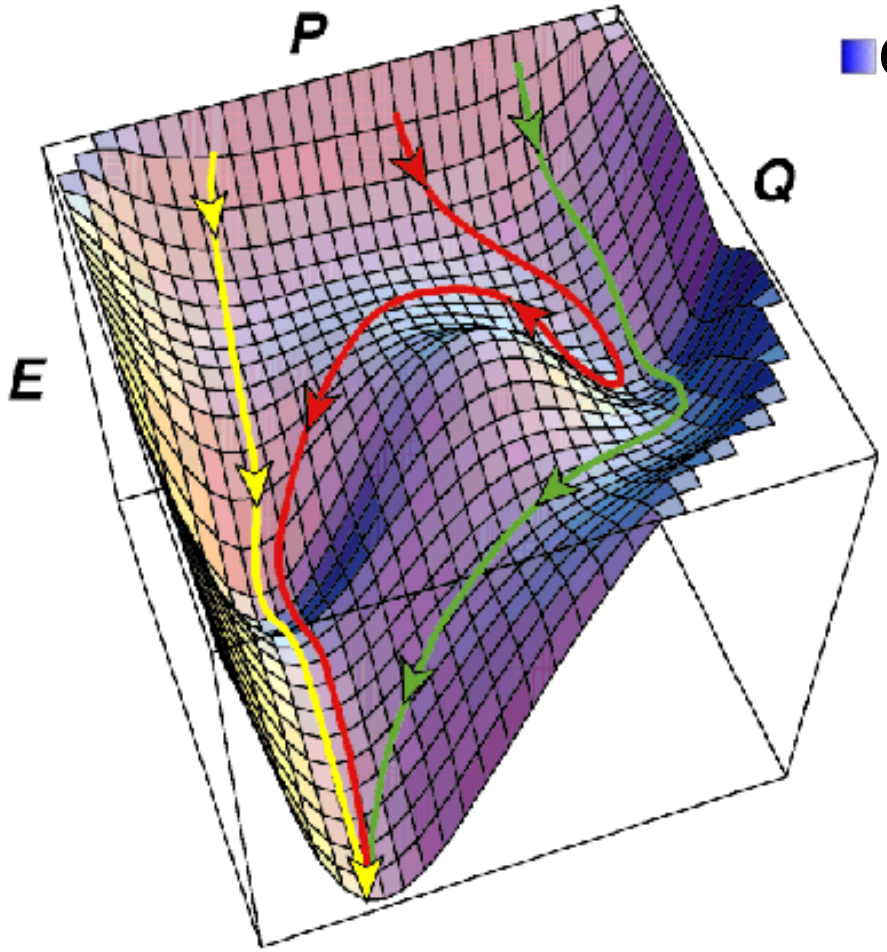
H-bonding term

Van der Waals term

Electrostatic term

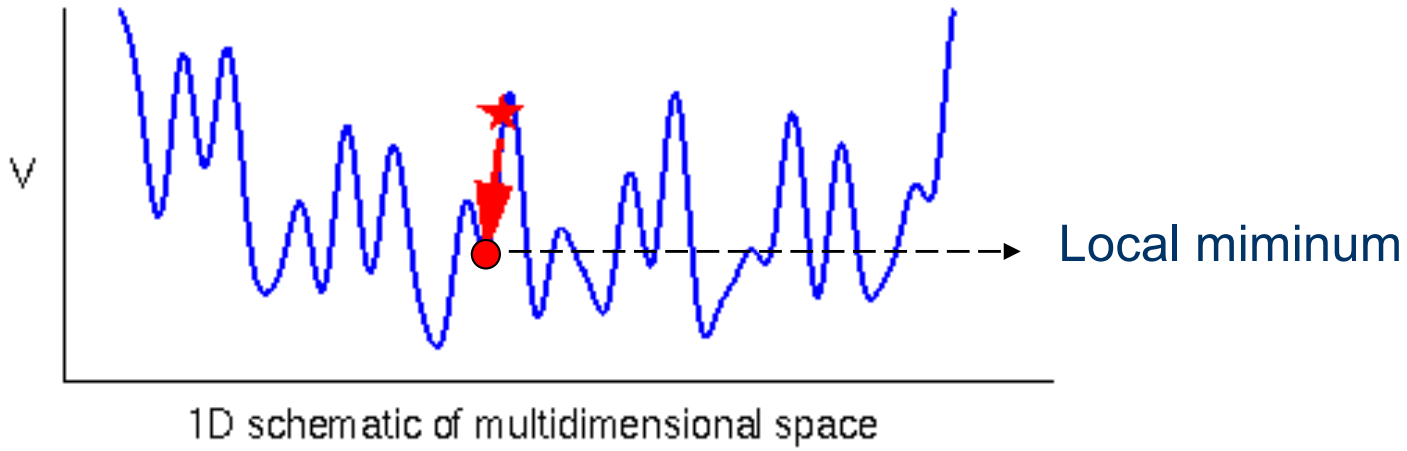
Search Potential Energy Surface

- We are interested in minimum points on Potential Energy Surface (PES)



- **Conformational search techniques**
  - Energy Minimization
  - Monte Carlo
  - Molecular Dynamics
  - Others: Genetic Algorithm, Simulated Annealing

Energy Minimization



- Energy minimization
- Methods
  - First-order minimization: Steepest descent, Conjugate gradient minimization
  - Second derivative methods: Newton-Raphson method
  - Quasi-Newton methods: L-BFGS

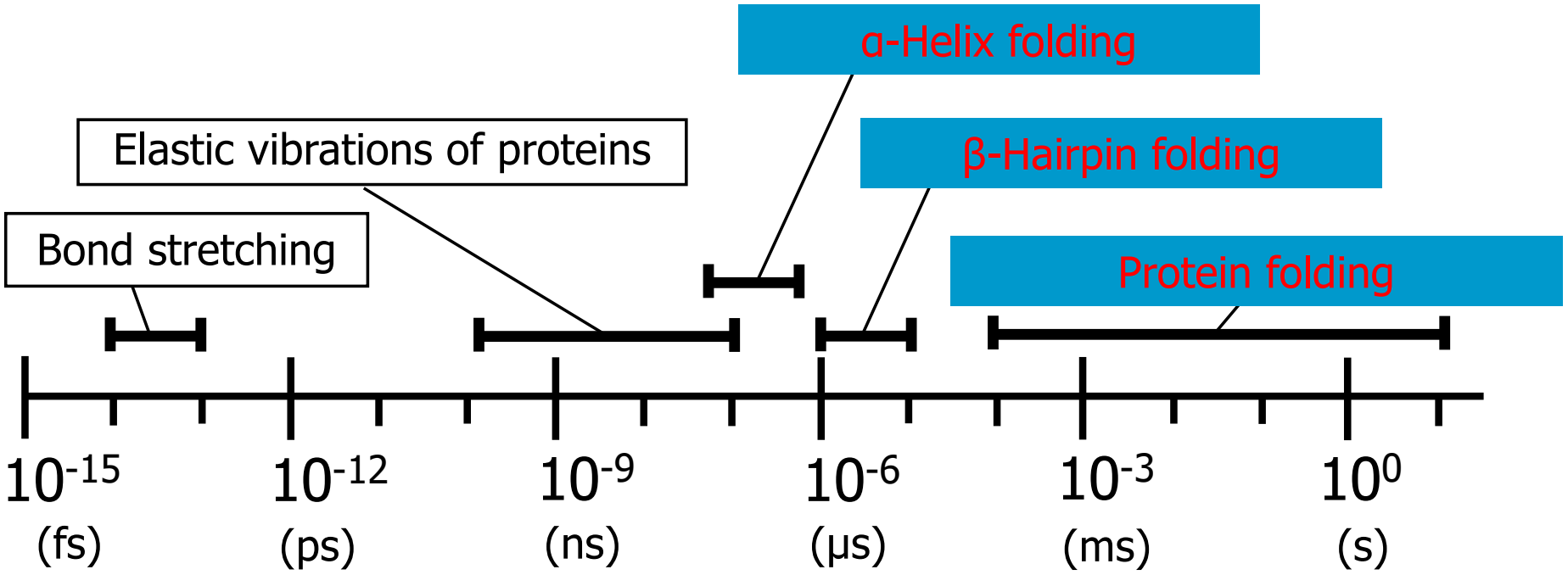
Molecular Dynamics

- **MD Software**
  - CHARMM (Chemistry at HARvard Molecular Mechanics) is a program for macromolecular simulations, including energy minimization, molecular dynamics and Monte Carlo simulations.
  - NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems.

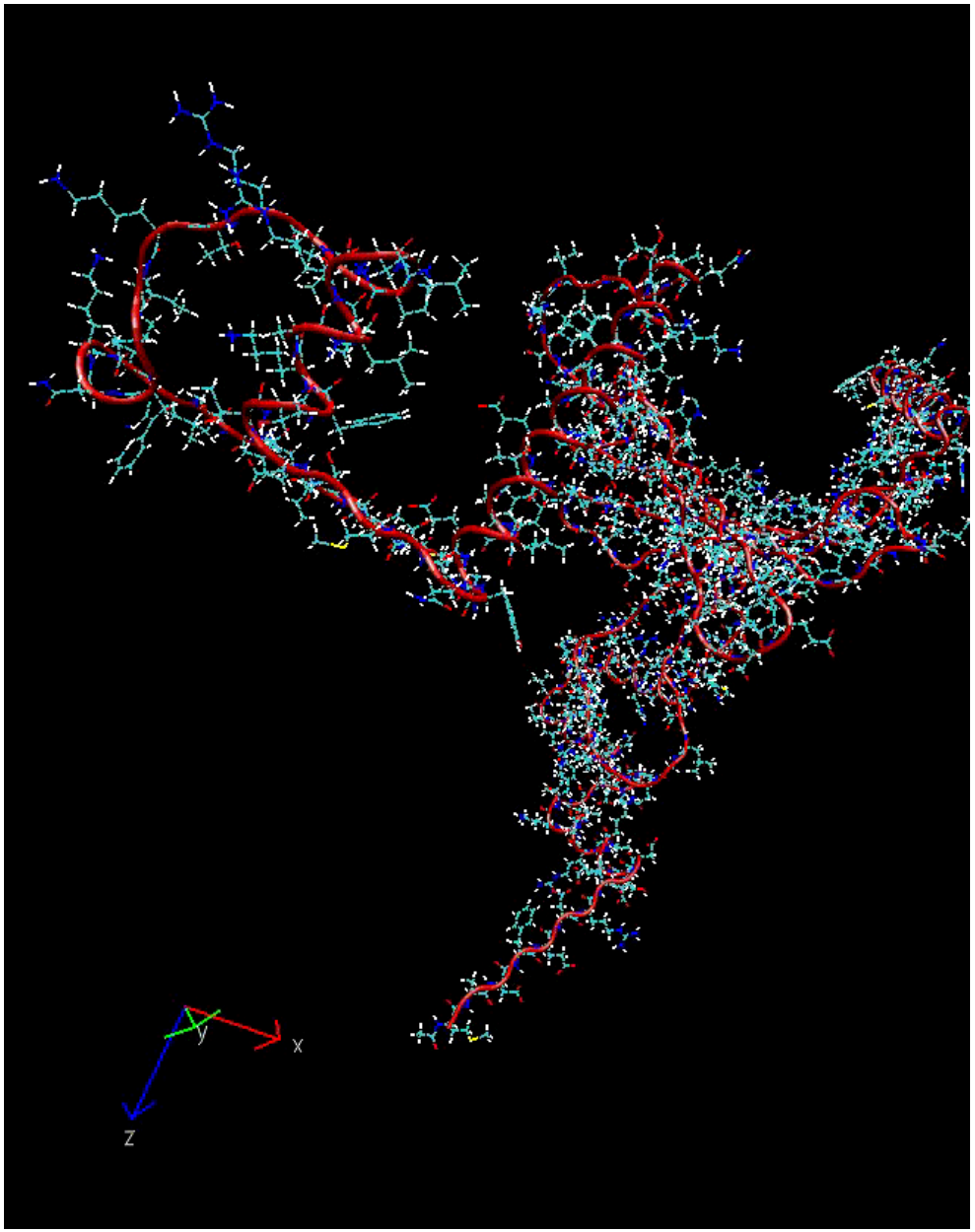
    http://www.ks.uiuc.edu/Research/namd/
- **Application in PSP**
  - Advantage: Deterministic; Provide details of the folding process
  - Limitation: The protein folding reactions take place at ms level, which is at the limit of accessible simulation times

It is still difficult to simulate a whole process of a protein folding using the conventional MD method.

Time Scales of Protein Motions and MD



MD Time Scale



MD is fun!

A small protein folding movie: simulated with NAMD/VMD

Other Conformational Search Algorithms

- **Global optimization algorithms**

  "Optimization" refers to trying to find the global energy minimum of a potential surface.

  - Genetic Algorithm (GA)
  - Simulated Annealing (SA)
  - Tabu Search (TS)
  - Ant Colony Optimization (ACO)

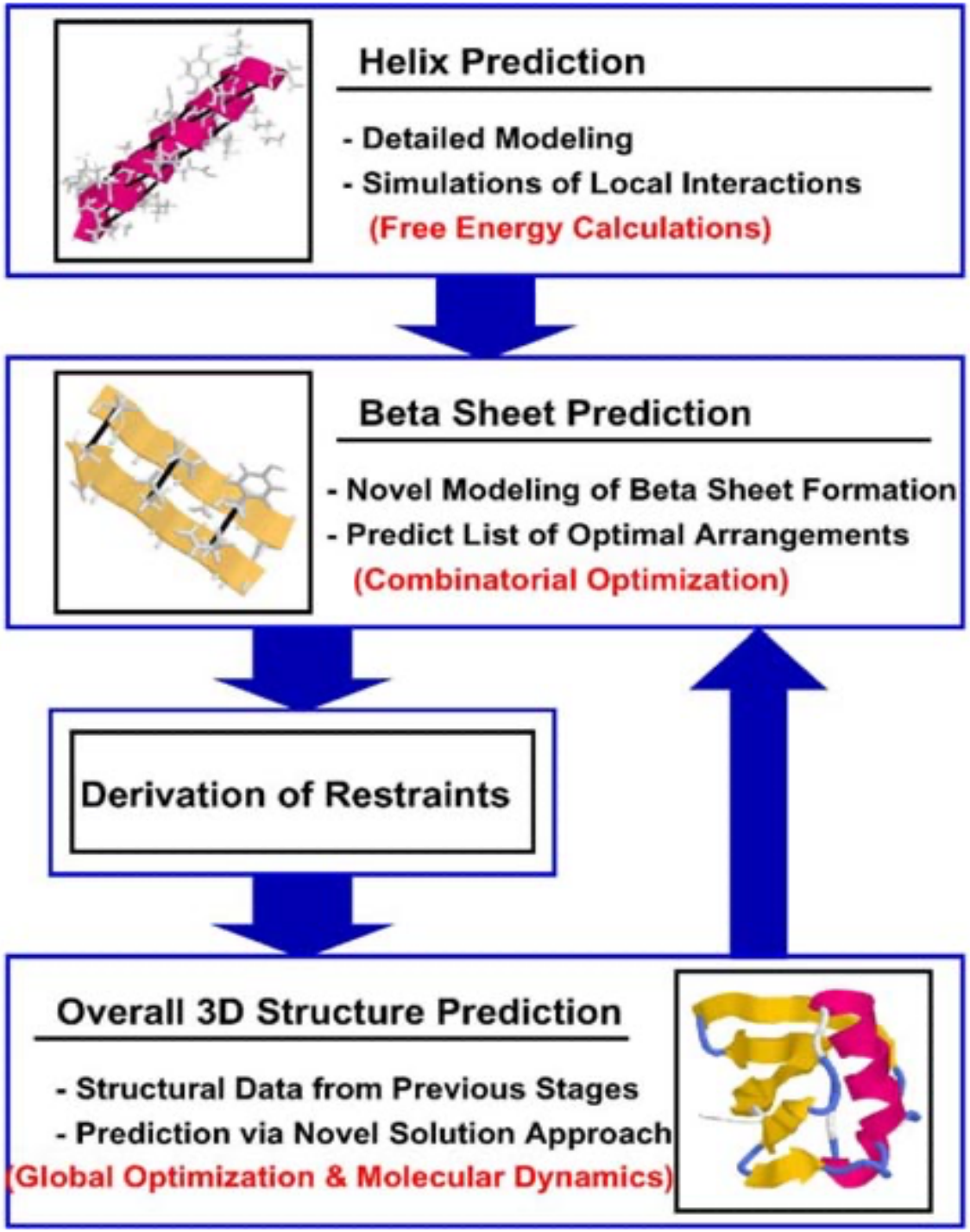
- A model system: Lennard Jones clusters

Applications of MM methods in PSP

- **Application in PSP**
  - Combination of several conformational search techniques
  - Recent developments
    - ➢ Simplified force field: united residue force field
    - ➢ Segment assembly

      Secondary structure prediction are quite reliable, so conformation can be produced by assemble the segments
- ***Ab initio* PSP software**
  - **Rosetta** is a five-stage fragment insertion Metropolis Monte Carlo method
  - **ASTRO-FOLD** is a combination of the deterministic αBB global optimization algorithm, and a Molecular Dynamics approach in torsion angle space
  - **LINUS** uses a Metropolis Monte Carlo algorithm and a simplified physics-based force field

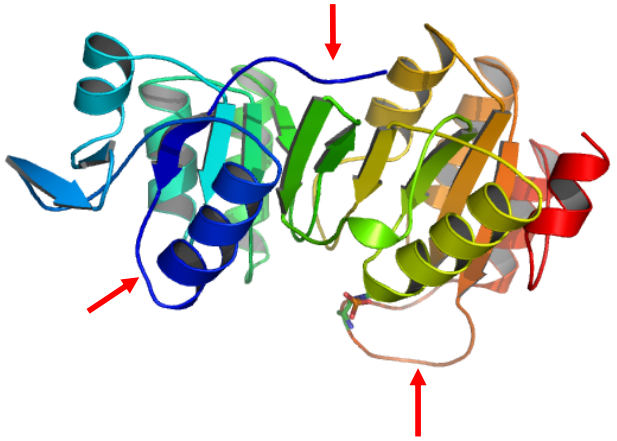
ASTRO-FOLD


Helix Prediction
- Detailed Modeling
- Simulations of Local Interactions
(Free Energy Calculations)
Beta Sheet Prediction
- Novel Modeling of Beta Sheet Formation
- Predict List of Optimal Arrangements
(Combinatorial Optimization)
Derivation of Restraints
Overall 3D Structure Prediction
- Structural Data from Previous Stages
- Prediction via Novel Solution Approach
(Global Optimization & Molecular Dynamics)

Ab Initio Protein Loop Prediction

- **Protein loop**

  Protein loops are polypeptides connecting more rigid structural elements of proteins like helices and strands.



- **Challenge in Loop Structure Prediction**
  - Loop is important to protein folding and protein function even though their size is small, usually <20 residues
  - Loops exhibit greater structural variability than helices and strands
  - Loop prediction is often a limiting factor on fold recognition methods