NAMS
91-27

ALMOST MARKOV PROCESSES IN MONTE CARLO SIMULATION OF BIOLOGICAL MOLECULES

by

Robert H. Swendsen
Department of Physics
Carnegie Mellon University
Pittsburgh, PA 15213

Djamal Bouzida
Department of Physics
Carnegie Mellon University
Pittsburgh, PA 15213

Shankar Kumar
Department of Biological Sciences
University of Pittsburgh
Pittsburgh, PA 15260

WORKSHOP ON CALCULUS OF VARIATIONS AND
NONLINEAR MATERIAL BEHAVIOR
November 1-4, 1990

Research Report No. 91-121-NAMS-27

July 1991

Almost Markov Processes in Monte Carlo Simulation of Biological Molecules

Robert H. Swendsen
Department of Physics, Carnegie Mellon University,
Pittsburgh, Pennsylvania 15213

Djamal Bouzida
Department of Physics, Carnegie Mellon University,
Pittsburgh, Pennsylvania 15213

Shankar Kumar
Department of Biological Sciences, University of Pittsburgh,
Pittsburgh, Pennsylvania 15213

Abstract

Computer simulation is an important tool for investigating properties of biological molecules. This paper presents newly-developed optimization techniques that take into account the inhomogeneity and anisotropy essential to the structure and function of biological molecules to produce highly efficient simulations.

During the past few decades, computer simulations have become an increasingly important tool for the study of thermodynamic systems [1]. Extremely rapid advances in both hardware and simulation techniques have improved the accuracy of such calculations and broadened the range of problems considered.

An important area of computer simulation at the borders of medicine, biology, chemistry, and physics is the study of macromolecules [2, 3]. Because of the large number of atoms and the complexity of the interactions, even today's largest supercomputers are not yet sufficiently powerful to deal with all the problems of interest. Consequently, new approaches to these problems are needed to increase the efficiency of these calculations.

In this paper, we present a new approach we have developed. We have achieved substantial improvement in both the efficiency and reliability of the simulations, but our approach has also opened up some interesting mathematical questions about the nature of an approximation we have introduced.

The computer simulation methods currently in use (molecular dynamics and Monte Carlo) were originally developed for treating fluids. They naturally made use of the simplifying feature that fluids are isotropic and homogeneous. Unfortunately, macromolecules have neither property, which leads to a large decrease in efficiency.

A Monte Carlo simulation stores an explicit configuration of the system, in this case, a large molecule. The algorithm is a stochastic process that consists of moving atoms, or groups of atoms, to simulate the thermal motion of the system. It is usual to consider only Markov processes, in which the transition probabilities depend only on the current configuration of the system, and not on its history. This enables the use of a theorem that if a condition on the transition probabilities is satisfied (detailed balance) and the system is ergodic (any configuration can be reached from any other configuration in a finite number of steps with non-zero probability), then the simulation will correctly reproduce the thermal distribution (in the limit of infinite computer time).

In setting up a simulation, there are generally one or two global parameters that must be specified to optimize the efficiency. These are normally determined during a preliminary simulation, and then fixed during the simulation proper.

In order to take into account the inhomogeneity and the local anisotropy, we have used data accumulated during the simulation to determine parameters for optimizing moves of individual atoms. This necessarily relaxes the Markov condition, because the new transition probabilities depend on the recent history of the simulation [4]. Empirical studies demonstrate that the almost Markovian nature of the simulation does not introduce any systematic errors, but a mathematical proof is still missing.

To develop the new method, we approximate the local potential by an anisotropic simple harmonic oscillator of the form

$$H = \frac{1}{2} \sum_{ij} k_{ij} x_i x_j \tag{1}$$

where k_{ij} represents the spring constants. A transformation matrix, D_{ij}, is

used to generate moves in an ellipsoid given by

$$d_i = \sum_j D_{ij} v_j \tag{2}$$

where v_j is a random vector chosen from a unit sphere. The optimum jump size scales as the contours of constant energy and is determined by a dimensionless parameter, F, given by

$$F^2 = \frac{1}{2}\beta \sum_{ij} k_{ij} d_i d_j \tag{3}$$

where $\beta = \frac{1}{k_B T}$. F is a scale factor chosen to optimize the efficiency of the simulation. The effective values of k_{ij} are determined from the simulation using

$$[\Delta E d_l d_m] = \frac{1}{2} \sum_{ij} k_{ij} [d_i d_j d_l d_m] \tag{4}$$

where ΔE denotes the change of energy for an attempted move d_i and the square brackets indicate an average over all attempted moves, whether or not they were accepted. The matrix D_{in} is then obtained from

$$D_{in} = F \sqrt{\frac{2}{\beta \lambda_n}} V_{in} \tag{5}$$

where λ_n is an eigenvalue of K_{ij} and V_{in} is the corresponding normalized eigenvector. D_{ij} is updated periodically to adapt to the changing local environment of each atom. This method, which we denote as Dynamically Optimized Monte Carlo ($DOMC$) turns out to be remarkably robust when applied to a wide variety of Hamiltonians.

As a simple test for systematic error, we simulated a one-dimensional simple harmonic oscillator, and used DOMC with very short cycles. Averaging over only 1 MC-step/cycle, we find a large systematic error of 42%. However, with even 2 MC-steps/cycle, the error is reduced to 3%, and for 3 MC-steps/cycle, it is just over 1%. No systematic error was measurable for 4 MC-steps/cycle. Since we have generally used more than 10 MC-steps/cycle in our work, the systematic error appears to be negligible.

We have already applied this method to a large number of systems, including highly anisotropic oscillators and biomolecules ranging from adenosine (one of the bases of DNA) to complex proteins such as the Bovine

Pancreatic Trypsin Inhibitor ($BPTI$). Our programs have included additional refinements beyond the scope of this paper to take long-wave-length fluctuations into account more effectively. Results have been very good, with large improvements in efficiency.

Work is continuing to extend this approach and prospects are extremely promising. Efforts to analyze the almost Markovian nature of these algorithms and provide a clear understanding and estimation of the systemmatic errors is an important aspect of this research.

Acknowledgment: We would like to acknowledge support by the National Science Foundation Grant No. DMR-9009475[RHS and DB] and NIH Grant No.GM25671 [SK]. We would also like to thank the Pittsburgh Supercomputer Center for supercomputer time under Grant DMB890026P [SK].

References

[1] *Applications of the Monte Carlo Method in Statistical Physics*, Vol. 36, 2nd ed., edited by K. Binder (Springer, Berlin, 1987).

[2] C. L. Brooks III, M. Karplus and B. M. Pettitt, "Proteins: A theoretical perspective of dynamics, structure and thermodynamics", Advances in Chemical Physics, Vol. LXXI, edited by I. Prigogine and S. Rice, John Wiley and Sons, New York, 1988.

[3] J. A. McCammon and S. C. Harvey, "Dynamics of proteins and nucleic acids", Cambridge University Press, New York, 1987.

[4] S. Shumway and J. P. Sethna (Bull. Am. Phys. Soc. **35** 500 (1990)) have presented a novel technique for optimizing MC moves. It is different from ours, but it also uses an almost Markovian process to optimize the simulations.

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTA

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1991	

4. TITLE AND SUBTITLE

Almost Markov Processes in Monte Carlo Simulation of Biological Molecules

5. FUNDING NUMBERS

6. AUTHOR(S)

Robert H. Swendsen, Djamal Bouzida & Shankar Kumar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Carnegie Mellon University
Department of Mathematics
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

NAMS-27

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Computer simulation is an important tool for investigating properties of biological molecules. This paper presents newly-developed optimization techniques that take into account the inhomogeneity and anisotropy essential to the structure and function of biological molecules to produce hgihly efficient simulations.

14. SUBJECT TERMS

15. NUMBER OF PAGES
4

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL