Introduction

Author for correspondence:
Radek Erban
e-mail: erban@maths.ox.ac.uk

Multi-resolution simulations of intracellular processes

Radek Erban[1], Sarah Harris[2] and Raffaello Potestio[3,4]

[1]Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK
[2]School of Physics and Astronomy and Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, UK
[3]Physics Department, University of Trento, via Sommarive, 14 38123 Trento, Italy
[4]INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, 38123 Trento, Italy

(iD) RE, 0000-0001-8470-3763; RP, 0000-0001-6408-9380

This is an introduction to the Thematic Issue of *Interface Focus* containing papers by speakers of the Theo Murphy International Scientific Meeting on 'Multi-resolution simulations of intracellular processes'

The role of computational methods in life sciences is becoming increasingly prominent. Experiment is not only supported by numerical investigation, it is often necessary to substitute computational studies to explore resolutions in time and space that would otherwise be impossible to probe. Furthermore, the 'computer experiment' allows the test of hypotheses through the construction of models whose fundamental properties and interactions are fully under the researcher's control, thus removing all accessory and/or unnecessary detail which hinders interpretation.

The Theo Murphy International Scientific Meeting on 'Multi-resolution simulations of intracellular processes' gathered scientists from all over the world to discuss pivotal aspects of such computer-aided studies of the physical mechanisms underlying biological processes. It was held at the Kavli Royal Society Centre in Chicheley Hall (UK) on 24–25 September 2018. Particular emphasis was placed on the intrinsic multi-scale nature of the properties and processes which characterize living systems from the atomistic level upwards. *In silico* methods developed to study biological matter need, therefore, to cover a range of length and time scales which, from the ångström to the metre, from the femtosecond to the second, span 10–15 orders of magnitude. A variety of different strategies has been pursued for correspondingly different tasks; however, in the past few years, the need has emerged to *integrate* some of these techniques into multiple-resolution methods capable of bridging different scales. The aim of this workshop was thus to illustrate the state of the art in this research area, pinpoint some of the crucial problems and challenges, and discuss the potential for a coordinated, far-reaching long-term community strategy for multi-scale, multi-resolution biological simulations.

The most fundamental differences among the computational methods employed in the study of biological systems arise from the level of resolution a given model has to achieve in order to investigate a particular property or process, ranging from *ab initio* methods through all-atom molecular dynamics to very coarse-grained stochastic models. The most detailed approaches covered during our Scientific Meeting included quantum mechanics. Crnjar *et al.* [1] start their discussion of ligand-gated ion channel models with first principle methods based on the density functional theory, a common computational quantum mechanical approach. A similar quantum mechanical starting point is mentioned in Machado *et al.* [2], which use double-stranded DNA as their illustrative example.

Both contributions [1,2] discuss all-atom molecular dynamics (MD, or molecular mechanics) as a coarser description of the studied system. MD can under certain assumptions provide all important information on the system properties and the same philosophy (that a model written in one resolution is good

enough) is often applied to models written in terms of a much coarser stochastic approach. However successful in many cases, this 'layered' view of the hierarchies of scales is often too simplistic and does not account for the intrinsic multi-scale nature of complex biological systems, which may involve interplay and feedback loops among the various scales. In order to tackle these specific problems, concurrent multi-resolution methods have been developed, where the same system is described with two or more models at different resolution in the same set-up. These models enable simulations of large systems with relatively low computational expenditure, while at the same time preserving a high accuracy in predefined, relevant subdomains of interest. Such a philosophy is shared by a number of contributions. Zavadlav et al. [3] consider methods for coupling atomistic and supramolecular water models with applications to solvated proteins. Gunaratne et al. [4] analyse this approach using two theoretical heat baths, for which one can prove the convergence of the detailed molecular dynamics model to a coarser stochastic modelling approach. They model a biomolecule at a much coarser bead-spring approach. A computational study of bead-spring polymer models, based on the classical MD, is then presented in Giunta & Carbone [5], who show that both Rouse-like and Zimm-like dynamics can be observed for polymers at the interface between two immiscible liquids.

Current limits of all-atom MD are tested by Farafonov & Nerukh [6], who present modelling of the whole virus capsid of an MS2 bacteriophage particle (without its genome) using all-atom MD. For such relatively large systems, a coarse-grained modelling approach is often the method of choice in applications. Berg & Peter [7] present such a simulation approach, based on the coarse-grained Martini model, which they use for understanding protein–protein interactions, while Pasquali et al. [8] study an RNA model, where each nucleotide is described by 6 or 7 beads. Continuing further up the scales, Floyd et al. [9] present their MEDYAN simulation platform which enables them to simulate the whole network of proteins, including actin and myosin filaments. At the tip of this pyramid, we find the work by Wijeratne & Vavourakis [10], who present their in silico cancer simulator based on solid and fluid mechanics equations.

A further aspect of the computer-based study of molecular and cellular biology is that of machine learning. This galaxy of powerful methods is opening up entire new lines of research, ranging from applications in bioinformatics (like genome sequencing and structure prediction) to automated data analysis. So far, machine learning methods have been applied most successfully to improve the efficiency of existing techniques. In many cases, these methods have also been employed in a fully novel manner, thus providing a further instrument in the researcher's toolbox. A deep learning approach to the structural analysis of proteins is presented by Giulini & Potestio [11], which can be used to identify mechanically relevant regions of the molecule. Lee et al. [12] then discuss nonlinear information fusion algorithms and data-driven higher-dimensional embeddings.

We are now experiencing the pervasiveness of social media and other communication instruments. The immense volume of information traffic on our laptops and mobiles, however, does not always imply that knowledge is acquired effectively—and its quality is often difficult to assess. It is thus of paramount importance to develop effective strategies to communicate the fundamental underpinnings of the scientific progress and technological developments, especially to the younger generations. This effort is carried out by Taly et al. [13], who discuss an interdisciplinary workshop developed for high school and junior undergraduate students in France. It covers topics in biology, chemistry, physics, mathematics and computer science, which are all needed to understand MD and its applications to simulations of biological systems.

One of the common themes mentioned in discussions during our Scientific Meeting was the immense amounts of data which can now be generated so easily it often surpasses a researcher's archival capacity. Therefore, distribution and sharing of large datasets are challenging issues. Data-sharing benefits the scientific community by making best use of one's research results, e.g. by enabling other groups to perform subsequent independent analysis after publication. For this reason, and to guarantee and enforce good scientific practices, our last contribution is an opinion article by Riccardi et al. [14], who conclude that it would be extraordinarily beneficial to develop a central—yet not centralized—platform for the storage of biophysics data and software, very much in line with similar efforts carried out in the soft matter and material science communities.

References

1. Crnjar A, Comitani F, Melis C, Molteni C. 2019 Mutagenesis computer experiments in Pentameric ligand-gated ion channels: the role of simulation tools with different resolution. *Interface Focus* **9**, 20180067. (doi:10.1098/rsfs.2018.0067)

2. Machado MR, Zeida A, Darré L, Pantano S. 2019 From quantum to subcellular scales: multi-scale simulation approaches and the SIRAH force field. *Interface Focus* **9**, 20180085. (doi:10.1098/rsfs.2018.0085)

3. Zavadlav J, Marrink SJ, Praprotnik M. 2019 SWINGER: a clustering algorithm for concurrent coupling of atomistic and supramolecular liquids. *Interface Focus* **9**, 20180075. (doi:10.1098/rsfs.2018.0075)

4. Gunaratne RS, Wilson DB, Flegg MB, Erban R. 2019 Multi-resolution dimer models in heat baths with short-range and long-range interactions. *Interface Focus* **9**, 20180070. (doi:10.1098/rsfs.2018.0070)

5. Giunta G, Carbone P. 2019 Cross-over in the dynamics of polymer confined between two liquids of different viscosity. *Interface Focus* **9**, 20180074. (doi:10.1098/rsfs.2018.0074)

6. Farafonov VS, Nerukh D. 2019 MS2 bacteriophage capsid studied using all-atom molecular dynamics. *Interface Focus* **9**, 20180081. (doi:10.1098/rsfs.2018.0081)

7. Berg A, Peter C. 2019 Simulating and analysing configurational landscapes of protein–protein

contact formation. *Interface Focus* **9**, 20180062. (doi:10.1098/rsfs.2018.0062)

8. Pasquali S, Frezza E, Barroso da Silva FL. 2019 Coarse-grained dynamic RNA titration simulations. *Interface Focus* **9**, 20180066. (doi:10.1098/rsfs.2018.0066)

9. Floyd C, Papoian GA, Jarzynski C. 2019 Quantifying dissipation in actomyosin networks. *Interface Focus* **9**, 20180078. (doi:10.1098/rsfs.2018.0078)

10. Wijeratne PA, Vavourakis V. 2019 A quantitative *in silico* platform for simulating cytotoxic and nanoparticle drug delivery to solid tumours. *Interface Focus* **9**, 20180063. (doi:10.1098/rsfs.2018.0063)

11. Giulini M, Potestio R. 2019 A deep learning approach to the structural analysis of proteins. *Interface Focus* **9**, 20190003. (doi:10.1098/rsfs.2019.0003)

12. Lee S, Dietrich F, Karniadakis GE, Kevrekidis IG. 2019 Linking Gaussian Process regression with data-driven manifold embeddings for nonlinear data fusion. *Interface Focus* **9**, 20180083. (doi:10.1098/rsfs.2018.0083)

13. Taly A, Nitti F, Baaden M, Pasquali S. 2019 Molecular modelling as the spark for active learning approaches for interdisciplinary biology teaching. *Interface Focus* **9**, 20180065. (doi:10.1098/rsfs.2018.0065)

14. Riccardi E, Pantano S, Potestio R. 2019 Envisioning data sharing for the biocomputing community. *Interface Focus* **9**, 20190005. (doi:10.1098/rsfs.2019.0005)