Journal of Computational Physics 246 (2013) 275–288

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Exploring large macromolecular functional motions on clusters of multicore processors

José R. López-Blanco [a], Ruymán Reyes [b], José I. Aliaga [b,*], Rosa M. Badia [c], Pablo Chacón [a], Enrique S. Quintana-Ortí [b]

[a] Instituto de Química Física Rocasolano, Spanish National Research Council (CSIC), 28006 Madrid, Spain
[b] Depto. de Ingeniería y Ciencia de Computadores, Universidad Jaume I, 12071 Castellón, Spain
[c] Computer Sciences Department, Barcelona Supercomputing Center (BSC-CNS), 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 21 December 2012
Received in revised form 12 March 2013
Accepted 15 March 2013
Available online 27 March 2013

Keywords:
Computational biology
Macromolecular machines
Normal modes
Eigenvalue problems
Krylov-subspace method
Clusters of multicore processors

ABSTRACT

Normal modes in internal coordinates (IC) furnish an excellent way to model functional collective motions of macromolecular machines, but exhibit a high computational cost when applied to large-sized macromolecules. In this paper, we significantly extend the applicability of this approach towards much larger systems by effectively solving the computational bottleneck of these methods, the diagonalization step and associated large-scale eigenproblem, on a small cluster of nodes equipped with multicore technology. Our experiments show the superior performance of iterative Krylov-subspace methods for the solution of the dense generalized eigenproblems arising in these biological applications over more traditional direct solvers implemented on top of state-of-the-art libraries. The presented approach expedites the study of the collective conformational changes of large macromolecules opening a new window for exploring the functional motions of such relevant systems.

1. Introduction

The most important components of living cells are proteins and nucleic acids which are formed by long unbranched chains of aminoacids and nucleotides, respectively. They usually assemble together into large macromolecular machines to support the main biological functions. For example, the ribosomal machinery produces new proteins according to the genetic code; the chaperonin proteins assist the folding process of these new-formed proteins; and tubulin and actin filaments support the cellular shape. At the molecular level, the biological activity of these components is frequently tackled by analyzing their dynamics and interactions. However, the direct experimental observation of the functional motions is quite complex and computational studies are often the only way to predict conformational changes. Molecular Dynamics (MD) simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids using available 3D atomic structures. Unfortunately, the large size of the macromolecules and the long time scale of their motion render MD simulations too costly or even prohibitive.

In recent years, the use of coarse-grained models (CG), i.e. reduced representations of biopolymer structures obtained by grouping atoms into simplified entities or pseudo-atoms, has naturally provided significant computational savings by reducing the number of system variables [32]. In this context, CG merged with normal mode analysis (NMA) has arisen as a powerful and popular approach to simulate collective motions of macromolecular complexes at extended time scales [6]. In

* Corresponding author. Tel.: +34 964728297.
E-mail addresses: jrlopez@iqfr.csic.es (J.R. López-Blanco), rreyes@uji.es (R. Reyes), aliaga@uji.es (J.I. Aliaga), rosa.m.badia@bsc.es (R.M. Badia), pablo@chaconlab.org (P. Chacón), quintana@uji.es (E.S. Quintana-Ortí).

0021-9991/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcp.2013.03.032

biomolecular simulations, NMA plays an important role to analyze structural fluctuations around a well-defined conformation. In particular, CG-NMA has been successfully applied to calculate biologically relevant motions of proteins and nucleic acids [19] even at very low resolutions [11], X-ray refinements [12], flexible fitting of atomic structures into electron microscopy density maps [18], efficient generation of conformational pathways [15], and ligand docking [10,22]. These results and many others validate the use of CG-NMA modeling in describing molecular flexibility and demonstrate how this approach has become, in practice, a powerful alternative to costly atomistic simulations.

We recently extended the applicability of such approaches to larger molecules by formulating the NMA in internal coordinates (ICs), i.e. using the dihedral angles of the macromolecular chain instead of the Cartesian coordinates (CCs) of the atoms [21]. Although ICs reduce the number of degrees of freedom (DOF) in comparison to CCs, the diagonalization step within this procedure remains a major computational bottleneck for large molecules, even when using CG-reduced approximations. Biologically relevant molecular machines such as ribosome complexes, actin filaments, microtubules, viral capsids and many others either are beyond the applicability range or can only be addressed by very aggressive CG approximations (e.g. a single pseudo-atom per molecule). Thus, additional methodological work must be performed to extend the applicability to both larger macromolecular systems and more detailed representations.

In this paper we address the parallel simulation of the motions of large macromolecular complexes via IC-NMA, leveraging clusters of computers equipped with multicore processors to tackle the diagonalization bottleneck. Specifically, we make the following contributions:

- For this biological relevant application, we illustrate the advantages of the Krylov subspace method for the solution of *dense* generalized eigenproblems over more conventional "direct" solvers based on LAPACK. While the superior performance of Krylov subspace methods was hinted in [2] for general problems of moderate size on multicore processors and hardware accelerators, here we extend the analysis to the large-scale eigenproblems arising in biomolecular simulation using clusters of multicore processors.
- We improve the performance of all the message-passing eigensolvers by combining MPI with the SMPSs [29] runtime to obtain an enhanced version of the Cholesky factorization, a crucial and highly parallel operation for all eigensolvers considered in this work (including the direct solver).
- The drastic reduction of computing time obtained with the proposed parallel diagonalization scheme is demonstrated by simulating the collective motions of a representative set of large macromolecular machines. In particular, our experimental study includes a detailed performance evaluation as well as a survey of the essential functional motions of some key biological systems extracted with IC-NMA.

The rest of the paper is structured as follows. In Section 2 we briefly review the IC-NMA method for the simulation of macromolecules, exposing the underlying eigenvalue problem that currently conforms the major computational bottleneck of this approach. In Section 3 we introduce the Krylov-subspace iterative method for the solution of generalized symmetric definite eigenproblems, together with its explicit and implicit variants, and discuss their arithmetic cost and memory requirements. Next, in Section 4, we describe the implementation of these two variants using existing message-passing linear algebra libraries for clusters of computers, as well as an alternative parallelization of the Cholesky factorization that leverages intranode concurrency using the SMPSs runtime. In Section 5 we evaluate the performance of the different methods, including a conventional "direct" solver from LAPACK, which can be considered as the state-of-the-art for the solution of this type of problems. In that section we also include several 3D animations showing selected motions of large macromolecular machines that were obtained using IC-NMA and the fast parallel methods presented here. Finally, we close the paper offering some concluding remarks and a discussion of future work in Section 6.

2. Simulation of macromolecules via IC-NMA

In NMA the relevant collective motions are described by approximating both potential and kinetic energies as quadratic functions of the atomic positions and velocities, respectively. This simplification allows the decomposition of the motion into a series of deformation vectors which encode the potential displacement directions. Such vectors, or modes, are obtained by diagonalizing the second derivative matrices of both potential (Hessian) and kinetic energies. Every mode exhibits a characteristic frequency that is inversely proportional to the deformation energy cost. High frequency modes represent stiff localized displacements, whereas low frequency modes correspond to soft collective conformational changes. The latter deformations have been closely related to functional motions [19,36], and it is also well-known that there is a good correspondence with the essential motions extracted from atomistic MD [24,27,1].

Following the seminal works of Go and others [16,20], we have implemented a complete mathematical NMA framework using dihedral angles as variables. This new multipurpose tool chest, named iMOD [21], exploits the benefits of classical NMA formulations in internal coordinates (ICs) while extending them to cover multiscale modeling, even for huge macromolecular structures. iMOD has two major advantages over current Cartesian approaches: it implicitly preserves model geometry to minimize potential distortions and substantially reduces the number of DOFs to improve efficiency. The software also includes several atomic CG representations but the ICs are always defined by both the canonical backbone dihedral angles, i.e. ϕ and ψ in proteins (see Fig. 1 for details), and the 6 rotational/translational DOFs between chains.