Journal of Molecular Graphics and Modelling 68 (2016) 68–77

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

High performance computing for three-dimensional agent-based molecular models

G. Pérez-Rodríguez [a], M. Pérez-Pérez [a], F. Fdez-Riverola [a], A. Lourenço [a,b,*]

[a] ESEI – Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, Universidad de Vigo, 32004 Ourense, Spain
[b] CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-0668 Braga, Portugal

ARTICLE INFO

Article history:
Received 23 November 2015
Received in revised form 26 May 2016
Accepted 7 June 2016
Available online 8 June 2016

Keywords:
Cellular systems
Agent-based modelling
Three-dimensional model representation
Individual molecule simulation
Distributed computing

ABSTRACT

Agent-based simulations are increasingly popular in exploring and understanding cellular systems, but the natural complexity of these systems and the desire to grasp different modelling levels demand cost-effective simulation strategies and tools.

In this context, the present paper introduces novel sequential and distributed approaches for the three-dimensional agent-based simulation of individual molecules in cellular events. These approaches are able to describe the dimensions and position of the molecules with high accuracy and thus, study the critical effect of spatial distribution on cellular events. Moreover, two of the approaches allow multi-thread high performance simulations, distributing the three-dimensional model in a platform independent and computationally efficient way.

Evaluation addressed the reproduction of molecular scenarios and different scalability aspects of agent creation and agent interaction. The three approaches simulate common biophysical and biochemical laws faithfully. The distributed approaches show improved performance when dealing with large agent populations while the sequential approach is better suited for small to medium size agent populations.

Overall, the main new contribution of the approaches is the ability to simulate three-dimensional agent-based models at the molecular level with reduced implementation effort and moderate-level computational capacity. Since these approaches have a generic design, they have the major potential of being used in any event-driven agent-based tool.

1. Introduction and motivation

New advances in super-resolution and super-localisation techniques have allowed experimental molecular biophysics and biochemistry to go beyond ensemble measurements and obtain data at the single molecule level [4,15]. Such experiments are able to track down key motions, reactions, and interactions of individual molecules with high temporal and spatial resolution. However, the acquisition of such data is very time-consuming, partially because the techniques are not yet advanced enough to allow the simultaneous observation of a wide range of molecule types [26,30].

An alternative to the use of experimental techniques is to assemble molecular models *in silico* and to use simulation techniques to explore their behaviour. Such computational models have the potential to elucidate structure and auto-organisation between molecules as well as complex molecular interplay that are difficult to observe *in vivo* or *in vitro*. In this context, the challenge presented to computational methodologies is to embrace the natural complexity of cellular systems as faithfully as possible [51]. A realistic model, depending on the cellular system at hand and the questions to be asked, should cope with spatial and temporal scales of various orders of magnitude and different levels of modelling detail. Biologically relevant time scales range from nanoseconds to microseconds for the internal dynamics of individual molecules; cellular dimensions are between 300 nm for the smallest bacterial cells and 100 μm for large eukaryotic cells; and, the atomistic structural description of molecules requires spatial resolution in the nanometre range [14]. Moreover, it is important to model the molecules and the environment as volumes in order to have full awareness of the spatial location of the molecules, and the implications over biophysics (e.g. collisions) and biochemical rules (e.g. reaction radius) [33].

So, although it is conceivable to model cellular processes at single molecule level, such fine grain simulation demands

* Corresponding author at: ESEI – Department of Computer Science, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain.
E-mail addresses: gprodriguez2@esei.uvigo.es (G. Pérez-Rodríguez), mpperez3@esei.uvigo.es (M. Pérez-Pérez), riverola@uvigo.es (F. Fdez-Riverola), analia@uvigo.es (A. Lourenço).

http://dx.doi.org/10.1016/j.jmgm.2016.06.001
1093-3263/

considerable computing power. In this context, general purpose Graphical Processing Unit (GPU) technology and multi-core CPU processors are being used to parallelise biomolecular simulations [20,21,23,40]. However, such frameworks do not yet support three-dimensional modelling and are not meant to be deployed in general biological research settings, specifically in research centres or labs that do not have access to high performance computing clusters, parallel architectures and GPU hardware, nor have the technical expertise to write efficient software for these environments. So, it is often the case that computational biologists resort to coarser resolution approaches to simulate large biological systems [41,43]. Coarser models keep a reduced and essential number of degrees of freedom and interactions, which decreases the computational requirements of the simulation and allows the execution of simulations in computers with moderate computational capacity. Still, this reduction can go so far and, in practice, single cell coarse models also demand the implementation of simulation optimisation algorithms.

In this context, this work aims to contribute to the simulation of individual molecules in cellular events in two ways: to enable the realistic simulation of the spatial location, diffusion and interaction of molecules in three-dimensional, continuous environments; and, to improve the generic template of agent-based approaches so that the computational costs of such realistic simulations are affordable. Most notably, the aim is to enable realistic simulation in computers with a moderate-level computational capacity, i.e. computers broadly available in biological labs and that do not required advanced programming skills. Hence, our work proposes solutions that may be applied to virtually any model at the molecular level and may be used in any event-driven agent-based tool.

The next sections describe the capabilities of existing ABM software for biomolecular simulation, the solutions devised for scalable three-dimensional single molecule simulation, and the analysis of performance results for the proposed approaches.

2. Related work: existing approaches for biomolecular modelling and simulation

Agent-based models (ABM) are a well-known and favoured modelling strategy for biomolecular systems [7]. Generically, these models are composed by a population of heterogeneous agents, which represent the molecules under study, including their shape, size and interaction logic. Biomolecular events unfold on an explicit and specific environment (e.g. representing the cytoplasmic environment) where agents act autonomously, executing some sort of itinerary (e.g. molecular diffusion). Each agent class is implemented to represent the features and behavioural responses of a specific type of molecules (e.g. enzymes and metabolites). Agents interact with one another following common biochemical and biophysical assumptions (e.g. enzymatic kinetics), and their behaviour may adapt depending on the perceived situation, and most notably the influence of the immediate surroundings (e.g. abundance of substrate).

2.1. ABM for biomolecular modelling

Individual particle or agent-based modelling is one of the current favoured alternatives for biomolecular representation, as it is naturally suited to describe single molecule behaviour and help understanding phenotypic heterogeneity and cell-to-cell variability.

Among the earliest biomolecular models one may find Cellulat, an agent-based intracellular signalling model [18], and Epitheliome, which represents the growth and repair characteristics of epithelial cell populations [49]. Today, the range of agent-based model applications is quite broad. For example, these models have been used to represent the Ras–MAPK [17] and NF-kB [38] intracellular signalling pathways, *Escherichia coli* cytoplasm dynamics [30], bacterial phenotypic switching [46], epithelial host-pathogen interactions [45], cancer systems biology [29], development of restenosis in blood vessels [12], autophagy dynamics and sub-mitochondrial heterogeneity [8], intracellular phosphorus heterogeneity in cultured phytoplankton [16], oxygen metabolism in aerobic-anaerobic respiration [5], and the design of cellulase systems [3].

A deep description of general agent-based modelling approaches and common "recipes" used in biomolecular modelling are out of the scope of this work. We recommend reading the following state-of-the-art reviews for gaining further understanding about the general aspects of ABM and computational social science [7,10], the use of ABM to model biological complexity across biological scales [1,22], the construction of biological ABM under the Systems Biology perspective [31,34] and existing models of biomolecules in cellular environments [14].

2.2. Software platforms for ABM simulation

Commonly, ABM software follows the framework and library paradigm, i.e. a framework that provides a set of standard concepts for designing and describing ABMs, and a software library implementing the framework and providing simulation tools. Swarm was one of the first ABM software [19]. It was written in Objective-C and served as inspiration to most of the more recent software. The well-known Recursive Porous Agent Simulation Toolkit (Repast) started as a Java implementation of Swarm, but evolved on its own [35]. The project has released several software toolkits and development environments (in Java, Python and NET) and recently Repast has been superseded by a significant development named Repast Simphony, or Repast-S [36]. The Multi-Agent Simulation of Neighbourhoods (MASON) is another prominent toolkit, which was designed as a smaller and faster alternative to Repast, focused on computationally demanding models with many agents executed over many iterations [27].

Alternative development exists and NetLogo, from the Logo family of platforms, is one of the best representatives [44]. Here, the purpose is mainly educational, more specifically to provide a high-level platform that allows non-skilled users to build and learn from simple ABMs. Nevertheless, the platform now contains many sophisticated capabilities (behaviours, agent lists, graphical interfaces, etc.) and it has been used in biomedical applications.

Throughout the years, several studies have analysed and compared the evolution of these platforms in terms of conceptual basis, programming experience, execution speed, development priorities, and ease of use, i.e. the intricacy of implementing ABMs with them [25,28,39,47]. NetLogo stands out for models that are compatible with its paradigm of short-term, local interaction of agents and a grid environment and are not extremely complex. Java Swarm is outperformed by more recent, alternative Java platforms. Overall, although current multi-agent platforms demonstrate ability for modelling and understanding phenomena of increasing complexity, new releases and new platforms keep emerging. So, it is not straightforward to choose a platform, in particular to those looking to develop domain-specific software. This decision is usually left to survey articles, past domain application experiences and platform publicity.

2.3. Biomolecular-specific agent-based simulation

In the case of biomolecular simulation, one may observe that usually development does not rely on general ABM platforms. Well-established biomolecular simulators such as ReaDDy [42], Smoldyn