Hybrid Probabilistic Roadmap and Monte Carlo Methods for Biomolecule Conformational Changes

Li Han [1]

Keywords: Conformation space, conformational changes, Monte Carlo, probabilistic roadmaps.

1 Introduction

Biomolecule conformational changes play important roles in biological functions and are an integral part of various important structural problems addressed in computational biology such as ligand binding and (mis)folding of proteins and nucleic acids. The current structural biology experimental techniques are not yet as powerful for providing information on dynamic molecular motion as on static molecular structures; but there has been a large body of work using computational approaches to study conformational changes. Even with the growth of computer power and the impressive results generated from computational study, computer modeling and simulation of biomolecules remain challenging in general, since molecular systems follow complex physical laws and involve high dimensional conformation spaces. In this work, we have developed a hybrid Probabilistic Roadmap and Monte Carlo planner for biomolecule conformation study. The planner has been designed to integrate the strengths of both probabilistic roadmap and Monte Carlo simulation methods, and its effectiveness has been demonstrated in our promising preliminary results.

2 Prior Work

Monte Carlo (MC) methods have been successfully applied to statistical physics and chemistry where the motion of complex systems of thousands of atoms is studied. Given an initial conformation of a molecular system, a new conformation is generated through a perturbation of the initial conformation. If the energy of the new conformation is lower than that of the initial one, the new conformation is accepted. Otherwise, the new conformation is accepted with some probability: the higher the new energy, the lower the acceptance rate. Such a method can be used to generate a trajectory of molecular conformations obeying the Boltzmann distribution. While the algorithmic simplicity and theoretical grounding have made MC appealing to a wide variety of problems, MC methods are quite computationally expensive. For large molecules, random Monte Carlo moves generally result in high rejection rates and long running time. Furthermore, each simulation run is generally treated independently from other runs, without using any knowledge generated from prior runs. Regarding the high rejection rate problem, it has been identified that perturbing one or few atoms at each step generally leads to a low percentage of rejections. This type of moves also facilitates efficient computation of molecular structures and energy[4] since a small number of perturbed atoms will partition a molecule chain into a small number of subchains, where the perturbation only causes non-rigid relative motion between the subchains and only the energy terms involving atoms of different subchains need to be updated. As for reusing prior computation results to improve the simulation efficiency, we believe that probabilistic roadmap (PRM) methods provide a natural framework.

[1]Department of Mathematics and Computer Science, Clark University, Worcester, MA 01610, USA. E-mail: lhan@clarku.edu

PRM [2] was originally developed for robot motion planning problems and has been successfully adapted to studying computational biology problems[1, 3]. Given a molecular system, the general methodology of PRM planners is to construct a graph(roadmap) to capture the properties of the conformation space. Roadmap vertices are randomly sampled conformations with their inclusion in the roadmap determined by some acceptance criteria, and roadmap edges correspond to transitions between 'nearby' vertices found with simple local planning methods such as the linear interpolation, again with inclusion of edges subject to some acceptance conditions such as the Metropolis criteria. After generating a roadmap, numerous trajectories between pairs of conformations can be extracted from the graph. This is one distinct feature of PRMs as compared to Monte Carlo simulation, where each successful simulation run generates one trajectory and many simulation runs simply fail. The success of PRM can be partly attributed to its efficient reuse of nodes and edges in the roadmap.

3 Our Methods

PRM planners generally need to sample and connect a large number of conformations in order to build a roadmap that reflects the properties of the conformation space. Conventionally, roadmap nodes are generated independently from each other. Such an approach need to compute the energy of each conformation from scratch and move most, if not all, atoms to connect each pair of conformations, which contribute to long computation times and low acceptance rates.

We have recently developed a hybrid Probabilistic Roadmap and Monte Carlo (PRM-MC) planner to combine the favorable properties of PRM – capturing the connectivity of conformations and facilitating the reuse of already identified conformations and conformational changes – with those of MC – efficiently updating molecular conformations and energy after perturbations of a small number of atoms at each step. First the planner generates a small number of seed conformations. These seed conformations can be experimentally determined structures stored in PDB or random conformations generated from some conformation space sampling schemes. Then the planner uses Monte Carlo simulation to generate and connect conformations originated from the perturbation of the seed conformations. Finally the planner uses PRM type connection strategies to try to find more energetically feasible conformational changes that have not been identified in the Monte Carlo step. Our preliminary simulation results have indicated that the PRM-MC planner can efficiently generate roadmaps useful for the study of molecule conformation space and conformational changes.

References

[1] N. M. Amato, Ken A. Dill, and G. Song. Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 2–11, 2002.

[2] L. Kavraki, P. Svestka, J. C. Latombe, and M. Overmars. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans. Robot. Automat.*, 12(4):566–580, August 1996.

[3] S.M. LaValle, P.W. Finn, L.E. Kavraki, , and J.C. Latombe. A randomized kinematics-based approach to pharmacophore-constrained conformational search and database screening. *Journal of Computational Chemistry*, 21(9):731–747, 2000.

[4] I. Lotan, F. Schwarzer, D. Halperin, and J.C. Latombe. Algorithm and data structure for efficient energy maintenance during monte carlo simulation of proteins. *Journal of Computational Biology*, 2003.