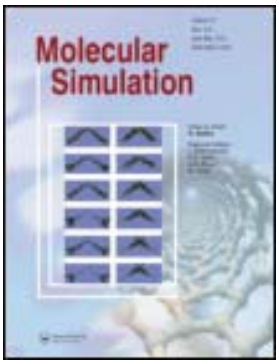

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmos20

Molecular Dynamics Simulation on a Parallel Computer

H. Heller [a] , H. Grubmüller [a] & K. Schulten [a]
[a] Beckman Institute and Department of Physics, University of Illinois, 405 N. Mathews Ave., Urbana, IL, 61801, U.S.A.
Version of record first published: 23 Sep 2006.

To cite this article: H. Heller , H. Grubmüller & K. Schulten (1990): Molecular Dynamics Simulation on a Parallel Computer, Molecular Simulation, 5:3-4, 133-165

To link to this article: http://dx.doi.org/10.1080/08927029008022127

Molecular Simulation, 1990, Vol. 5, pp. 133–165
Reprints available directly from the publisher
Photocopying permitted by license only

MOLECULAR DYNAMICS SIMULATION ON A PARALLEL COMPUTER

H. HELLER, H. GRUBMÜLLER and K. SCHULTEN*

Beckman Institute and Department of Physics, University of Illinois, 405 N. Mathews Ave., Urbana, IL 61801 U.S.A.

(*Received March 1989, accepted October 1989*)

For the purpose of molecular dynamics simulations of large biopolymers we have built a parallel computer with a systolic loop architecture, based on Transputers as computational units, and have programmed it in occam II. The computational nodes of the computer are linked together in a systolic ring. The program based on this topology for large biopolymers increases its computational throughput nearly linearly with the number of computational nodes. The program developed is closely related to the simulation programs CHARMM and XPLOR, the input files required (force field, protein structure file, coordinates) and output files generated (sets of atomic coordinates representing dynamic trajectories and energies) are compatible with the corresponding files of these programs. Benchmark results of simulations of biopolymers comprising 66, 568, 3 634, 5 797 and 12 637 atoms are compared with XPLOR simulations on conventional computers (Cray, Convex, Vax). These results demonstrate that the software and hardware developed provide extremely cost effective biopolymer simulations. We present also a simulation (equilibrium of X-ray structure) of the complete photosynthetic reaction center of *Rhodopseudomonas viridis* (12 637 atoms). The simulation accounts for the Coulomb forces exactly, i.e. no cut-off had been assumed.

KEY WORDS: Molecular dynamics simulation, parallel computers, parallel programming, Transputer, photosynthetic reaction center.

1. INTRODUCTION

A major concern of molecular biology is to understand the structure-function relationship of biological polymers, mainly proteins and nucleic acids. For a long time it had been tacitly assumed that the function of a biopolymer can be revealed from its static structure, i.e. from a precise knowledge of the equilibrium positions of its atoms together with a knowledge of typical atomic charges and chemical properties like hydrogen bonding. However, during the past decade it has been realized that further properties, which are not evident from the characteristics of the separate constituents of biopolymers, are required to understand function. Such properties are, for example, thermal mobilities of atoms, activated motions of constituent groups, local electric fields and dielectric relaxation.

The properties mentioned are often very difficult to measure experimentally even for small subsections of biopolymers, let alone for the whole polymer. It appears that the required information can be obtained only by computer simulations of biopolymers. Currently, many groups are developing a software basis in order to allow an increasingly faithful representation of biopolymers by computer programs. These programs are likely to contribute to biology and biotechnology beyond the scope of

*To whom correspondence should be sent.

elucidating the structure-function relationship: they might guide the synthesis of new materials, e.g. in conjunction with genetic engineering methods, as well as predict the properties of materials, e.g. drug specificities or mechanical properties. A further long-range goal of computer simulations is to predict secondary, tertiary and quaternary structures of proteins from their primary structure (amino acid sequence).

The prospects of computer simulations in molecular biology rest on the availability of suitable computer resources. In fact, simulations until now are limited to rather small biopolymers (of a few thousand atoms) and short simulation periods (of a few nanoseconds). Furthermore, the cardinal issue of a faithful representation of biopolymers by computer simulation is closely linked to the availability of computational resources: realistic descriptions of forces acting between the constituents of biopolymers, e.g. a proper description of Coulomb forces without 'cut-off', require enormous computer time; simulations must also represent enough of the surrounding medium, e.g. lipids of biological membranes and water, in order to achieve realistic descriptions.

Our study of the photosynthetic reaction center of the bacterium *Rhodopseudomonas viridis* [1] is a case in point. We have found [2,3,4] that consideration of all atoms of the molecule and an adequate representation of electrical interactions is a prerequisite for reliable simulations. The photosynthetic reaction center is a protein complex embedded in a cellular membrane and contains 12 large prosthetic groups [5]; the whole system encompasses about 12 600 atoms. The protein complex converts the energy of the sunlight into an electric membrane potential. The two primary processes which are responsible for the high efficiency of photosynthetic energy conversion last 3 ps and 140 ps, respectively. If one wishes to include in a simulation of the primary processes some of the surrounding membrane, water and ions, one would have to simulate about 30 000 atoms over a period of a few hundred picoseconds. The necessary computations are not yet feasible except at some extreme cost, as is explained below.

Computer simulations are faced with a serious computational barrier which can be illustrated by the following estimate of the requirements on computer time: In order to determine the forces between all atoms of a protein with 12 600 atoms, i.e. of the photosynthetic reaction center, without 'cut-off', about 98 s on a Cray-XMP are needed. Since the forces have to be re-evaluated at each integration step, the size of which has to be chosen 1 fs or shorter, a simulation describing a period of 1 ns requires at least one million steps, i.e. more than 1000 days of Cray time. A 'cut-off' of pair interactions to 10 Å reduces this time to about 19 days, but one may question the soundness of such approach. The numbers illustrate a well-known point, namely, that computational requirements for molecular dynamics simulations are prohibitive and, for many problems, exceed all available means.

In this article we want to show that this situation can be improved by employing parallel computers to simulate biopolymers. For the purpose of such demonstration we have built a parallel computer with a systolic ring architecture, the design of which will be outlined. We have developed also a program for protein simulations on this computer. Computer and program achieve the same rate of computation as much larger conventional vector machines, but for a small fraction of their cost. The parallel strategy, therefore, should make the method of computer simulations accessible to many researchers, allow simulations of larger numbers of atoms as well as of more realistic (and computer time consuming) force models.

Our suggestion is to delegate only the computationally most intensive phase of

molecular dynamics simulations, namely the force evaluation and the integration step[1], to a parallel computer and to employ existing simulation programs and graphics packages for an analysis of the simulated trajectories. To implement this suggestion a program on a parallel computer needs to interface with some existing simulation software using standard input and output files. The program described below provides such an interface specifically for the simulation programs CHARMM [6] and XPLOR [7, 8].

The parallel computer built and programmed by us is based on Transputers as computational units. The Transputer is a 32 bit processor with a 64 bit floating point coprocessor integrated on a single chip. We have chosen this chip for our parallel computer for reasons detailed further below. One advantage of the choice of the Transputer is that parallel computers, comparable to the one developed by us, can be obtained from commercial manufacturers. Therefore, molecular biologists not willing to build computer hardware, which we assume is the majority, can still use our simulation program. The reason why we decided to build our own computer rather than use a commercial machine is the following: We wanted to choose an optimal computer design in order to achieve for a minimum cost a rate of computation comparable to that of the best currently available supercomputers.

Our program has been written in occam II [9, 10, 11, 12], the language around which the Transputer had been designed. The language facilitates distribution of computational processes among Transputers as well as communication among these processes. In the initial phase of the research described here, occam II had been the only programming language for the Transputer. However, since that time more conventional languages, e.g. FORTRAN and C, have been ported to the Transputer. Molecular biologists who prefer these languages over a new language might want to incorporate the programming strategies presented below in these conventional languages. Such approach would actually allow to include elements of programs, written for sequential machines, into the program for a parallel Transputer-based machine.

The software development environment chosen by us has been the Transputer Development System (TDS), also described below, which runs on IBM personal computers. Recently, familiar operating systems, e.g. UNIX-like systems, have also been adapted to Transputer workstations. We point this possibility out, since the aim of this paper is not to propagate a particular parallel computer and a particular concurrent algorithm but rather to provide a convincing example which demonstrates the feasibility and the cost effectiveness of molecular dynamics simulations on parallel computers.

It must be pointed out that we succeeded in making molecular dynamics simulations more affordable because we had ready access to supercomputers such that programming strategies for large scale molecular dynamics simulations became familiar to us. Our experience that the use of supercomputers does not always lead to bigger appetite for expensive computational equipment, but rather can have the opposite effect, will not be an isolated one. Future development of parallel approaches to molecular dynamics simulations will require numerical experiments on conventional supercomputers.

[1]The latter step does not require much time; however, it is so closely linked to the force evaluation and, therefore, we do not want to separate the two.

A most recent review on large-scale molecular dynamics simulations of simple liquids using vector and parallel processors has been given by Rapaport [13]. This review emphasized, however, methods for handling systems of structureless particles with simple short-range interactions and, therefore, computational strategies discussed differ in most respects from the ones for biopolymers presented below. The differences are the following: (i) The force fields to be employed for biopolymers are more complex than those of condensed matter systems, i.e. there is a large family of different forces as explained briefly in Sect. 2; (ii) Furthermore, biopolymers most often are completely heterogeneous, i.e. no translational symmetry exists; (iii) The native structures of biopolymers are non-trivial, i.e. they cannot be determined by reasonable guesses and simulations, but rather need to be known beforehand from analyses of X-ray scattering data; (iv) The atoms in biopolymers have an intricate and heterogeneous pattern of chemical bonds which determines the force field and, hence, needs to be known to the simulation program. These aspects require computational approaches which differ from those taken for molecular dynamics simulations of condensed matter systems, e.g. liquids and crystals, and only few computational strategies can be shared between the approaches.

A most pertinent review on concurrent computation for molecular dynamics simulation covering algorithms for homogeneous as well as heterogeneous, e.g. biopolymer, systems has been provided by Fincham [14]. This review has influenced the work reported here, in particular, since it discussed the use of Transputers, a systolic loop architecture and respective algorithms for the evaluation of pair interactions, the latter constituting the most time-consuming task in molecular dynamics simulations.

In Section 2 we review briefly the computational aspects of protein simulation. In Section 3 we discuss the strategy of parallel computation which is dictated by the long range character of the Coulomb forces in proteins. In Section 4 we describe the parallel computer, i.e. its nodes and board design. In Section 5 we introduce our parallel program. In Section 6.1 we provide benchmark tests for actual molecular dynamics simulations comparing computational speeds with those of XPLOR running on various conventional machines producing the same output files (trajectories of various proteins). In Section 6.2 we discuss the costs of our computer system. This discussion reveals that the suggested parallel computation achieves biopolymer simulation for a much lower price than sequential computations. Finally, in Section 7, we present an application of parallel simulation, namely the relaxation of the complete structure of the photosynthetic reaction center, a system of 12 600 atoms, to an equilibrium geometry.

2. NUMERICAL TASKS IN MOLECULAR DYNAMICS SIMULATIONS

In this Section we review briefly the computational aspects of molecular dynamics simulations and discuss the relationship between our parallel algorithm and the programs CHARMM [6] and XPLOR [7, 8], to which our algorithm is closely related.

Computer simulations of biological macromolecules are based on a classical mechanical model of biomolecules. For the nuclei of the N atoms of a molecule the Newtonian equations of motion ($i = 1, 2, \ldots N$) are assumed to hold

$$m_i \ddot{\vec{r}}_i = -\nabla_i E(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) \tag{1}$$

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.