Stochastic Roadmap Simulation:
Efficient Representation and Algorithms for the Analysis of Molecular Motion

M. Serkan Apaydin[1]

Douglas L. Brutlag[1]

Carlos E. Guestrin[1]

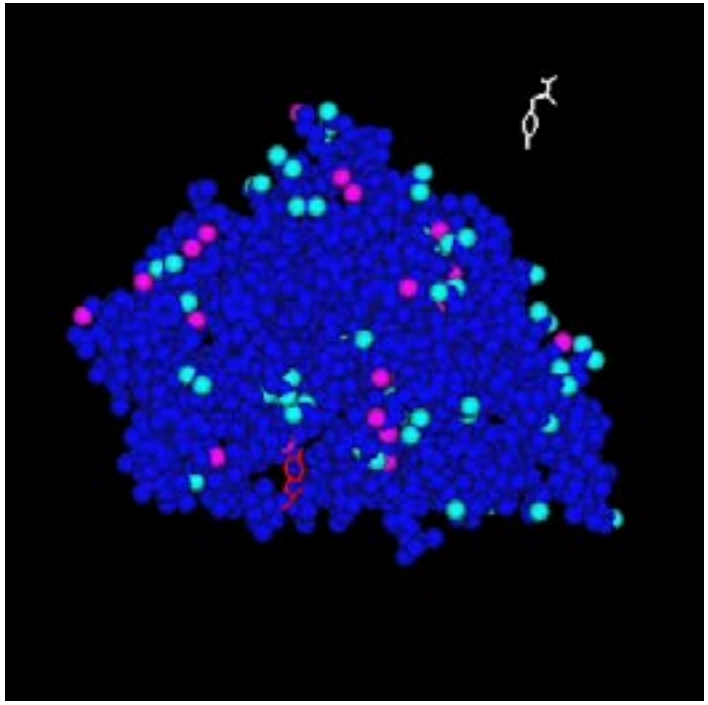
David Hsu[2]

Jean-Claude Latombe[1]

[1]Stanford University [2]University of North Carolina

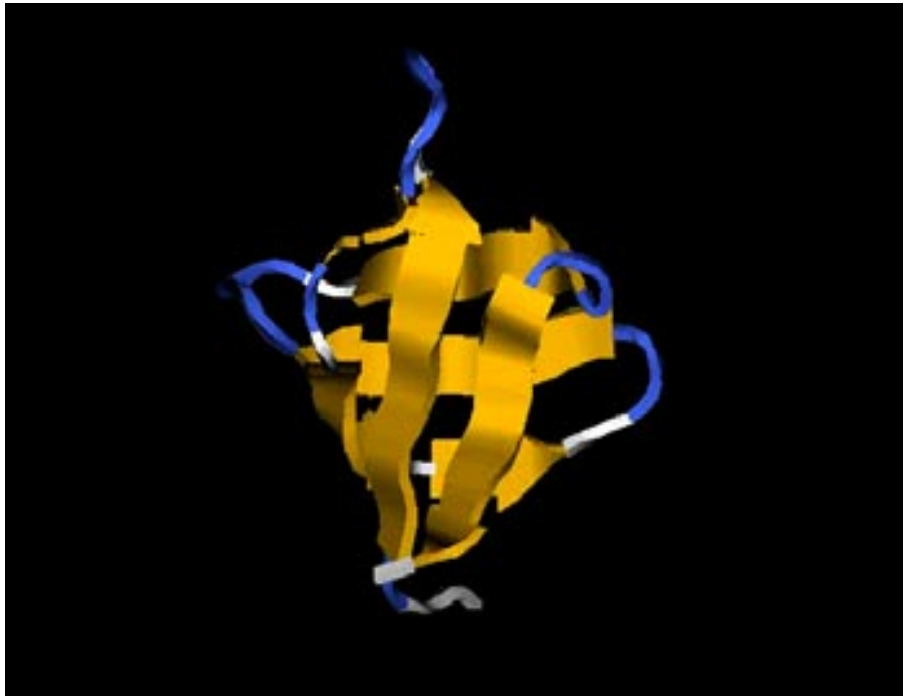
Molecular motion

Lactate Dehydrogenase - NAD and Oxamate

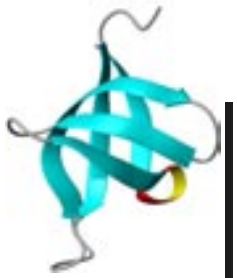


[Singh et al. ‘99]

HIV integrase

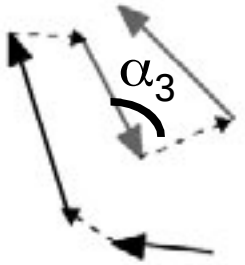


http://foldingathome.stanford.edu

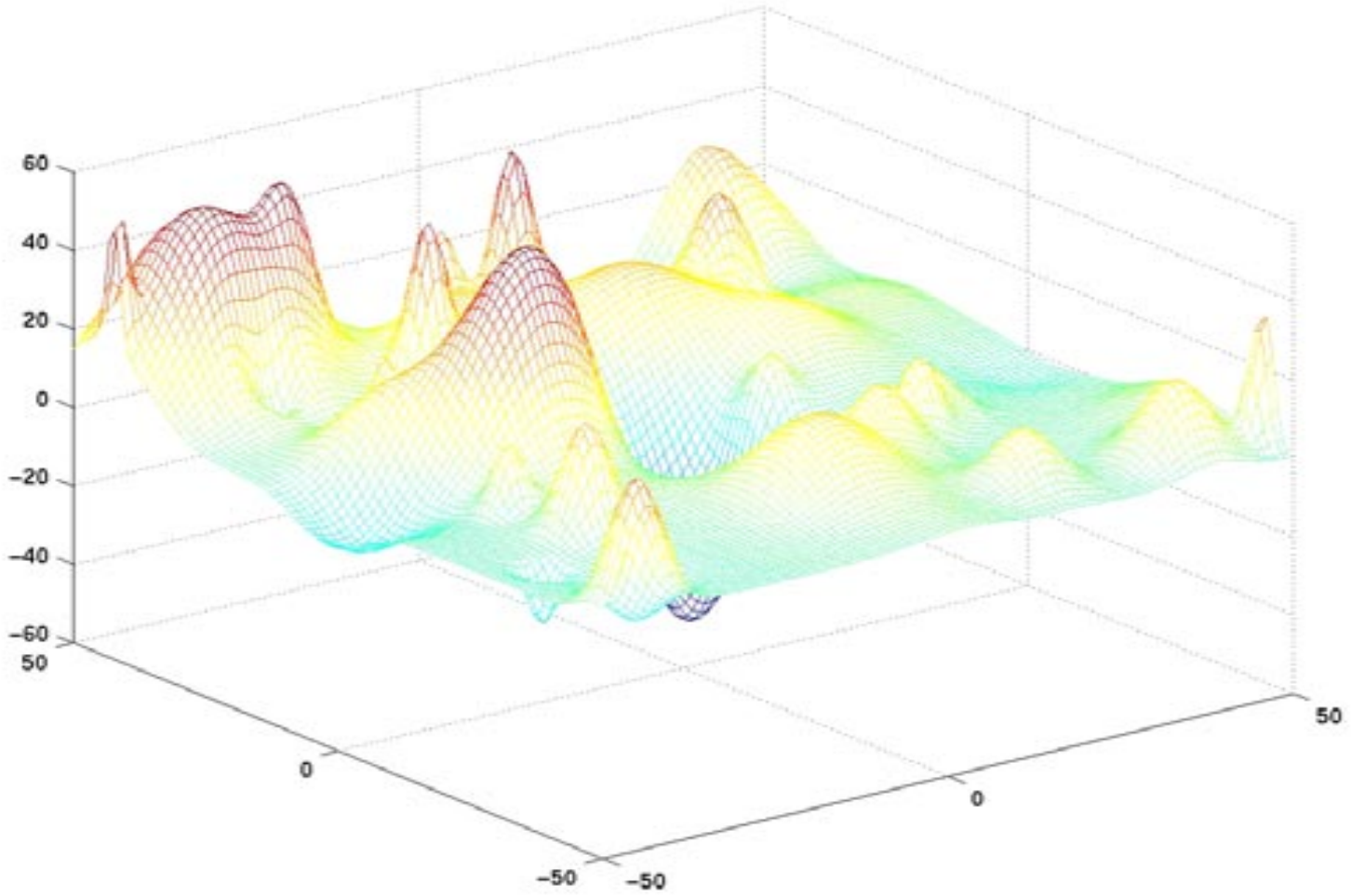


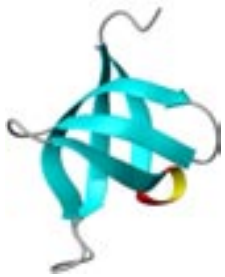
Studying molecular motion

- Molecule represented by parameters;
- Energy for each conformation;
- Use Monte Carlo(MC) or Molecular Dynamics(MD) simulation to study molecular motion.

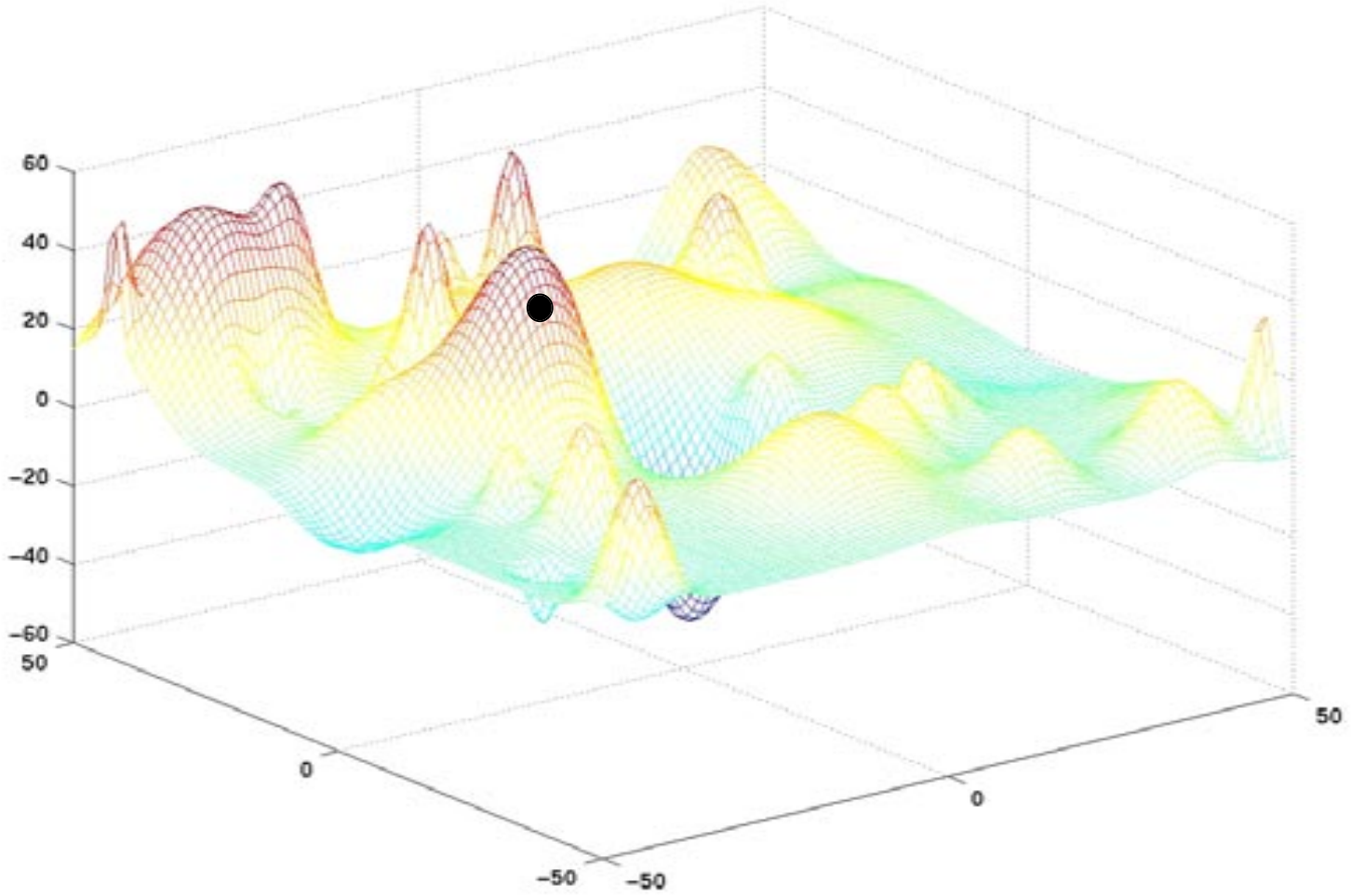


MC simulation

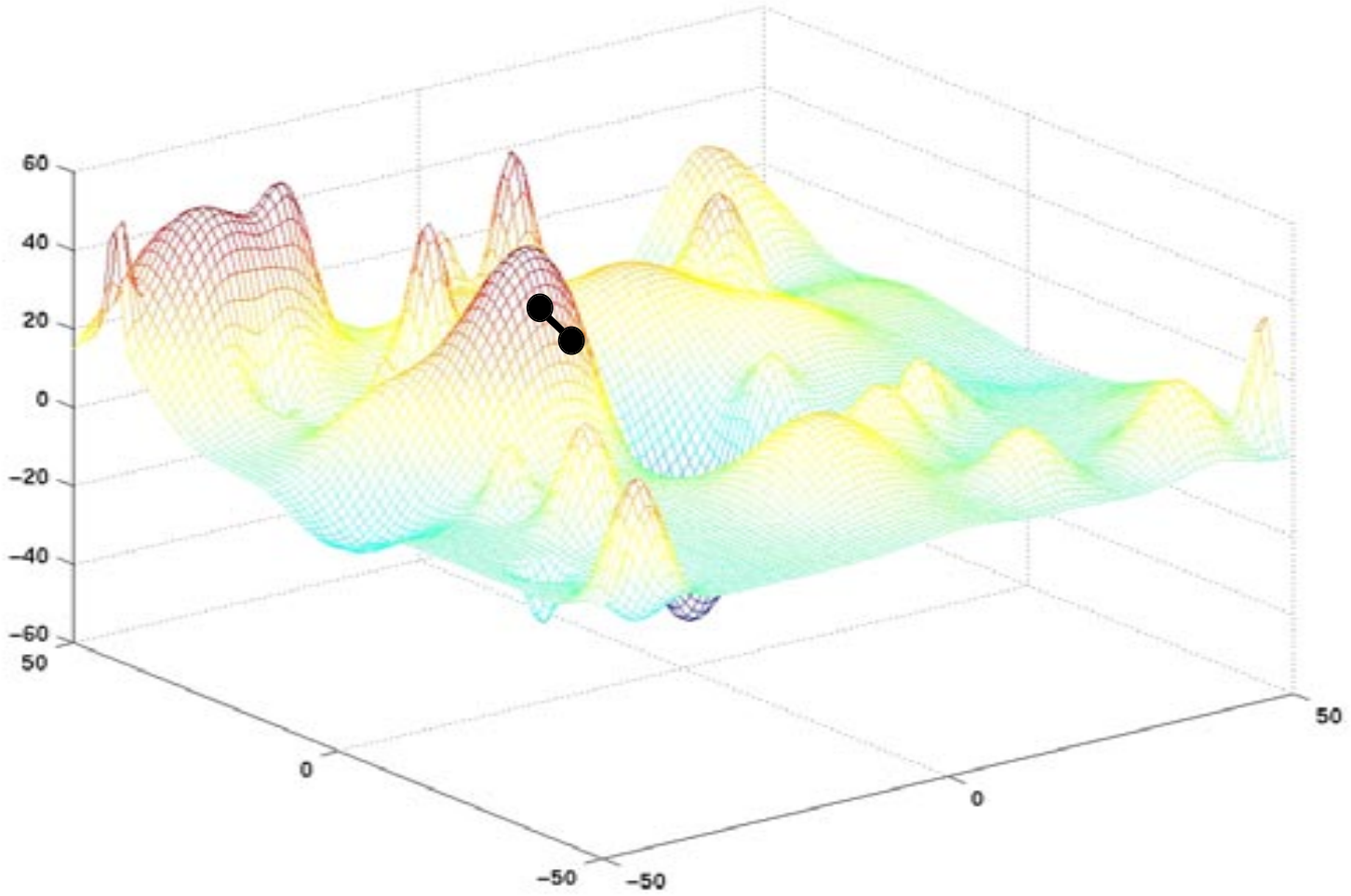




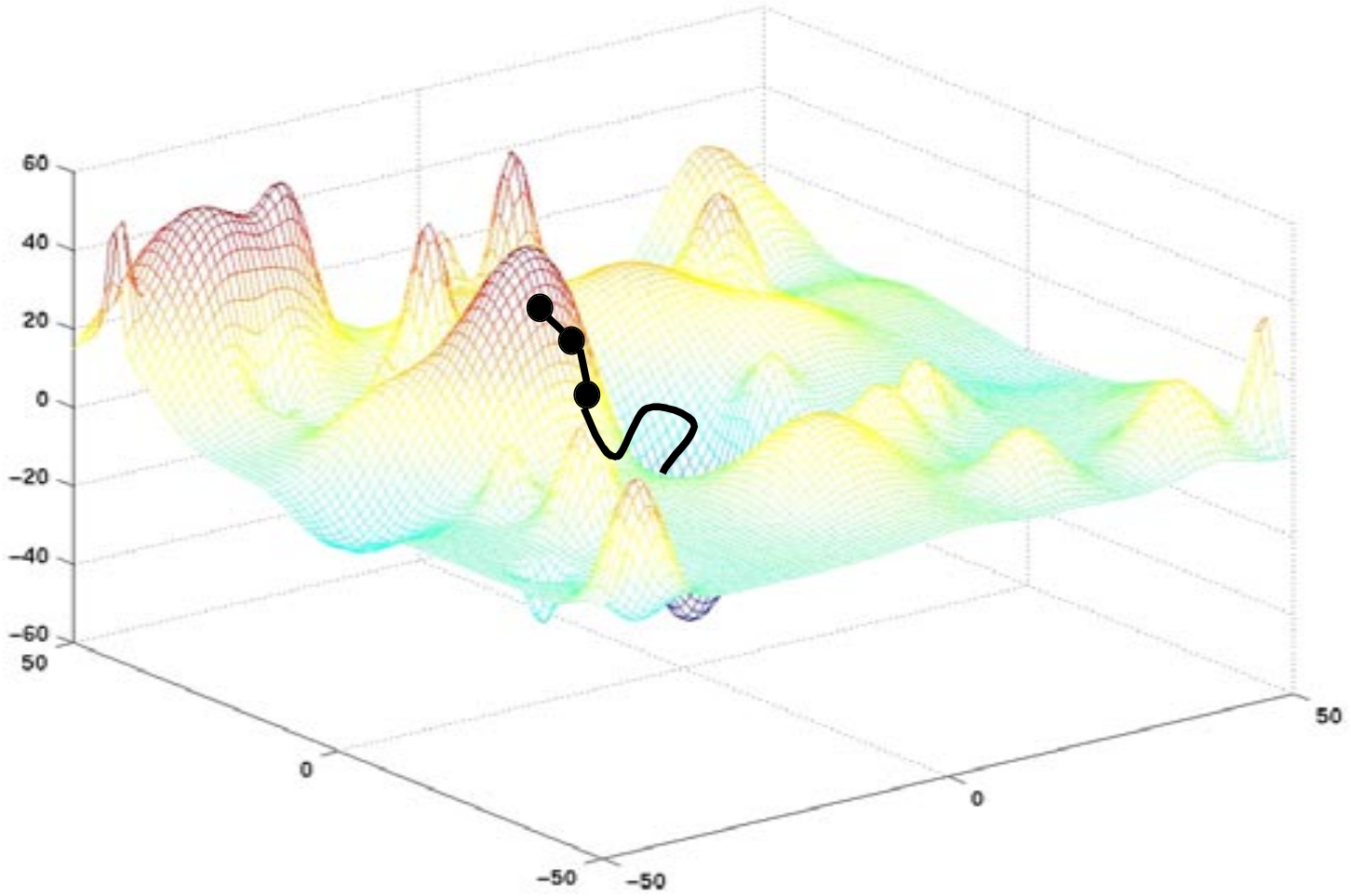
MC simulation



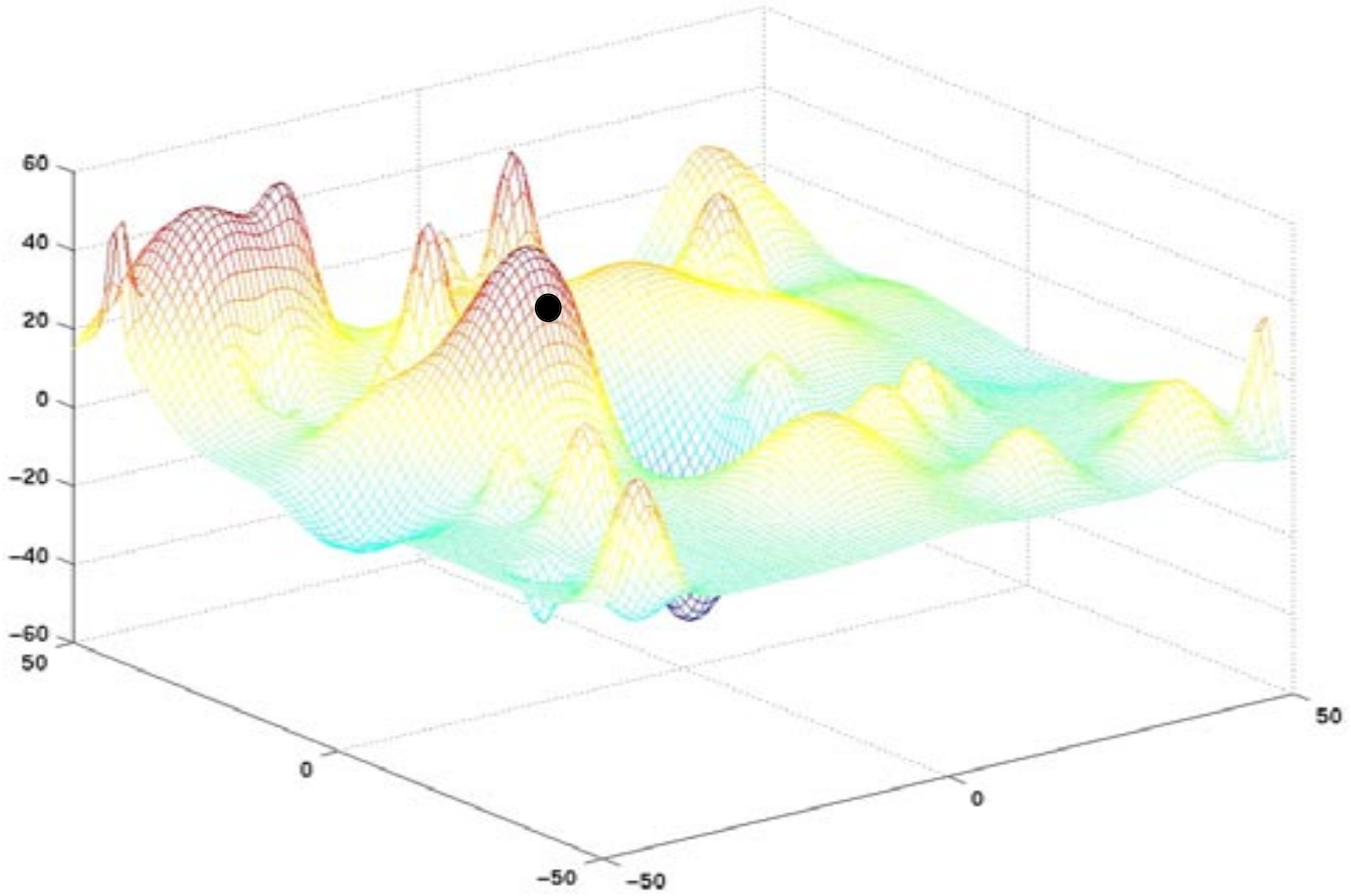
MC simulation



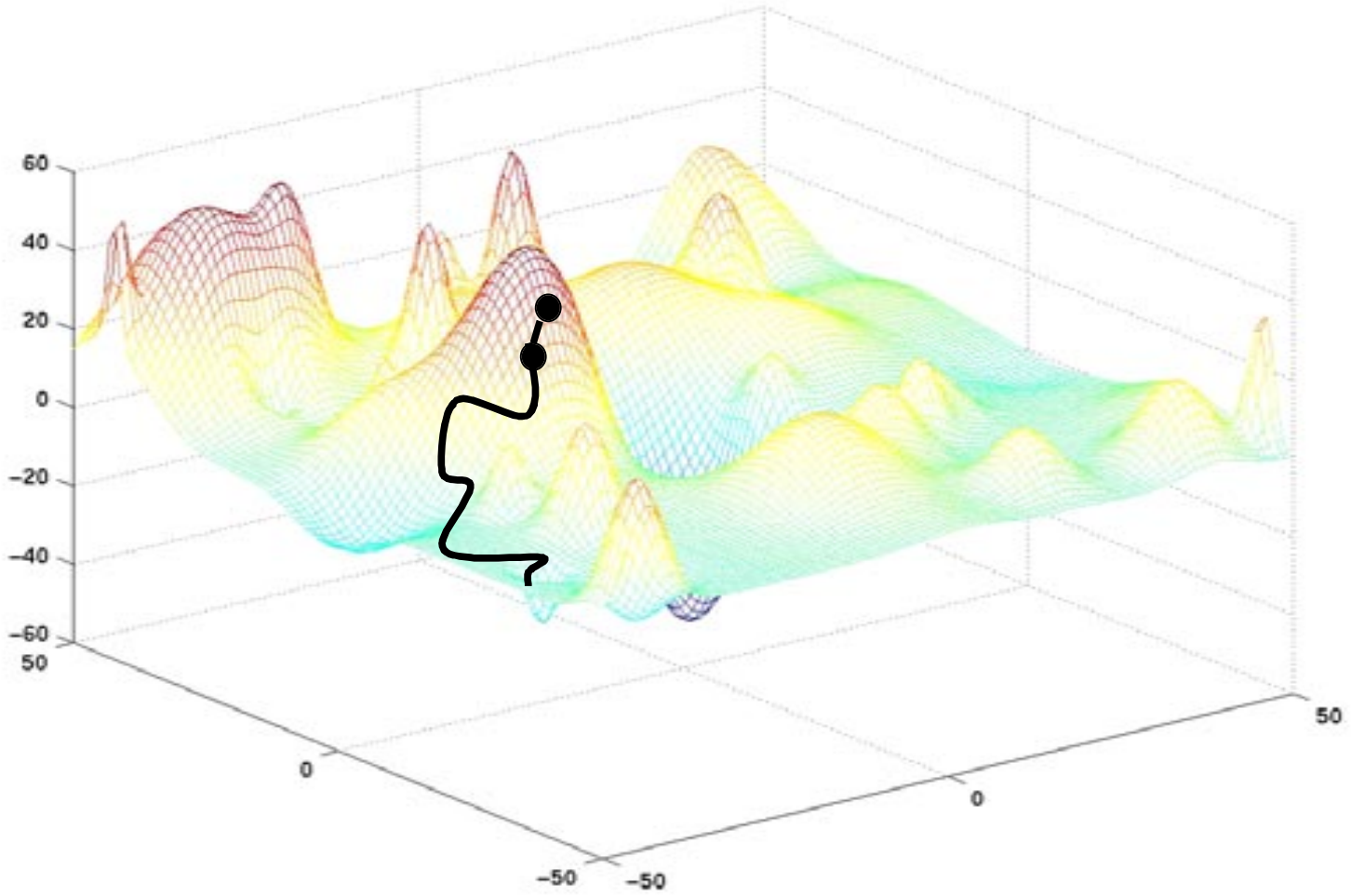
MC simulation

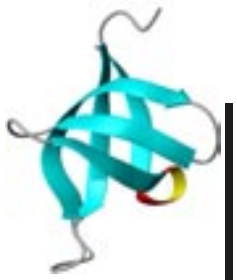


MC simulation



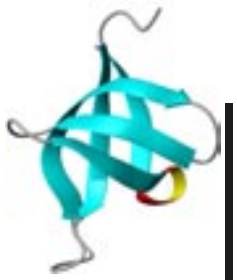
MC simulation



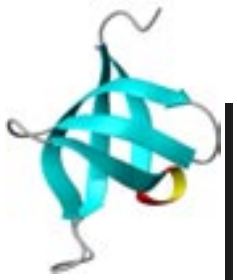


Problems with MC simulation

- Expensive to compute;
- Gets stuck at local minimum;
- Single path at a time!
- Similar problems with MD simulation.

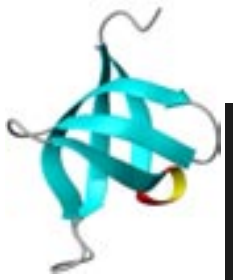


Stochastic Roadmap Simulation (SRS)



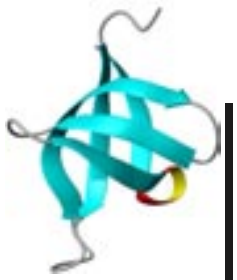
Stochastic Roadmap Simulation (SRS)

- Multiple paths at once;



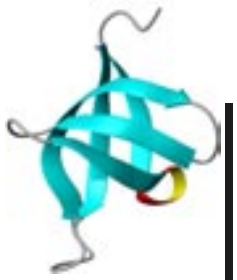
Stochastic Roadmap Simulation (SRS)

- Multiple paths at once;
- No local minimum problem;



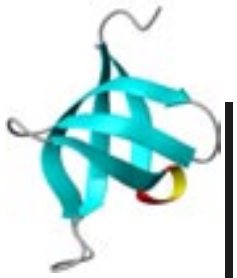
Stochastic Roadmap Simulation (SRS)

- Multiple paths at once;
- No local minimum problem;
- Molecular properties computed analytically;

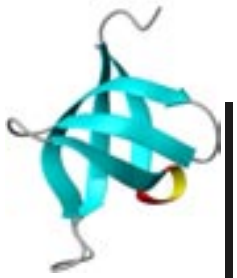


Stochastic Roadmap Simulation (SRS)

- Multiple paths at once;
- No local minimum problem;
- Molecular properties computed analytically;
- Converges to the same distribution as MC.

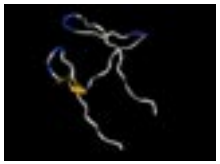


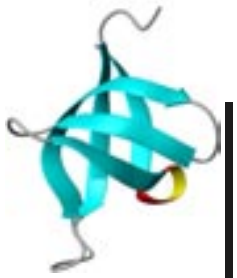
Roadmap construction



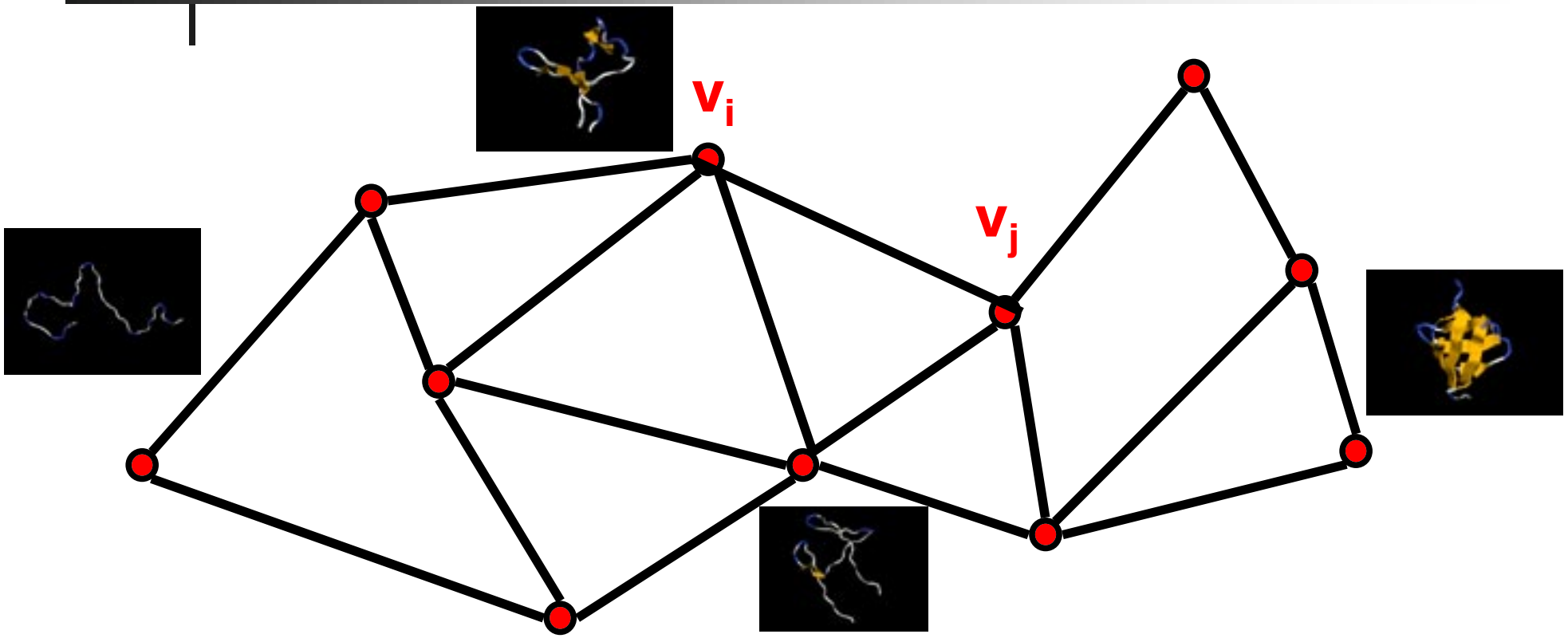
Roadmap construction

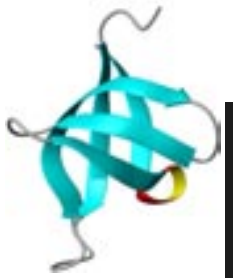
v_i



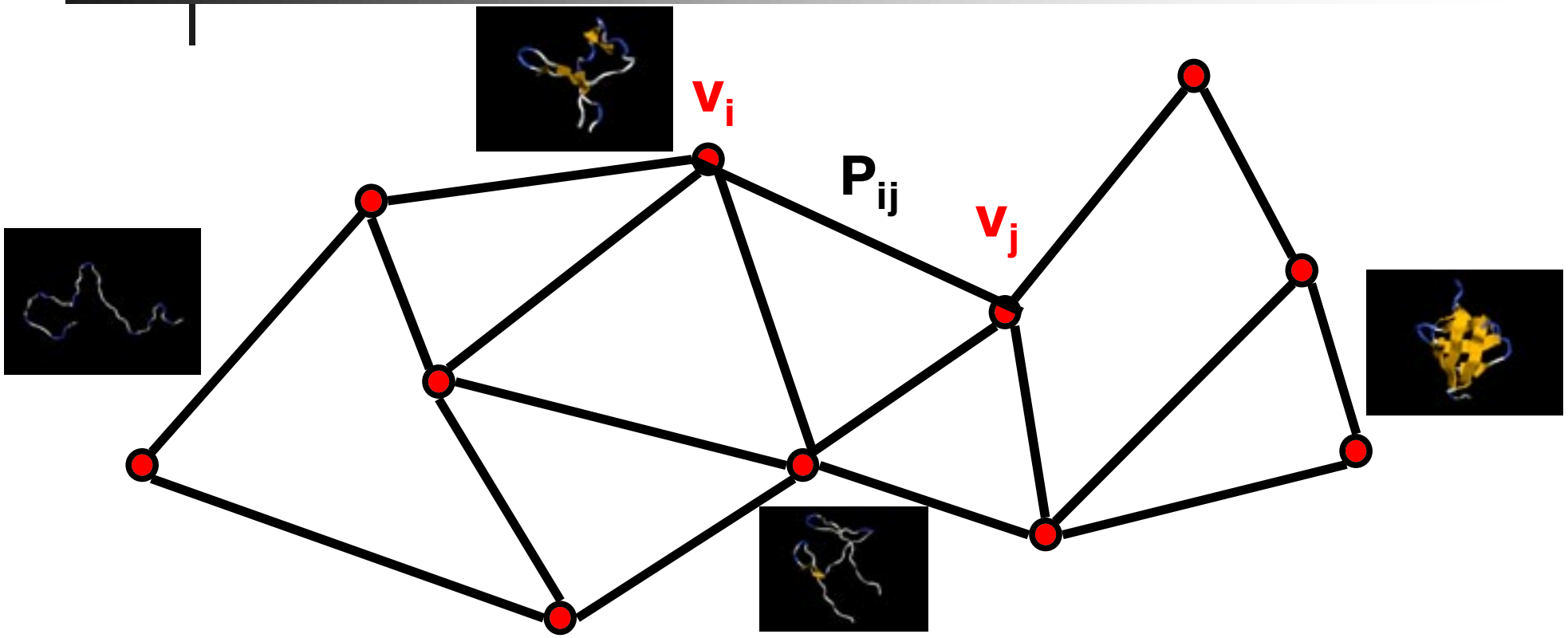


Roadmap construction

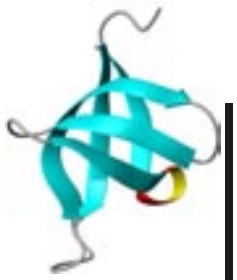




Roadmap construction



- Sample nodes from conformation space;
- Edge weights are probabilities.



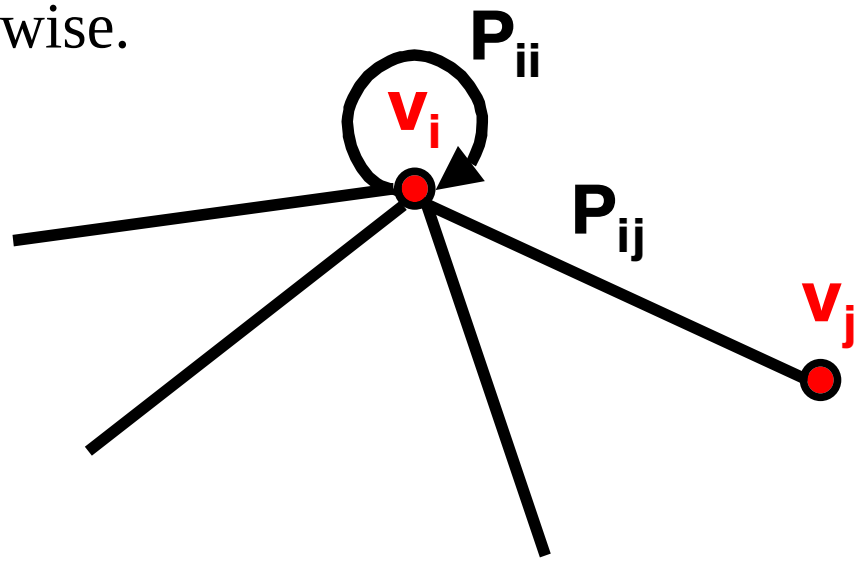
Edge probabilities

Follow Metropolis criteria:

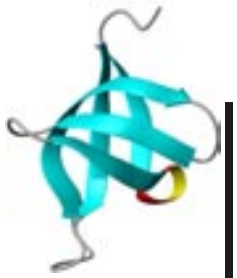
$$P_{ij} = \begin{cases} \dfrac{\exp(-\Delta E_{ij} / k_B T)}{|N_i|}, & \text{if } \Delta E_{ij} > 0; \\ \dfrac{1}{|N_i|}, & \text{otherwise.} \end{cases}$$

Self transition probabilities:

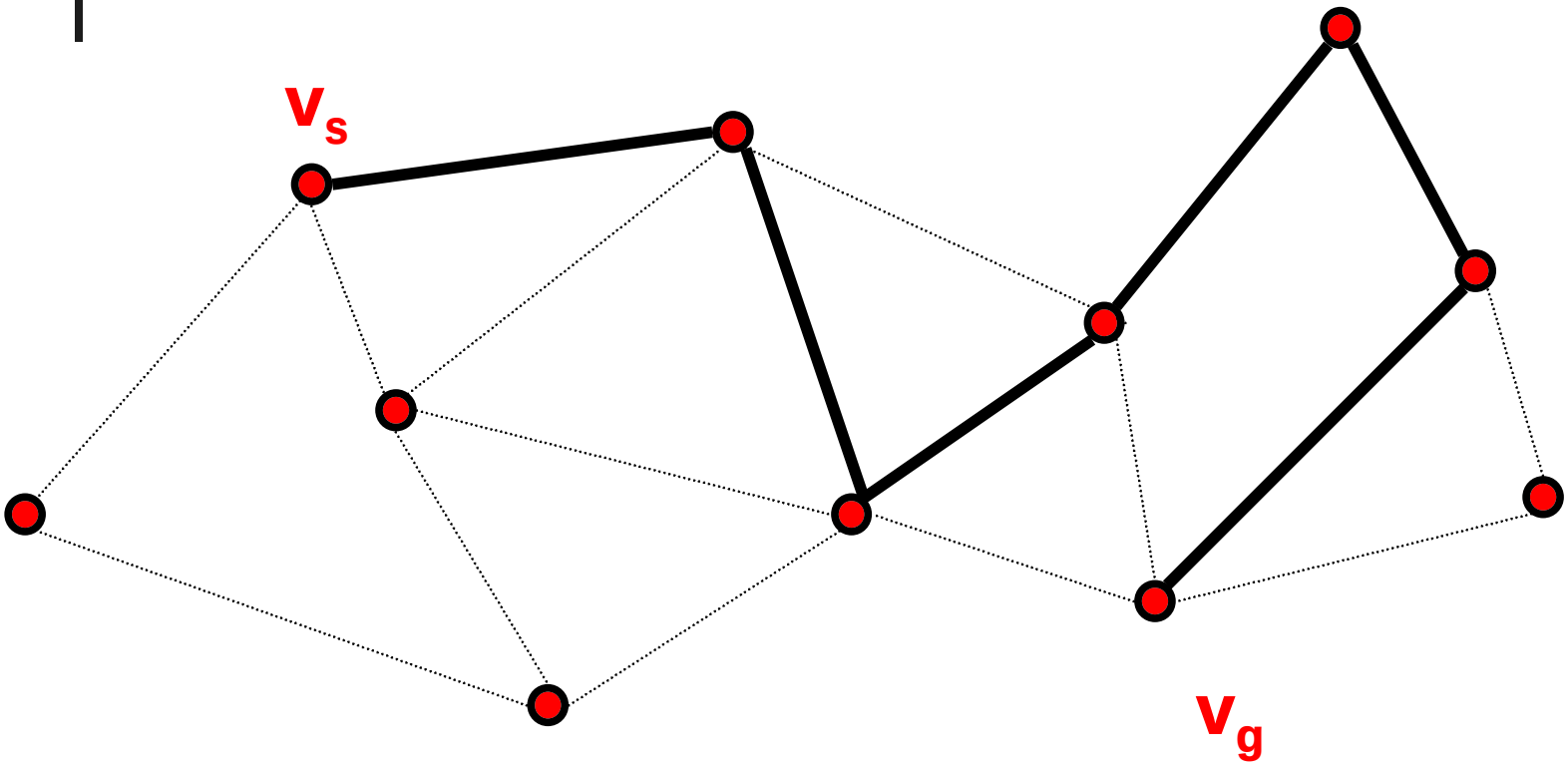
$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}$$

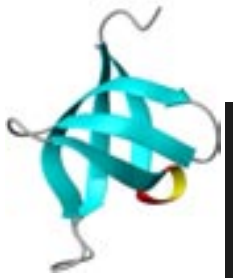


- Correspond to probabilities in Monte Carlo simulation;
- Different from roadmaps in previous work!

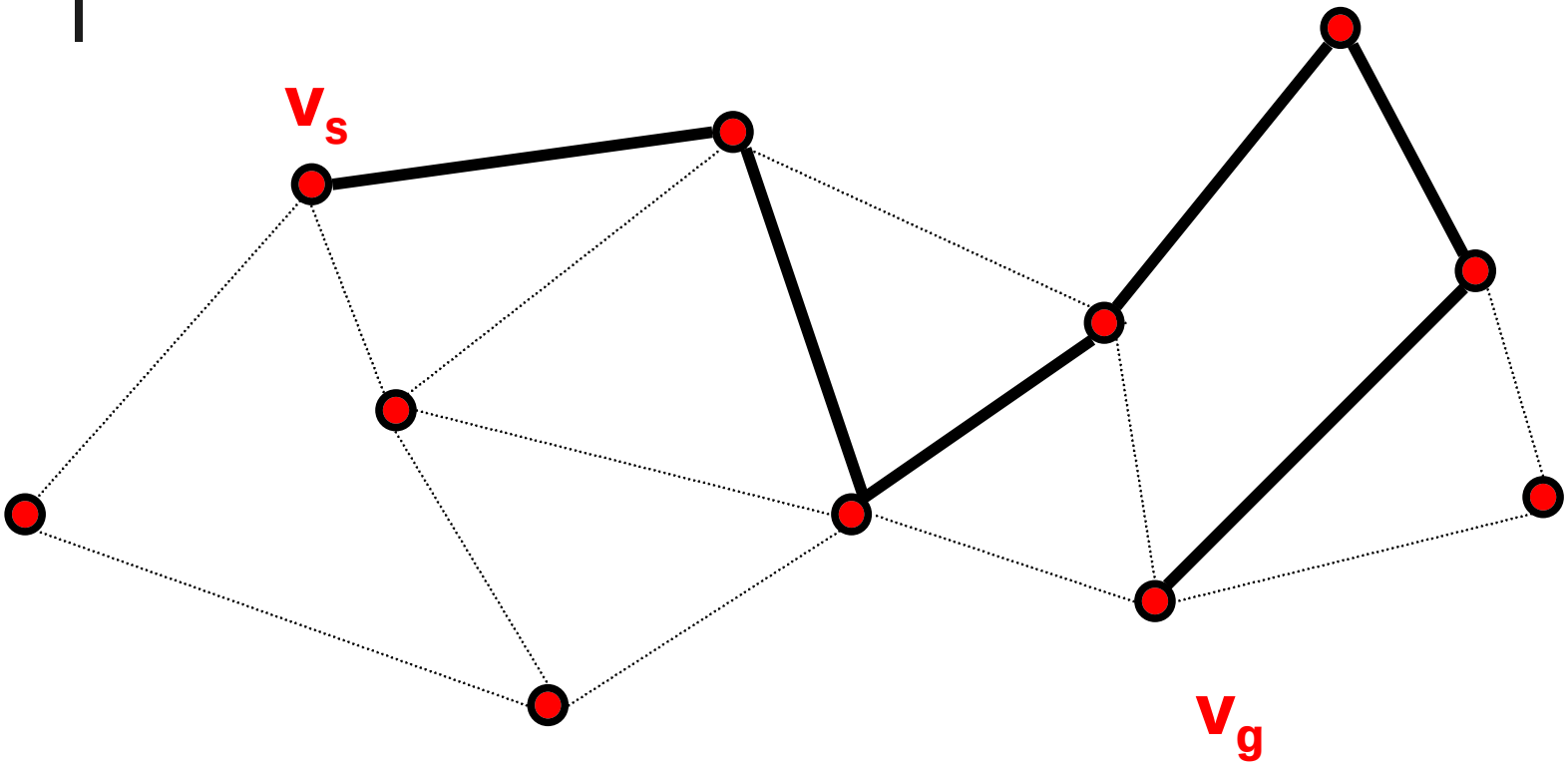


Relationship to MC simulation

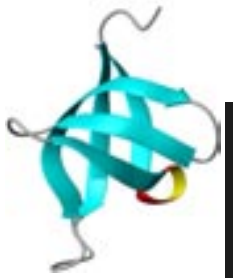




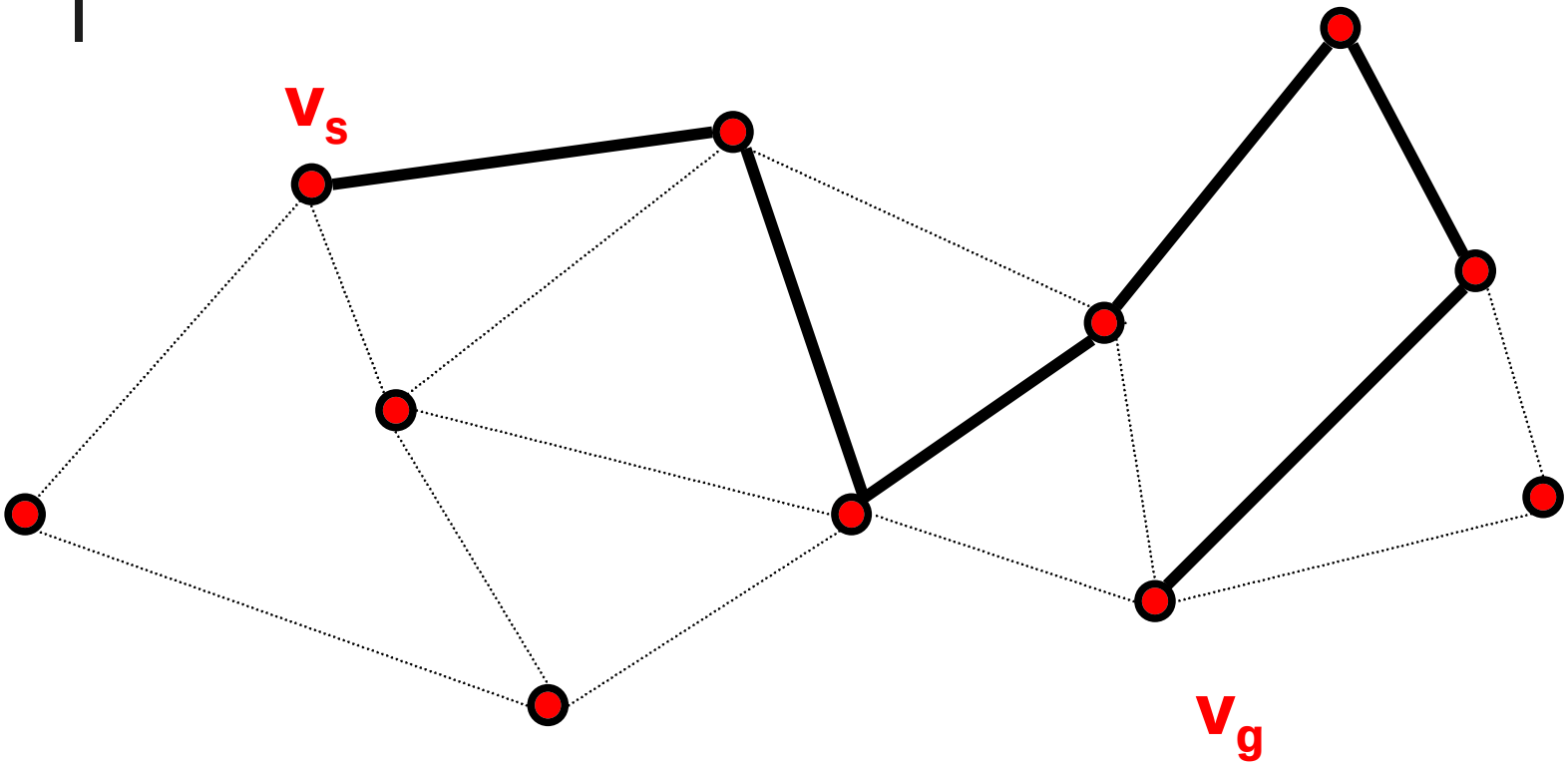
Relationship to MC simulation



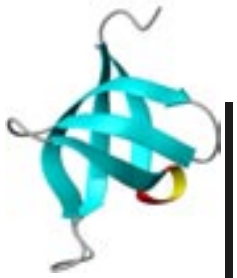
- Each path on graph = a path of MC simulation;



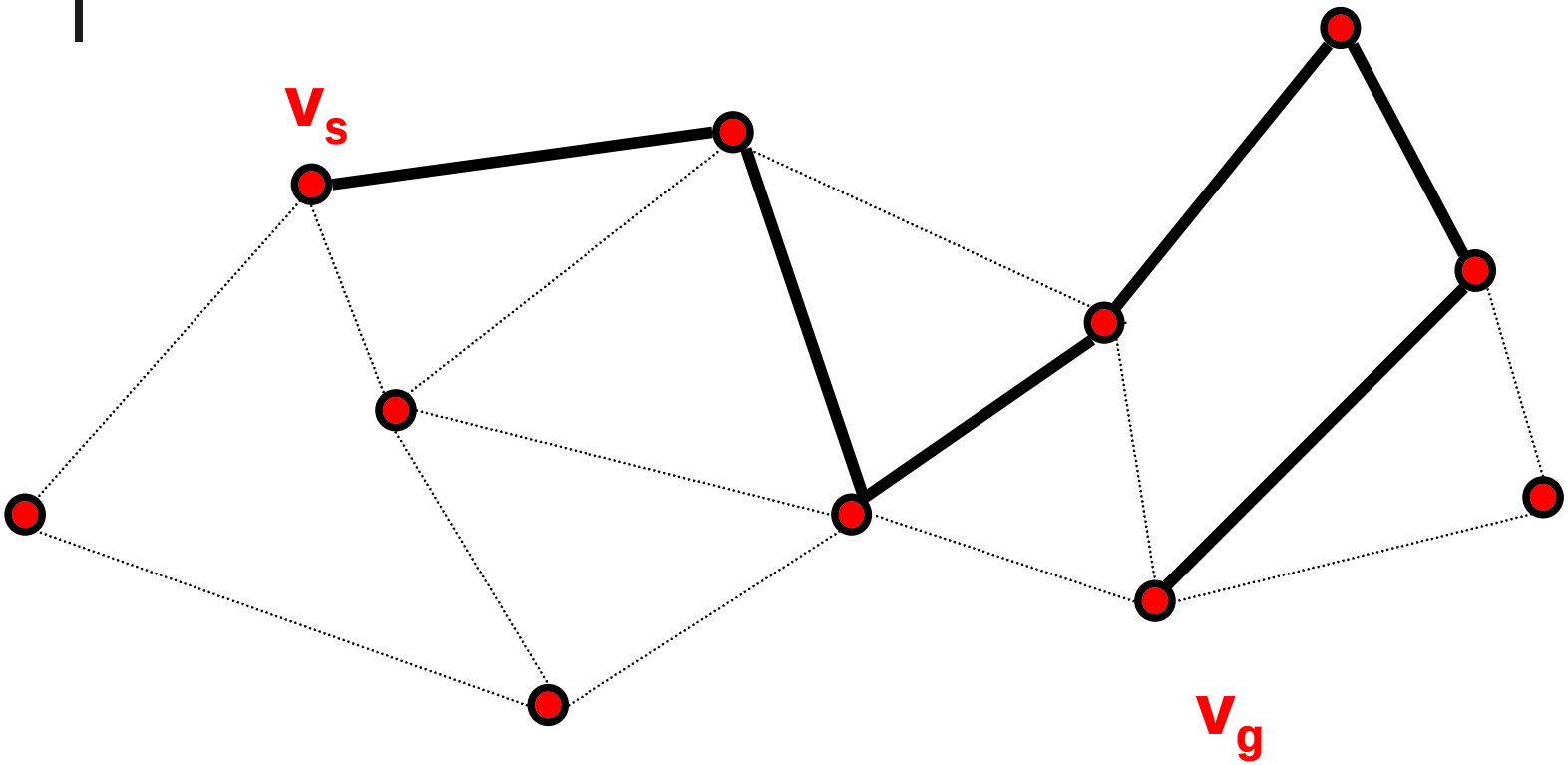
Relationship to MC simulation



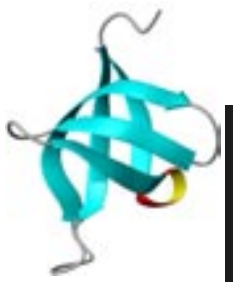
- Each path on graph = a path of MC simulation;
- Roadmap represents many MC simulation paths simultaneously;



Relationship to MC simulation

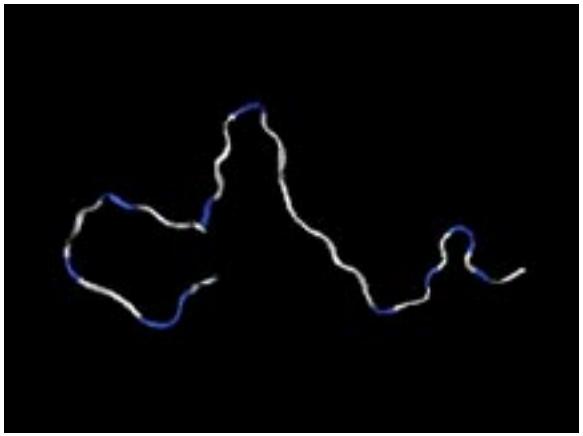


- Each path on graph = a path of MC simulation;
- Roadmap represents many MC simulation paths simultaneously;
- Proved: SRS and MC converge to the same distribution.

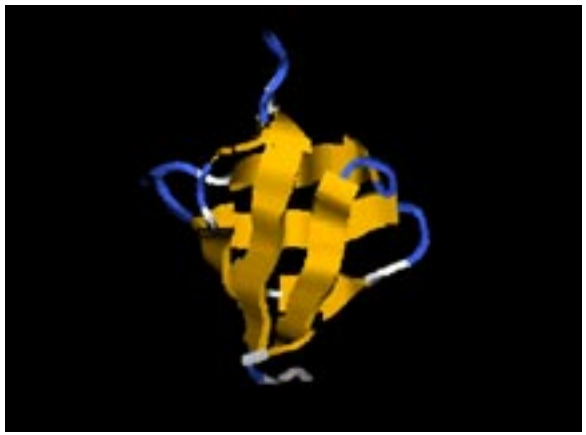


Application: probability of folding p_{fold}

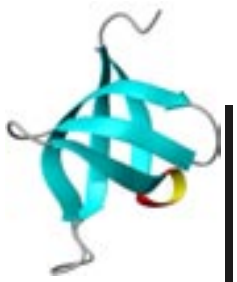
[Du et al. ‘98]



Unfolded set

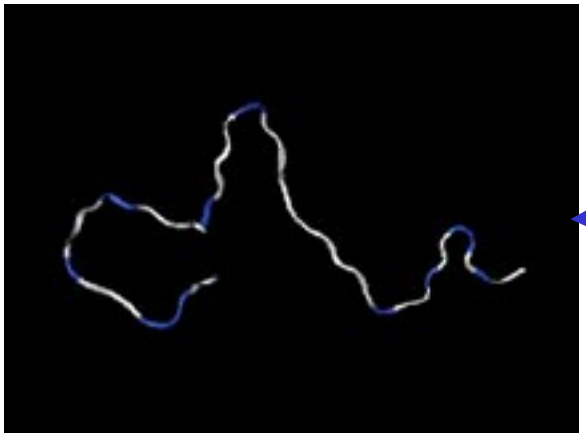


Folded set



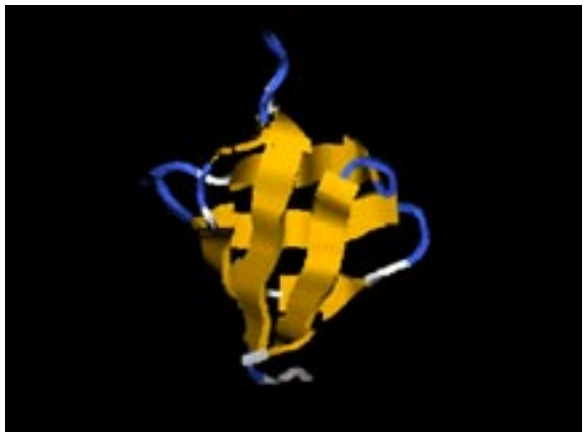
Application: probability of folding p_{fold}

[Du et al. ‘98]



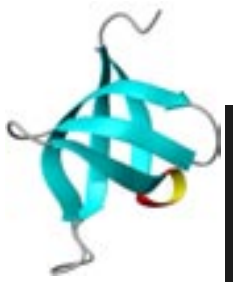
1- p_{fold}

p_{fold}



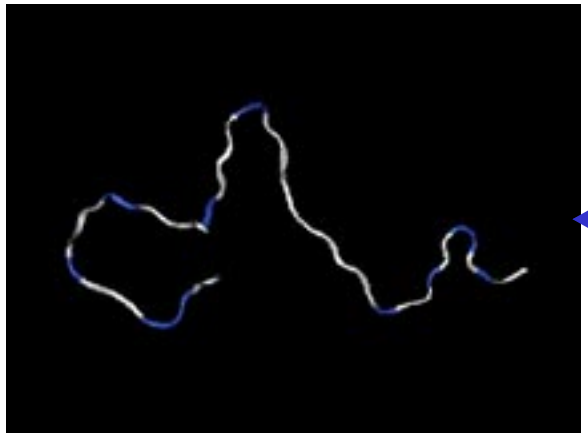
Unfolded set

Folded set



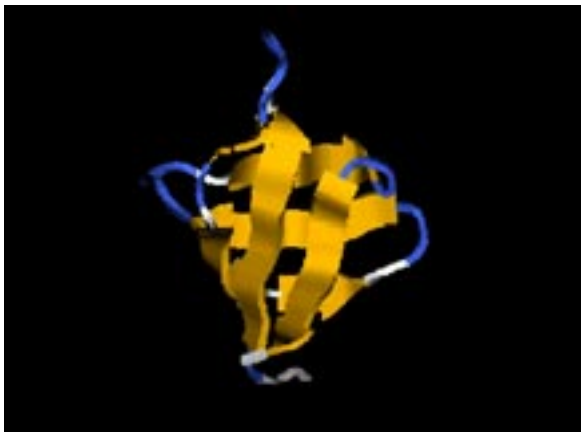
Application: probability of folding p_{fold}

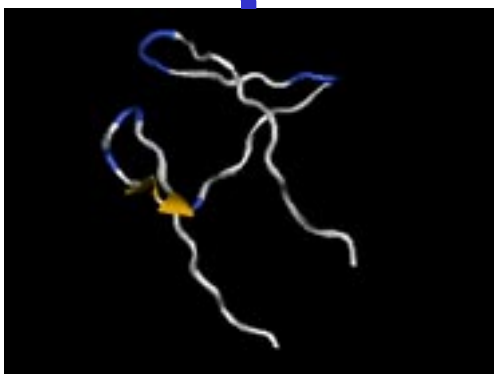
[Du et al. ‘98]



1- p_{fold}

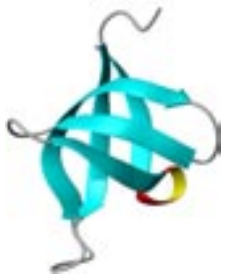
p_{fold}



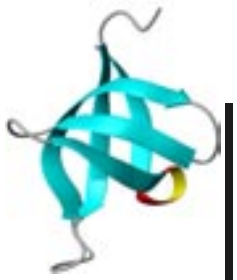


Unfolded set

Folded set

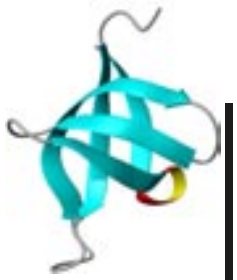


Computing p_{fold}: MC approach



Computing p_{fold}: MC approach

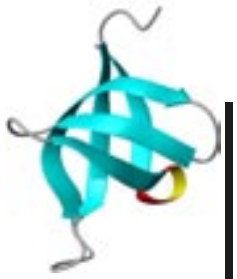
- For each conformation:
  - Perform many MC/MD simulations;
  - Count number of folded simulations.



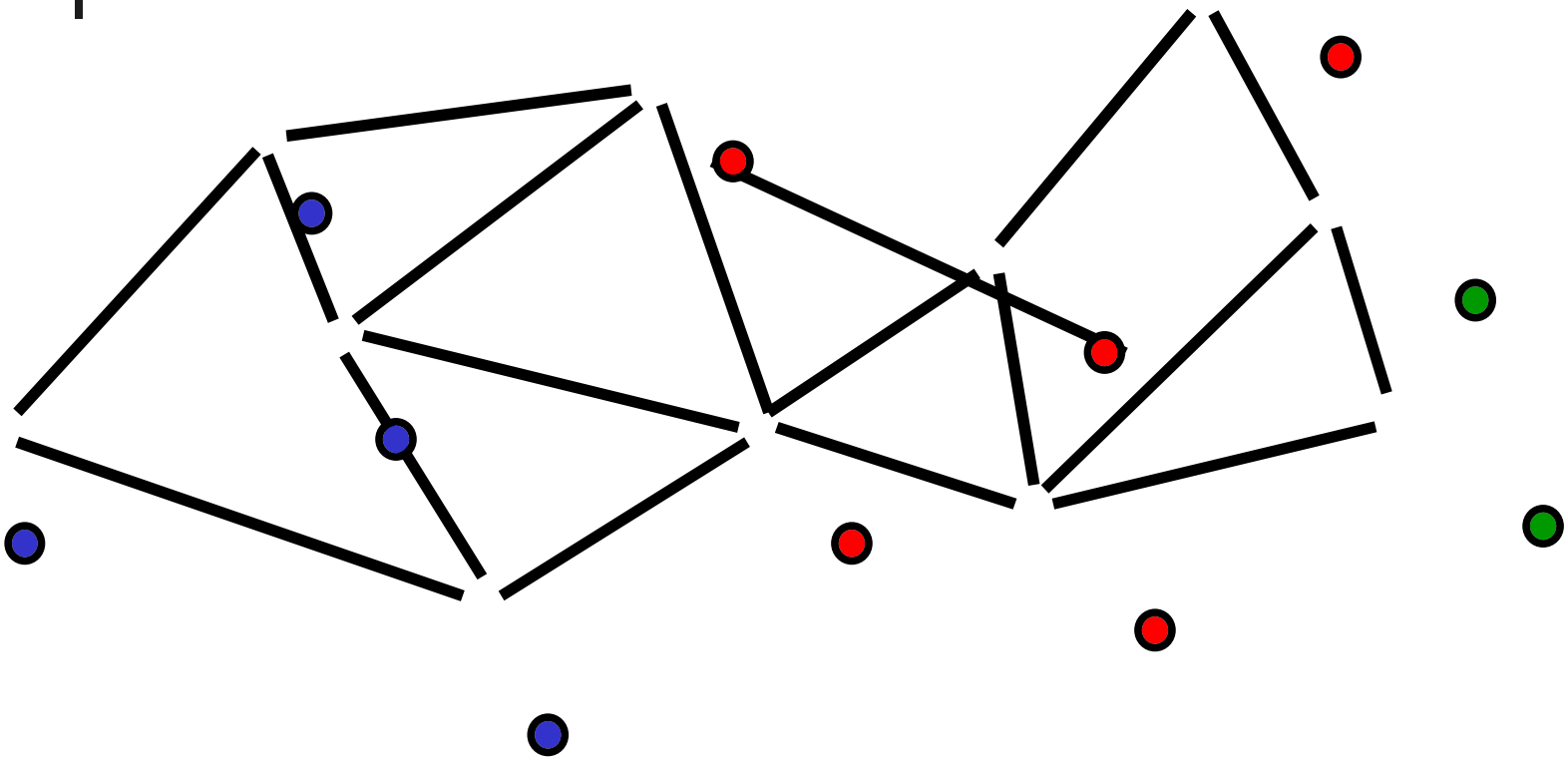
Computing p_{fold}: MC approach

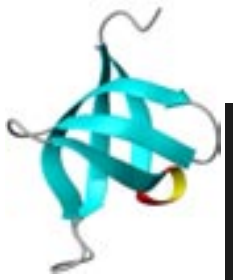
- For each conformation:
  - Perform many MC/MD simulations;
  - Count number of folded simulations.

- Too slow for any practical application!

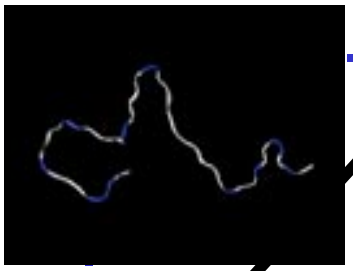


Roadmap for p_{fold} computation





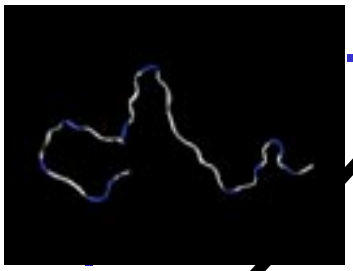
Roadmap for p_{fold} computation



Folded set

Unfolded set

Roadmap for p_{fold} computation



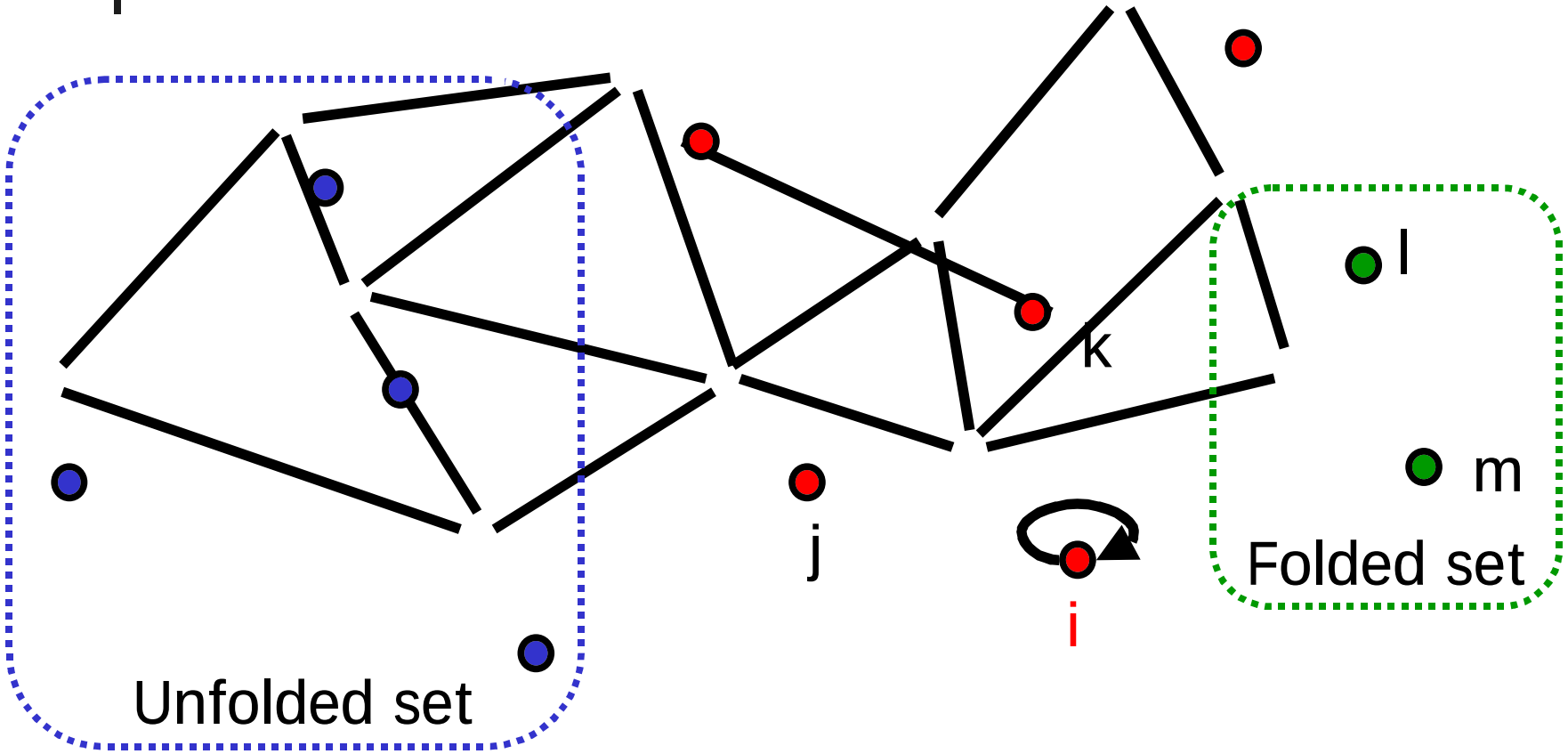
f_i

Folded set

Unfolded set

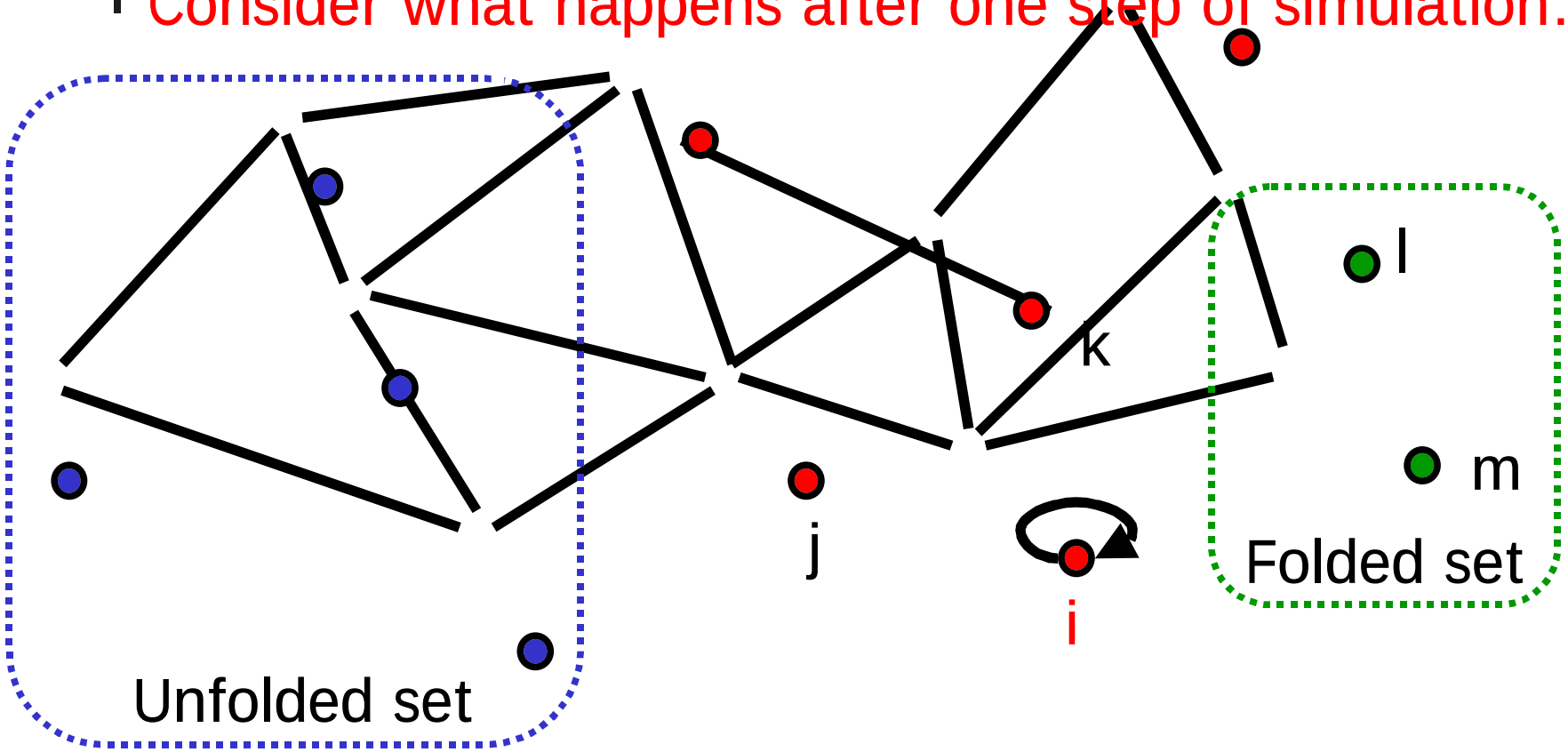
Want to compute: f_i = probability of folding starting from node i

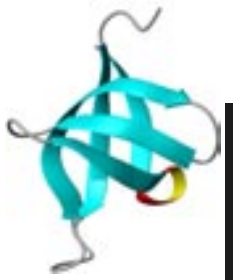
First step analysis



First step analysis

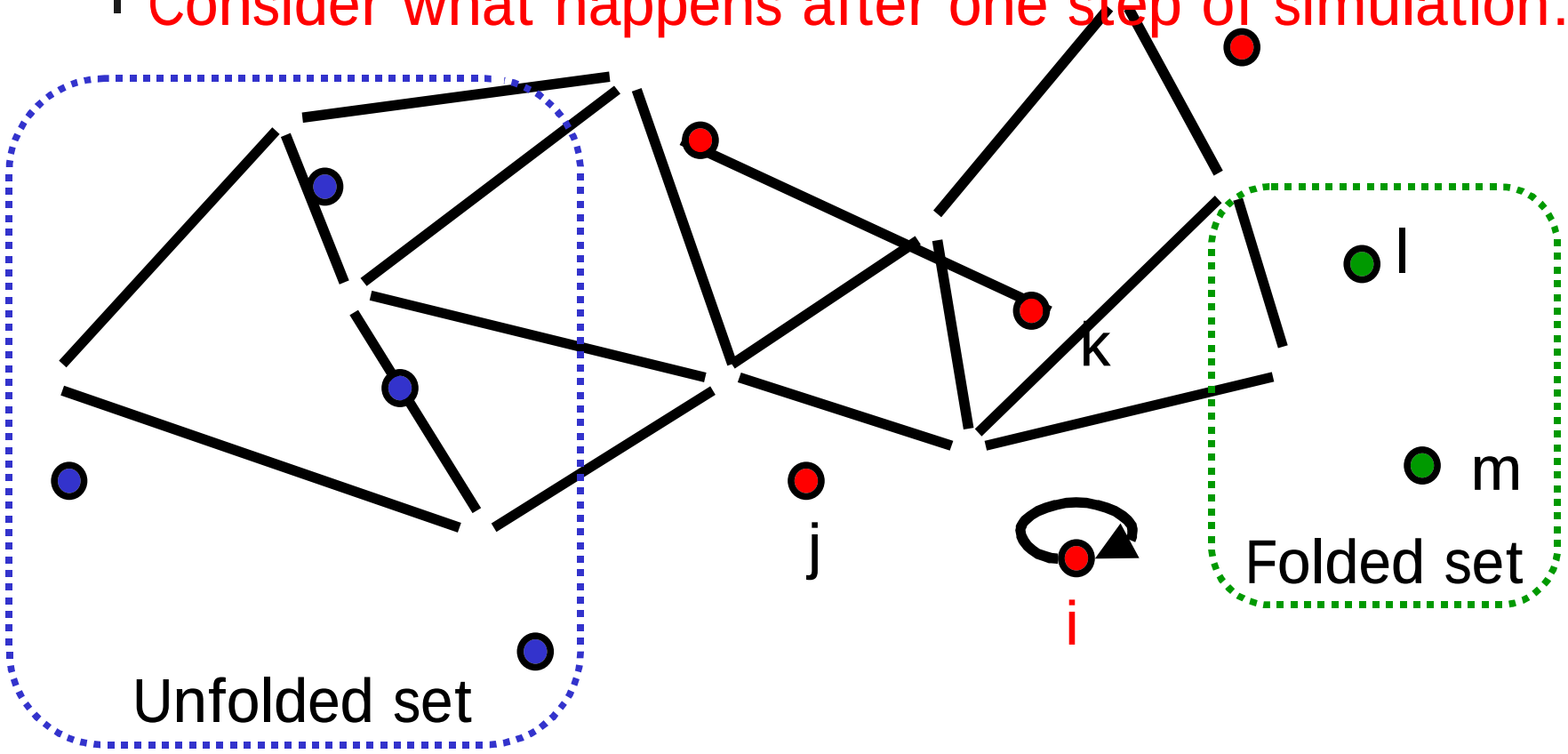
Consider what happens after one step of simulation:





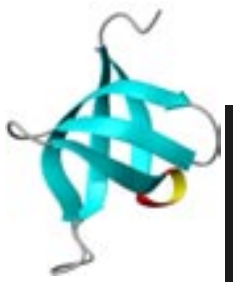
First step analysis

Consider what happens after one step of simulation:



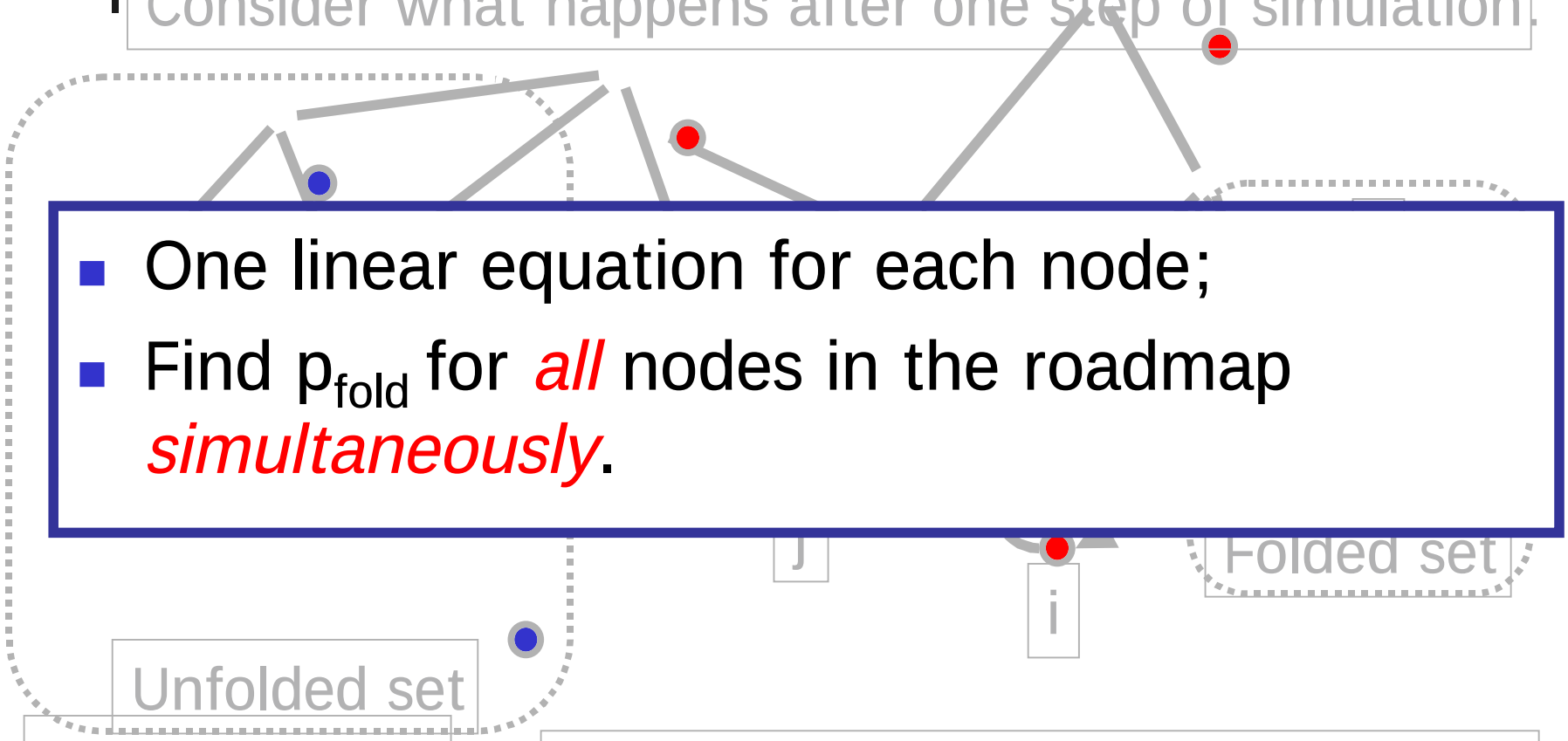
After one step: $f_i = P_{ii}\, f_i + P_{ij}\, f_j + P_{ik}\, f_k + P_{il}\, f_l + P_{im}\, f_m$

But $f_l = f_m = 1$: $f_i = P_{ii}\, f_i + P_{ij}\, f_j + P_{ik}\, f_k + P_{il}\, 1 + P_{im}\, 1$



First step analysis

Consider what happens after one step of simulation:



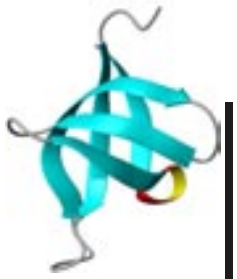
- One linear equation for each node;
- Find p_{fold} for *all* nodes in the roadmap *simultaneously*.

After one step:

$$f_i = P_{ii}\, f_i + P_{ij}\, f_j + P_{ik}\, f_k + P_{il}\, f_l + P_{im}\, f_m$$

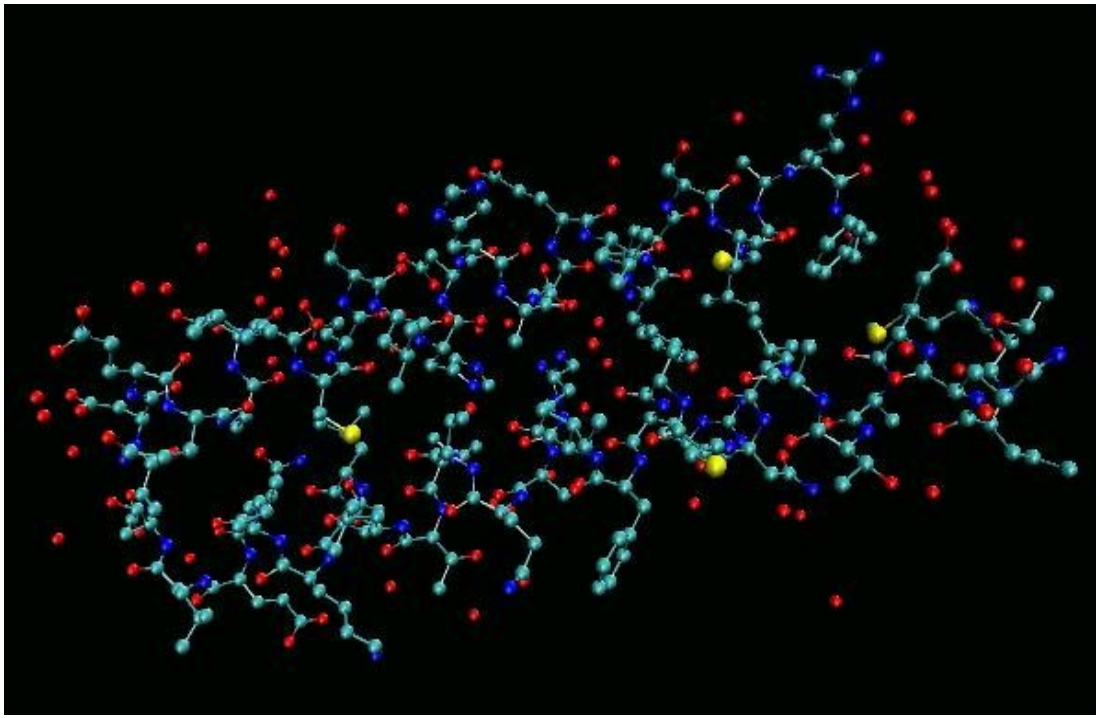
But $f_l = f_m = 1$:

$$f_i = P_{ii}\, f_i + P_{ij}\, f_j + P_{ik}\, f_k + P_{il}\, 1 + P_{im}\, 1$$

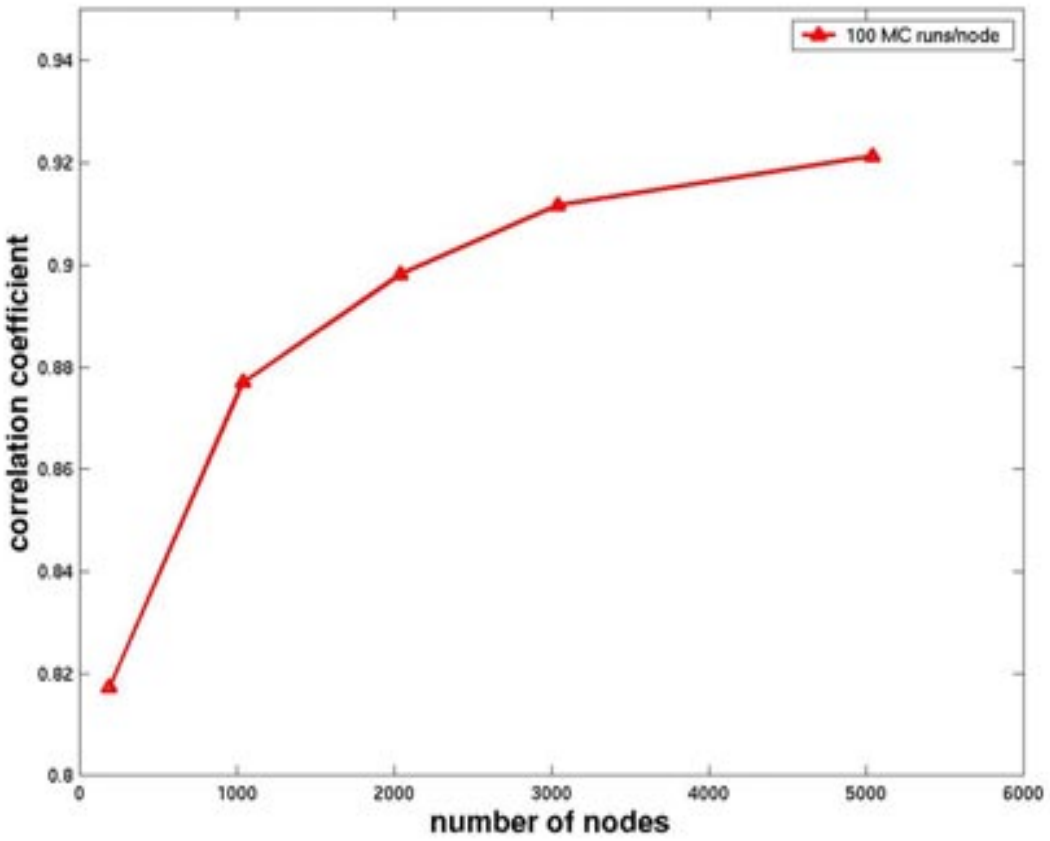


p_{fold} on real protein: 1ROP

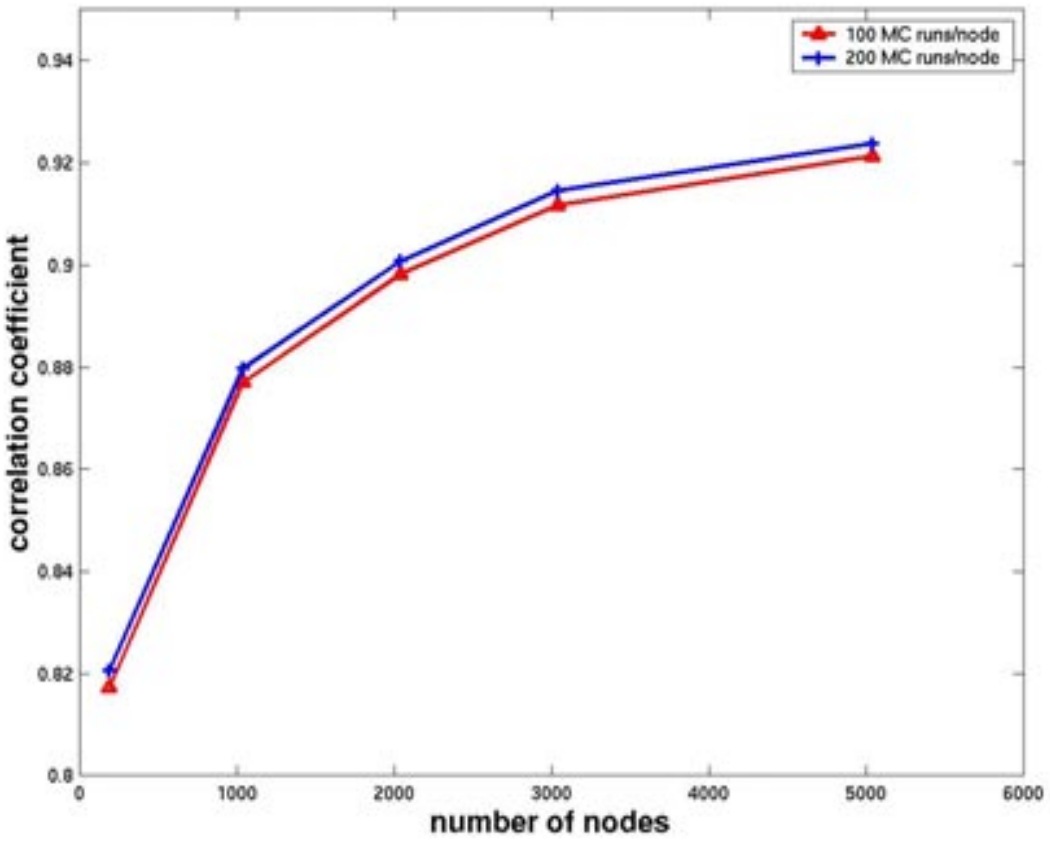
- Comparison:
  - SRS;
  - MC simulation for 36 starting conformations.

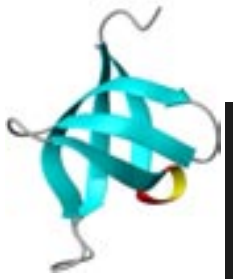


Correlation: SRS versus MC

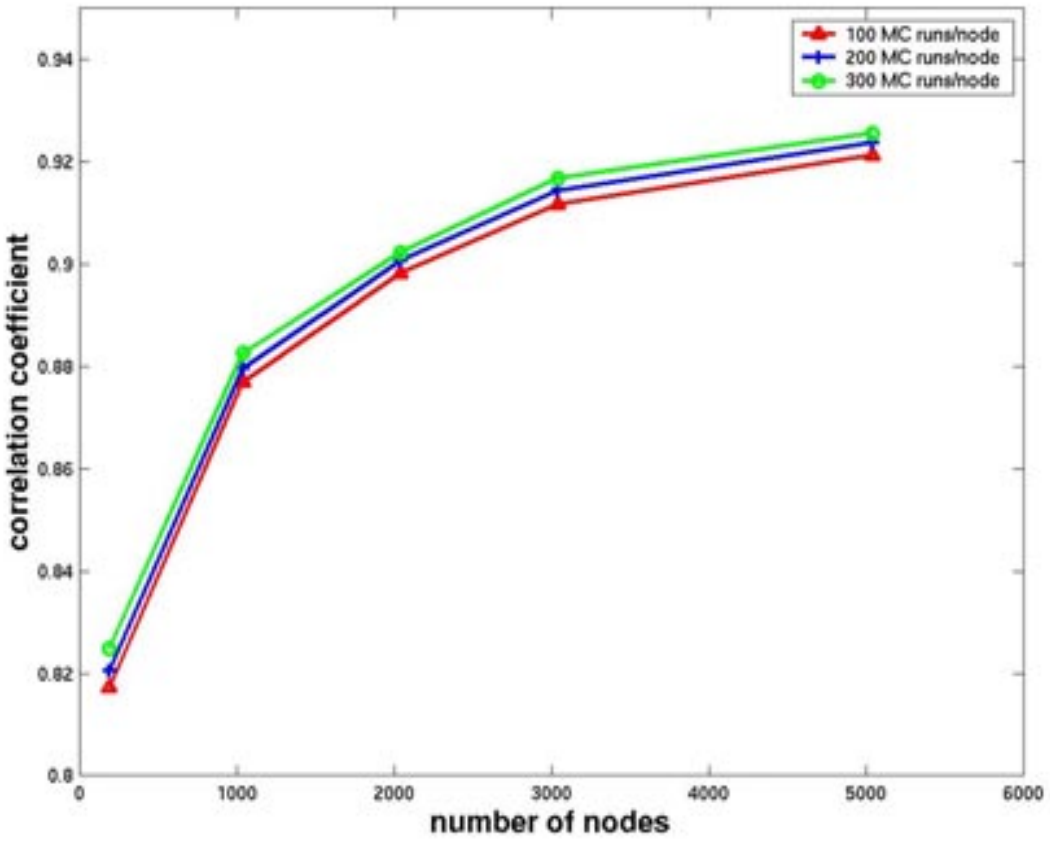


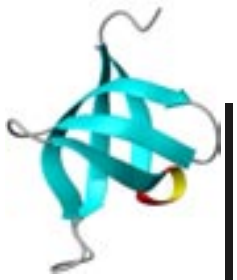
Correlation: SRS versus MC



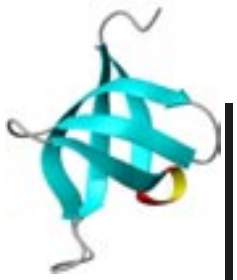


Correlation: SRS versus MC





Computation time on 1ROP



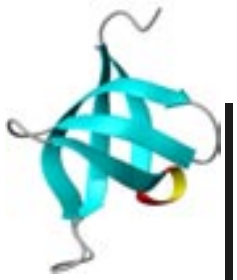
Computation time on 1ROP

Monte Carlo:

36 conformations

100 days of computer time

Over 10^9 energy computations



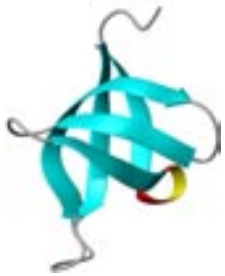
Computation time on 1ROP

Monte Carlo:

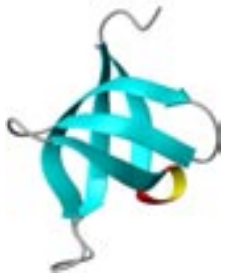
36 conformations	100 days of computer time	Over 10^9 energy computations

SRS:

5000 conformations	1 hour of computer time	5000 energy computations

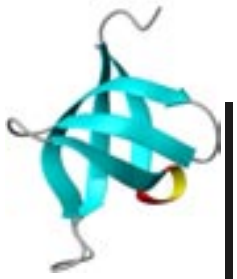


Conclusion



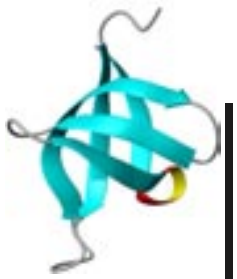
Conclusion

- Roadmap for analysis of molecular motion;



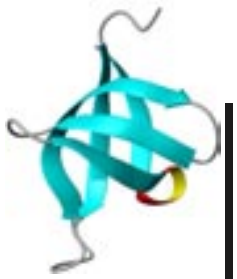
Conclusion

- Roadmap for analysis of molecular motion;
- Efficiently considers many MC paths simultaneously;



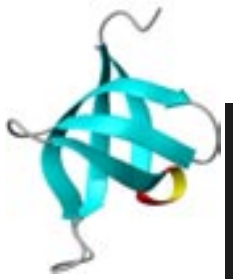
Conclusion

- Roadmap for analysis of molecular motion;
- Efficiently considers many MC paths simultaneously;
- Application to p_{fold}:
  - More accurate results;
  - Fewer energy computations;
  - *Many* orders of magnitude speed-up.



Conclusion

- Roadmap for analysis of molecular motion;
- Efficiently considers many MC paths simultaneously;
- Application to p_{fold}:
  - More accurate results;
  - Fewer energy computations;
  - *Many* orders of magnitude speed-up.
- Other applications:
  - Ligand-protein docking;
  - Order of formation of secondary structure.



Conclusion

"To conclude, we stress that we do not suggest using p_{fold} as a transition coordinate for practical purposes as it is very computationally intensive."

Du, Pande, Grosberg, Tanaka, and Shakhnovich "On the Transition Coordinate for Protein Folding" *Journal of Chemical Physics* (1998).

We believe that Stochastic Roadmap Simulation could make computations such as p_{fold} practical.