Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Multiscale biomolecular simulations in the exascale era

David Carrasco-Busturia[1,a], Emiliano Ippoliti[2], Simone Meloni[3], Ursula Rothlisberger[4] and Jógvan Magnus Haugaard Olsen[1]

Abstract

The complexity of biological systems and processes, spanning molecular to macroscopic scales, necessitates the use of multiscale simulations to get a comprehensive understanding. Quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulations are crucial for capturing processes beyond the reach of classical MD simulations. The advent of exascale computing offers unprecedented opportunities for scientific exploration, not least within life sciences, where simulations are essential to unravel intricate molecular mechanisms underlying biological processes. However, leveraging the immense computational power of exascale computing requires innovative algorithms and software designs. In this context, we discuss the current status and future prospects of multiscale biomolecular simulations on exascale supercomputers with a focus on QM/MM MD. We highlight our own efforts in developing a versatile and high-performance multiscale simulation framework with the aim of efficient utilization of state-of-the-art supercomputers. We showcase its application in uncovering complex biological mechanisms and its potential for leveraging exascale computing.

Addresses

[1] DTU Chemistry, Technical University of Denmark (DTU), Kongens Lyngby, DK-2800, Denmark

[2] Computational Biomedicine, Institute of Advanced Simulations IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, Jülich, DE-52428, Germany

[3] Dipartimento di Scienze Chimiche, Farmaceutiche ed Agrarie (DOCPAS), Università degli Studi di Ferrara (Unife), Ferrara, I-44121, Italy

[4] Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, CH-1015, Switzerland

Corresponding author: Olsen, Jógvan Magnus Haugaard (jmho@kemi.dtu.dk)

(Carrasco-Busturia D.), (Meloni S.), (Rothlisberger U.), (Olsen J.M.H.)

[a] Present address: Division of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden.

Current Opinion in Structural Biology 2024, **86**:102821

This review comes from a themed issue on Theory and Simulation/Computational Methods (2024)

Edited by **Paolo Carloni** and **Karissa Y. Sanbonmatsu**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2024.102821

Introduction

Biological processes extend across wide scales in space and time due to the hierarchical organization of biological matter [1]. The characteristic dimensions span from the molecular scale of a few ångströms with ultrafast electronic processes on the order of atto- and femtoseconds and rapid chemical reactions that occur within pico- to microseconds, to the macroscopic scale of cells and organs that are visible to the naked eye and where processes extend to seconds and even days and years. The complexity in living organisms largely stems from this hierarchical structure, where a local process may trigger a cascade of events across multiple spatial and temporal scales. Thus, multiscale approaches integrating different resolutions and methodologies are essential for capturing the entire spectrum of biological events [2,3].

The continuous development of multiscale methods in computational structural biology is driven by simultaneous advancements in algorithms, software implementations, and hardware technology that push the boundaries of molecular simulations in terms of accessible time scales and system sizes, capturing biological systems from the atomistic to the cellular level. In particular, quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulations have become increasingly important in the last decades for studying processes where a description of electronic degrees of freedom is paramount and thus beyond the capabilities of classical MD based on standard analytical force fields. Examples of such events encompass all types of chemical reactions, including proton-coupled electron

transfers [4,5], photochemistry [6], and the manifold of chemical transformations observed in enzymatic reactions [7,8], especially those involving transition metals [9]. A detailed characterization of these phenomena, including reactants, transition states, intermediates, and products, as well as the involved relative free energies, reaction rates, and binding affinities in both electronic ground and excited states, also has direct implications for drug design [10]. In QM/MM models of biomolecular systems, a smaller part (the QM subsystem), such as the active site of an enzyme or a chromophore embedded in a protein or lipid bilayer, is described at the quantum-mechanical level, while the remainder (the MM subsystem) is modeled using molecular mechanics. This multiscale strategy balances a detailed and accurate but computationally costly description of the essential part of a system with a coarser but computationally expedient approach for the much larger MM part. The dynamic aspect of QM/MM MD is crucial to accurately capture the behavior and function of complex proteins [11,12]. However, QM/MM MD simulations have a substantially higher computational cost than classical (i.e., MM) MD simulations, severely limiting the accessible time scales. Typically, for a density functional theory (DFT)-based QM/MM MD simulation with around 100 atoms in the QM subsystem, accessible time scales are limited to a few hundred picoseconds [8,13].

The advent of exascale computing marks a pivotal moment for all simulation-based scientific fields [14,15]. This remarkable technological achievement was accomplished by connecting thousands of computing nodes through high-speed network interconnects. Each node combines traditional general-purpose central processing units (CPUs) with powerful graphics processing units (GPUs). However, exascale supercomputer architectures also introduce new challenges since programming software applications for these heterogeneous machines requires a judicious decomposition of the computational work to exploit each component of the machines optimally [16]. Therefore, to fully exploit the computational power of present and future heterogeneous HPC architectures, a high degree of concurrent parallelism is needed, where different parts of the supercomputer work on different subdomains of the computational model, each described within a different theoretical method. Together, exascale computing and the development of novel computational methods and software can enable longer and more accurate simulations on larger and more complex systems. This opens up new opportunities for discovery and innovation in the life and health sciences, potentially revolutionizing areas such as drug design and bioengineering.

The heterogeneous CPU/GPU technology highlighted above is taken to the next level with modular supercomputer architectures that integrate a variety of hardware technologies into interconnected partitions [17,18]. A prime example of this is the LUMI supercomputer [19], along with the upcoming exascale JUPITER supercomputer [20], both procured by the European HPC Joint Undertaking (EuroHPC JU). Currently, LUMI consists of two primary *number-crunching* partitions: a general-purpose CPU partition and a high-performance GPU-accelerated partition. Additionally, it features an interactive data analytics partition and an accelerated storage partition. Looking ahead, the integration of prospective technologies like quantum and neuromorphic computing into traditional HPC infrastructure is expected [17,18]. Indeed, the procurement of a quantum computing partition for LUMI is already underway [21], and the Jülich Supercomputing Center in Germany is already experimenting with this kind of integration [22]. Modular supercomputer architectures offer substantial benefits, particularly in allowing calculations to run on the most suitable hardware for the specific problem. Moreover, existing software packages that have yet to be optimized for new hardware remain useful for solving important scientific problems. Fully exploiting the capabilities of complex modular supercomputers for exceptionally challenging scientific problems, which require the full computing power of the machine, necessitates a new computational paradigm. Here, we highlight an approach recently introduced in the field of multiscale biomolecular simulations.

Multiscale methods, especially the QM/MM approach, have proven indispensable for exploring complex biological phenomena. Interestingly, QM/MM is not only a robust technique in itself but also offers a route to overcome some of the existing limitations in the development of software for atomistic simulations on modern hybrid computer architectures [14]. Beyond QM/MM, multiscale methods can also be extended to incorporate multiple layers ranging from coarse-grained (CG) to continuum models of matter where parts of a system, e.g., membrane regions that are sufficiently distant from an embedded protein of interest, are treated by techniques usually employed for meso- and macroscopic systems [23–27].

To reach the full potential offered by exascale supercomputers, it is imperative that multiscale interfaces scale efficiently, fully leveraging the extensive network of CPUs and GPUs. That is one of the main objectives of MiMiC (multiscale modeling in computational chemistry), a high-performance and versatile framework for multiscale simulations [28]. The *program-agnostic* MiMiC framework is designed to combine virtually any QM and MM (or other) program without compromising the computational efficiency and scalability of the simulation. Indeed, MiMiC, although still in its nascent stages, has demonstrated its ability to efficiently scale QM/MM MD simulations across thousands of CPU-based computing nodes [13,29]. Here we will showcase the MiMiC framework for biomolecular simulations,

illustrating its potentiality in leveraging state-of-the-art supercomputing resources and its capability to address complex biochemical and biophysical problems.

Current status of multiscale biomolecular simulation software

Multiscale QM/MM capabilities are often implemented by either extending dedicated QM or MM programs with the functionalities of the other [30–37] or by creating ad hoc interfaces between stand-alone QM and MM programs [38–41]. Additionally, some programs offer flexible interfaces that facilitate easy coupling with other programs [42–44]. Beyond these, there are integrative frameworks that do not contain inherent QM, MM, or similar functionalities, but instead depend entirely on other programs for these capabilities [45–59].

The interface between QM and MM components can be classified as tightly or loosely coupled, reflecting the degree of interdependence between the components. Loose coupling is preferred for its flexibility and ease of maintenance, enabling straightforward integration of independent programs. This approach supports the quick incorporation of new functionalities and advancements within individual programs without necessitating alterations to other coupled components. Crucially, loose coupling permits independent optimization of each program for peak performance. However, this flexibility can come at the cost of computational efficiency due to slower inter-program communication, particularly when compared to the tight coupling approach that allows for direct in-memory data sharing.

There are three communication mechanisms for exchanging data between programs, namely the file-, library-, and network-based approaches. Most commonly utilized for general interfaces and integrative frameworks is the file-based approach, which does not require modifications to the source code of the interfaced programs, making it the simplest to implement. Here, the main driver program generates input files for the interfaced programs, executes them, and subsequently reads their output files. However, it is notably the least efficient communication mechanism, primarily due to slow disk write/read operations and overhead associated with executing and terminating the interfaced programs. Both drawbacks are avoided in the library-based approach, where the interfaced programs are converted into a library that is linked to the main program, thus enabling efficient in-memory data exchange. The disadvantages are the rather intrusive modifications needed in the source codes of the interfaced programs, and the risk of ending up with a tight coupling focused on a few specific programs. Moreover, a critical aspect of multiscale implementations is parallel scalability and efficiency. The optimal parallel algorithms for QM and MM components, or even among different QM methods, may differ significantly and may not seamlessly integrate within the more tightly coupled environments.

The network-based approach, on the other hand, offers a solution that potentially circumvents the limitations inherent in both file- and library-based methods. It achieves a balance between flexibility and communication efficiency by enabling data exchange over a network. This facilitates seamless communication between programs running on different processors (CPUs and GPUs), across computing nodes, or even among supercomputer partitions. Such an approach retains the benefits of loose coupling, i.e., ease of integration and maintenance, while surpassing file-based approaches in performance through the use of fast network protocols and remote direct memory access (RDMA) technologies. Crucially, it can be implemented so as not to disrupt the normal execution of interfaced programs, thus preserving their performance. However, the network-based approach requires a more sophisticated design and management strategy to reduce latency and enhance throughput. Despite these challenges, with proper implementation, the network-based approach substantially improves the efficiency and scalability of multiscale simulations by facilitating high-speed communication between diverse computational models.

Toward multiscale biomolecular simulations on exascale architectures

For multiscale simulations to run efficiently on state-of-the-art supercomputers, first and foremost, it is imperative that the programs dedicated to a specific methodology, such as QM and MM, are able to take advantage of the strengths of both CPUs and GPUs. Moreover, these programs must be capable of using the vast parallel processing capabilities of exascale architectures. Considerable efforts are underway to push QM- and MM-based programs towards exascale [60–67] with support from EuroHPC JU through its HPC Centres of Excellence [68] and the Exascale Computing Project (ECP) led by the US Department of Energy [69].

For a multiscale simulation framework to fully exploit the capabilities of these QM and MM programs, it must seamlessly integrate with their optimal parallelization strategies without introducing inefficiencies. This entails enabling the concurrent execution of interfaced programs, eliminating the need for their repeated startup and shutdown, and reducing communication overhead to a minimum. Crucially, the calculation of subsystem interactions, such as those between QM and MM particles, needs to be highly parallel and efficient to prevent it from becoming a computational bottleneck. Implementing effective and automatic load balancing is also critical to ensure that computational resources are utilized optimally, thereby maximizing the throughput of individual simulations.

Figure 1

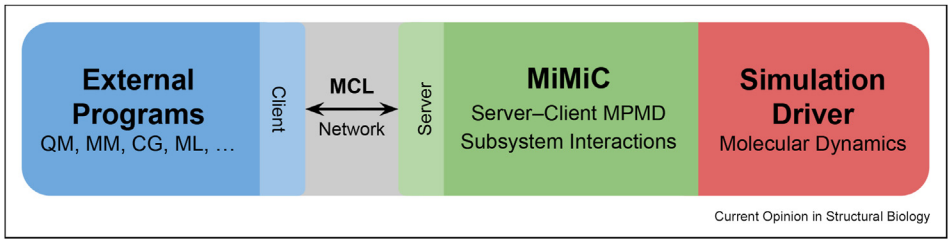


Illustration of the strategy used by the MiMiC framework.

The versatile and high-performance MiMiC framework for multiscale simulations was designed with the aim of addressing the criteria outlined above [28]. The strategy used by the MiMiC framework is illustrated in Figure 1. On one side, MiMiC connects to a simulation driver that manages the overall simulation process, including the integration of the equations of motion and the maintenance of temperature and pressure. On the other side, it interfaces multiple external programs using a client-server approach combined with a multiple-program multiple-data (MPMD) model. Each external program is tasked with calculations belonging to a given subsystem, while MiMiC calculates subsystem interactions. Importantly, the external programs run concurrently on separate computational resources using their own optimal parallelization capabilities. Communication between MiMiC and the external programs is facilitated by the MiMiC Communication Library (MCL), a dedicated lightweight library that ensures efficient network-based communication and simplifies the interfacing with external programs. An illustration of a MiMiC-based simulation workflow is shown in Figure 2. The scope of the MiMiC framework is broad, extending beyond QM/MM. This includes QM/QM, which integrates different levels of QM theory, QM/MM/CG, and even models that incorporate machine-learning (ML) techniques like ML/MM or QM/ML.

The MiMiC framework has been used to implement a DFT-based electrostatic-embedding QM/MM method, employing the CPMD program [70] as both the MD driver and QM engine, and GROMACS [61] as the MM engine [28,29]. In electrostatic embedding, the (external) electric field from the point charges in the MM subsystem is included in the Hamiltonian of the QM subsystem, thus directly polarizing its electronic density. The calculation of electrostatic QM/MM interactions is based on a dense grid representation of the electronic density, which is computationally expensive, especially for large systems, because the number of integrals over the electron density scales linearly with the number of atoms in the MM subsystem. The MiMiC framework has implemented an efficient approach that substantially speeds up the calculation essentially without compromising the accuracy of the forces, thus reducing the computational cost by about 80% for small systems (e.g., small solvated proteins) and up to 99% for larger systems (e.g., membrane-embedded proteins) [28]. Furthermore, MiMiC implements a hybrid shared- and distributed-memory (OpenMP/MPI) parallelization strategy to ensure that these calculations remain efficient and do not hinder performance, even under highly parallel conditions [29]. A powerful example of this is shown in Figure 3, where the extreme scalability of the CPMD program is harnessed to achieve strong scalability beyond 80,000 CPU cores with a parallel efficiency of 70% for a single MiMiC-based QM/MM MD simulation of the human isocitrate dehydrogenase-1 (IDH1) [13].

Table 1 shows examples of the computational performance and cost of MiMiC-based QM/MM MD simulations using CPMD and GROMACS. It is clear that the simulation throughput depends first and foremost on the chosen exchange-correlation functional and the size of the QM subsystem. This reflects the fact that the QM calculation is by far the most computationally demanding part of a QM/MM MD simulation. Hybrid exchange-correlation functionals are particularly expensive in a plane-wave basis set such as the one employed by the CPMD program. Still, due to the excellent parallelism in CPMD, the simulation throughput can be pushed to 4.8 ps/day for the small QM region (46 atoms) and 0.7 ps/day for the larger QM subsystem (142 atoms). Using instead a non-hybrid functional pushes the performance to 21 and 5.4 ps/day for the small and large QM subsystems, respectively.

Biological applications

The MiMiC framework has enabled a number of recent computational studies that demonstrate its utility and efficiency across a broad spectrum of systems with biological relevance. In this section, we showcase select examples that highlight the impactful contributions of MiMiC-based QM/MM MD simulations in advancing our understanding of complex biological processes.

Figure 2

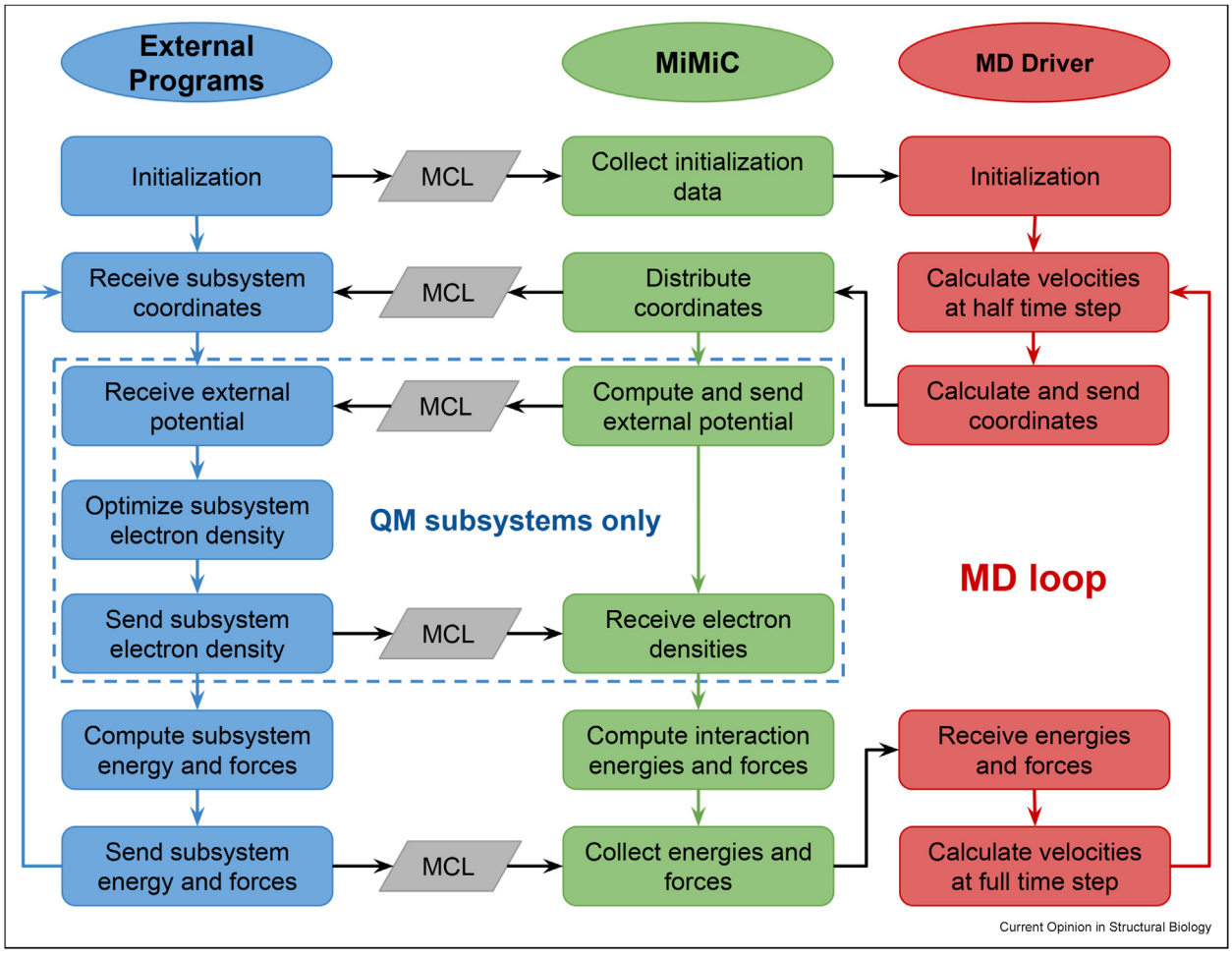
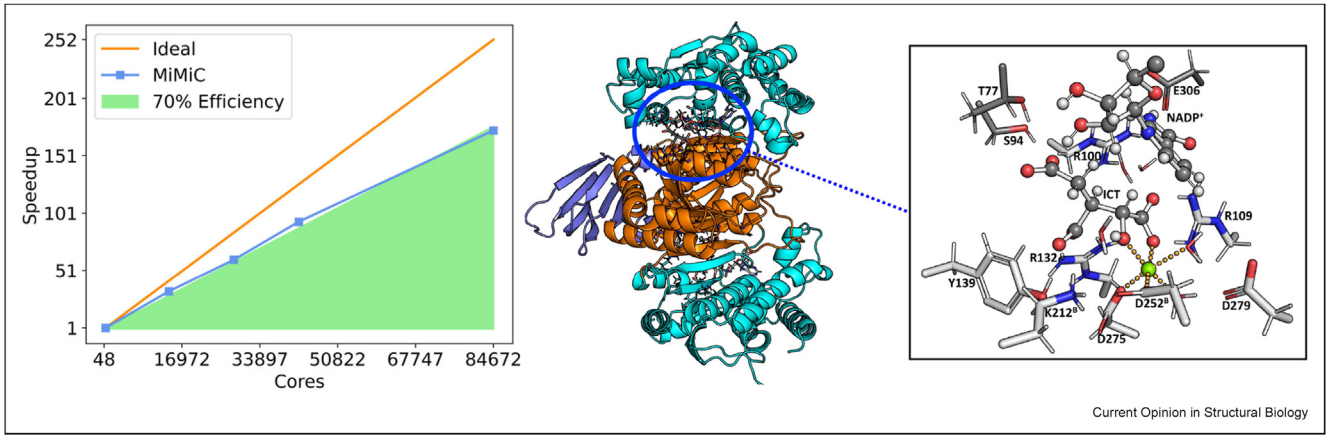


Illustration of MiMiC-based simulation workflow.

Figure 3



Parallel scalability and efficiency of MiMiC-based QM/MM MD simulations of IDH1. The IDH1 system has a total of 130,828 atoms with 142 QM atoms [13]. The simulations were run on the CPU partition of JUWELS [71]. The speedup is given in terms of the time per MD step normalized against a reference run using seven nodes. Adapted from Raghavan et al. [13], licensed under CC BY 4.0.

Table 1

Computational performance and cost of MiMiC-based QM/MM MD simulations. All simulations were run with a 0.5 fs timestep on the CPU partition of JUWELS [71] at the scaling limit (parallel efficiency ≥70%). The systems are p38α mitogen-activated protein kinase (169,550 atoms) and human isocitrate dehydrogenase-1 (130,828 atoms). We refer to the original work for full computational details [13].

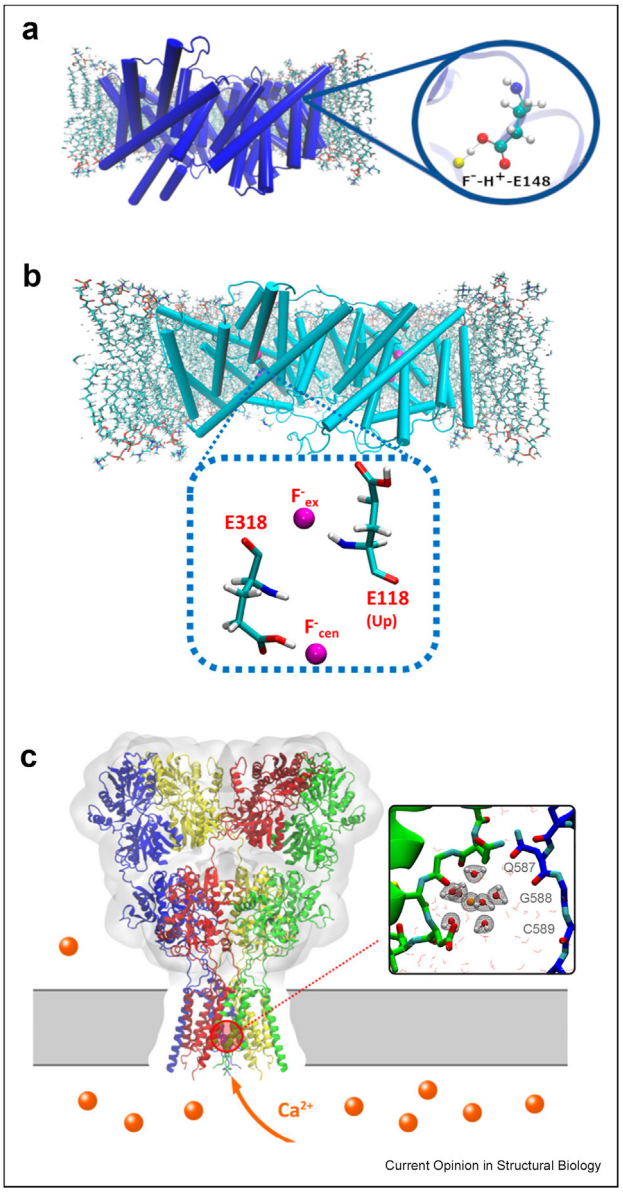
System	p38α		IDH1	
QM atoms	46		142	
XC functional	BLYP	B3LYP	BLYP	B3LYP
Throughput (ps/day)	21	4.8	5.4	0.7
Cost (node-hours/ps)	9	1280	480	60,480

Among the pioneering uses of MiMiC-based QM/MM MD simulations were studies of CLC proteins, a large family of anion channels and transporters. Chiariello et al. studied the molecular mechanism of fluoride inhibition of the anion/proton exchanger ClC-ec1 from *Escherichia coli* (Figure 4a) [72]. The use of QM/MM for this study was mandatory as ion translocation involves proton transfer processes. On the basis of QM/MM MD and well-tempered metadynamics (wtMTD) simulations at the B3LYP and BLYP levels of theory, Chiariello et al. were able to report proton affinities of the fluoride ion and the gating glutamate residue E148, thus providing valuable insights into transport inhibition. In a second study, the mechanisms of proton transfer and release by the fluoride/proton antiporter CLC^F-eca were investigated (Figure 4b) [73]. Employing the same simulation techniques, it could be shown that a triad is formed between fluoride, glutamate E318, and the gating glutamate E118, eventually releasing protons and fluoride as hydrogen fluoride.

In a recent study, the ligand iperoxo (routinely used in neuroimaging) targeting the human muscarinic acetylcholine receptor 2 was investigated [75]. The work focuses on the calculation of the drug unbinding rate constant k_{off}, a very difficult parameter to correctly estimate with force-field-based approaches. In fact, while methods based on modern force fields are nowadays able to predict accurate binding free energies and affinities, this is often not the case for rate constants such as k_{off}, which also require a correct estimation of the free energy of transition states. The study shows how sensitive this estimation is to values of the partial charges of the ligand and that while results obtained with a QM/MM MD simulation are in good agreement with experimental findings, standard force-field-based procedures lead to qualitatively wrong results, most likely due to the lack of explicit electronic polarization and charge transfer, which are included in QM/MM.

Another important application of QM/MM MD simulations is in scenarios where parts of a system cannot be adequately described using simplified analytical force fields. Notable examples are metal ions, in particular transition metals or divalent alkaline-earth ions. The absence of accurate parameters for the latter hinders a thorough understanding of many biological processes

Figure 4



Illustrations of biological systems that have been studied using the MiMiC framework. **a**: ClC-ec1 anion/proton antiporter embedded in a solvated lipid bilayer with a total of 150,925 atoms (19 QM atoms) [72]. Reprinted with permission from Ref. [72]. Copyright 2020 American Chemical Society. **b**: CLC^F-eca fluoride/proton antiporter embedded in a solvated lipid bilayer with a total of around 174,000 atoms (36 QM atoms) [73]. Reprinted with permission from Ref. [73]. Copyright 2021 American Chemical Society. **c**: AMPAR cation channel embedded in a solvated lipid bilayer where the transmembrane domain was included in the simulations (22 QM atoms) [74]. Figure by Schackert et al. [74], licensed under CC BY 4.0.

related to, e.g., ion channels, transporters, and pumps. Recently, Schackert et al. studied the mechanism of calcium permeation in α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs), which are key to rapid synaptic transmission in the central nervous system (Figure 4c) [74]. In that study, the calcium-binding sites within the channel, initially identified through classical MD, were further confirmed using QM/MM MD. This step was crucial for validating a newly developed classical force field specifically designed for calcium ions. Interestingly, they observed charge transfer between the calcium ion and the water molecules in the first solvation shell, which underlines the importance of a QM/MM description for such systems.

Outlook

The exascale era is poised to radically change life sciences. With the ability to conduct longer and more accurate simulations on larger and more complex molecular systems than ever before, we will be able to tackle scientific questions that are currently beyond our reach, deepening our understanding of biological processes. This opens up new opportunities for discovery and innovation, potentially revolutionizing fields such as drug design and bioengineering. However, fully realizing the potential of exascale computing is contingent upon overcoming substantial challenges in software design, algorithm optimization, and the efficient utilization of modular and heterogeneous HPC architectures.

The versatility of the MiMiC framework makes it ideal for pushing multiscale biomolecular simulations towards the exascale. It is capable of exploiting the results of the large-scale initiatives by EuroHPC JU and ECP for individual domains such as QM and MM. Indeed, work is already underway to couple a diverse set of QM programs to MiMiC, namely, Quantum ESPRESSO, CP2K, and DFT-FE, all of which are being developed to exploit state-of-the-art HPC technology [67]. Additionally, it is likely that MD simulations will benefit greatly from future ML-based force fields [76,77], which MiMiC is fully able to integrate into advanced simulation workflows [78]. For instance, hybrid ML/MM models facilitate long simulations of biological systems using ML-based force fields, achieving near QM/MM precision at a substantially reduced computational cost [79], while ML-enhanced free energy methods significantly accelerate QM/MM calculations of ligand binding affinities for drug discovery [80]. These and other novel methods are expected to aid in accurately predicting key biophysical properties like drug–protein binding free energies and complete free energy profiles. We anticipate that the integration of MiMiC-based multiscale simulations with ML and enhanced sampling techniques will further catalyze a *quantum leap* in the fundamental atomistic understanding of biological processes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

No data was used for the research described in the article.

Acknowledgments

JMHO thanks VILLUM FONDEN (Grant No. VIL29478) for financial support. UR gratefully acknowledges funding from the Swiss National Science Foundation via the NCCR MUST and the individual grants 200020-185092 and 200020-219440, as well as computing time from the Swiss National Computing Centre CSCS. EI thanks the Helmholtz European Partnering program ("Innovative high-performance computing approaches for molecular neuromedicine") for funding.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Schaffer LV, Ideker T: **Mapping the multiscale structure of biological systems**. *Cell Syst* 2021, **12**:622–635, https://doi.org/10.1016/j.cels.2021.05.012.

2. Palermo G, Bonvin AMJJ, Dal Peraro M, Amaro RE, Tozzini V: **Multiscale modeling from macromolecules to cell: opportunities and challenges of biomolecular simulations**. *Front Mol Biosci* 2020, **7**, https://doi.org/10.3389/fmolb.2020.00194.

3. Trovato F, Trylska J, Bond PJ, Wolynes PG: **Combining simulations, theory, and experiments into multiscale models of biological events**. *Front Mol Biosci* 2021, **8**, https://doi.org/10.3389/fmolb.2021.797754.

4. Hammes-Schiffer S: **Proton-coupled electron transfer: moving together and charging forward**. *J Am Chem Soc* 2015, **137**:8860–8871, https://doi.org/10.1021/jacs.5b04087.

5. Nottoli M, Cupellini L, Lipparini F, Granucci G, Mennucci B: **Multiscale models for light-driven processes**. *Annu Rev Phys Chem* 2021, **72**:489–513, https://doi.org/10.1146/annurev-physchem-090419-104031.

6. Toldo JM, do Casal MT, Ventura E, do Monte SA, Barbatti M: **Surface hopping modeling of charge and energy transfer in active environments**. *Phys Chem Chem Phys* 2023, **25**:8293–8316, https://doi.org/10.1039/d3cp00247k.

7. Sousa SF, Ribeiro AJM, Neves RPP, Brás NF, Cerqueira NMFSA, Fernandes PA, Ramos MJ: **Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms**. *Wiley Interdiscip Rev Comput Mol Sci* 2017, **7**, e1281, https://doi.org/10.1002/wcms.1281.

8. * Vennelakanti V, Nazemi A, Mehmood R, Steeves AH, Kulik HJ: **Harder, better, faster, stronger: large-scale QM and QM/MM for predictive modeling in enzymes and proteins**. *Curr Opin Struct Biol* 2022, **72**:9–17, https://doi.org/10.1016/j.sbi.2021.07.004.
Informative recent article summarizing the current state-of-the-art in QM and QM/MM modeling of enzymes and proteins.

9. Tzeliou CE, Mermigki MA, Tzeli D: **Review on the QM/MM methodologies and their application to metalloproteins**. *Molecules* 2022, **27**:2660, https://doi.org/10.3390/molecules27092660.

10. van der Kamp MW, Begum J: *QM/MM for structure-based drug design: techniques and applications*. John Wiley & Sons, Ltd; 2024:119–156, https://doi.org/10.1002/9783527840748.ch6.

11. Brunk E, Rothlisberger U: **Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states**. *Chem Rev* 2015, **115**:6217–6263, https://doi.org/10.1021/cr500628b.

12. Clemente CM, Capece L, Martí MA: **Best practices on QM/MM simulations of biological systems**. *J Chem Inf Model* 2023, **63**:2609–2627, https://doi.org/10.1021/acs.jcim.2c01522.

13. * Raghavan B, Paulikat M, Ahmad K, Callea L, Rizzi A, Ippoliti E, Mandelli D, Bonati L, De Vivo M, Carloni P: **Drug design in the Exascale Era: a perspective from massively parallel QM/MM simulations**. *J Chem Inf Model* 2023, **63**:3647–3658, https://doi.org/10.1021/acs.jcim.3c00557.
This article explores the prospects of exascale computing for QM/MM-based computationally-aided drug design.

14. Remmel A. *What exascale computing could mean for chemistry*, vol. 100. C&EN Global Enterprise; 2022. URL: https://cen.acs.org/physical-chemistry/computational-chemistry/exascale-computing-mean-chemistry/100/i31.

15. Choi CQ: **The beating heart of the world's first exascale supercomputer**. *IEEE Spectrum* 2022. URL: https://spectrum.ieee.org/frontier-exascale-supercomputer; 2022.

16. Di Felice R, Mayes ML, Richard RM, Williams-Young DB, Chan GK-L, de Jong WA, Govind N, Head-Gordon M, Hermes MR, Kowalski K, Li X, Lischka H, Mueller KT, Mutlu E, Niklasson AMN, Pederson MR, Peng B, Shepard R, Valeev EF, van Schilfgaarde M, Vlaisavljevich B, Windus TL, Xantheas SS, Zhang X, Zimmerman PM: **A perspective on sustainable computational chemistry software development and integration**. *J Chem Theory Comput* 2023, **19**:7056–7076, https://doi.org/10.1021/acs.jctc.3c00419.

17. Carpenter P, Utz U-H, Narasimhamurthy S, Suarez E: ***Heterogeneous high performance computing***. https://doi.org/10.5281/zenodo.6090425.

18. Suarez E, Eicker N, Moschny T, Pickartz S, Clauss C, Plugaru V, Herten A, Michielsen K, Lippert T: ***Modular supercomputing architecture***. https://doi.org/10.5281/zenodo.6508394.

19. ***LUMI***. https://www.lumi-supercomputer.eu/ [Date accessed: 2024-02-11].

20. ***JUPITER – exascale for Europe***. https://www.fz-juelich.de/en/ias/jsc/jupiter [Date accessed: 2024-02-11].

21. ***Czechia will host the European LUMI-Q quantum computer***. https://www.it4i.cz/en/about/infoservice/press-releases/czechia-will-host-the-european-lumi-q-quantum-computer [Date accessed: 2024-02-11].

22. ***JUNIQ – Jülich UNified infrastructure for quantum computing***. https://www.fz-juelich.de/en/ias/jsc/systems/quantum-computing/juniq-facility/juniq [Date accessed: 2024-02-20].

23. Jin J, Pak AJ, Durumeric AEP, Loose TD, Voth GA: **Bottom-up coarse-graining: principles and perspectives**. *J Chem Theory Comput* 2022, **18**:5759–5791, https://doi.org/10.1021/acs.jctc.2c00643.

24. Guo Y, Mofrad MRK, Tepole AB: **On modeling the multiscale mechanobiology of soft tissues: challenges and progress**. *Biophys Rev* 2022, **3**, https://doi.org/10.1063/5.0085025.

25. Bock LV, Gabrielli S, Kolář MH, Grubmüller H: **Simulation of complex biomolecular systems: the ribosome challenge**. *Annu Rev Biophys* 2023, **52**:361–390, https://doi.org/10.1146/annurev-biophys-111622-091147.

26. Borges-Araújo L, Patmanidis I, Singh AP, Santos LHS, Sieradzan AK, Vanni S, Czaplewski C, Pantano S, Shinoda W, Monticelli L, Liwo A, Marrink SJ, Souza PCT: **Pragmatic coarse-graining of proteins: models and applications**. *J Chem Theory Comput* 2023, **19**:7112–7135, https://doi.org/10.1021/acs.jctc.3c00733.

27. Fadda E, Makshakova O, Perez S: **Understanding glycobiology through multiscale molecular dynamics simulations: from basic principles to case studies**. *Elsevier* 2024:379–396, https://doi.org/10.1016/b978-0-12-819655-7.00006-0.

28. Olsen JMH, Bolnykh V, Meloni S, Ippoliti E, Bircher MP, Carloni P, Rothlisberger U: **MiMiC: a novel framework for multiscale modeling in computational chemistry**. *J Chem Theory Comput* 2019, **15**:3810–3823, https://doi.org/10.1021/acs.jctc.9b00093.

29. Bolnykh V, Olsen JMH, Meloni S, Bircher MP, Ippoliti E, Carloni P, Rothlisberger U: **Extreme scalability of DFT-based QM/MM MD simulations using MiMiC**. *J Chem Theory Comput* 2019, **15**:5601–5613, https://doi.org/10.1021/acs.jctc.9b00424.

30. Laio A, VandeVondele J, Rothlisberger U: **A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations**. *J Chem Phys* 2002, **116**:6941–6947, https://doi.org/10.1063/1.1462041.

31. Vreven T, Morokuma K, Farkas O, Schlegel HB, Frisch MJ: **Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints**. *J Comput Chem* 2003, **24**:760–769, https://doi.org/10.1002/jcc.10156.

32. Laino T, Mohamed F, Laio A, Parrinello M: **An efficient real space multigrid QM/MM electrostatic coupling**. *J Chem Theory Comput* 2005, **1**:1176–1184, https://doi.org/10.1021/ct050123f.

33. Brooks BR, Brooks III CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M: **CHARMM: the biomolecular simulation program**. *J Comput Chem* 2009, **30**:1545–1614, https://doi.org/10.1002/jcc.21287.

34. Salomon-Ferrer R, Case DA, Walker RC: **An overview of the amber biomolecular simulation package**. *Wiley Interdiscip Rev Comput Mol Sci* 2013, **3**:198–210, https://doi.org/10.1002/wcms.1121.

35. de la Lande A, Alvarez-Ibarra A, Hasnaoui K, Cailliez F, Wu X, Mineva T, Cuny J, Calaminici P, López-Sosa L, Geudtner G, Navizet I, Garcia Iriepa C, Salahub DR, Köster AM: **Molecular simulations with in-deMon2k QM/MM, a tutorial-review**. *Molecules* 2019, **24**, https://doi.org/10.3390/molecules24091653.

36. Neese F, Wennmohs F, Becker U, Riplinger C: **The ORCA quantum chemistry program package**. *J Chem Phys* 2020, **152**, 224108, https://doi.org/10.1063/5.0004608.

37. Seritan S, Bannwarth C, Fales BS, Hohenstein EG, Isborn CM, Kokkila-Schumacher SIL, Li X, Liu F, Luehr N, Snyder Jr JW, Song C, Titov AV, Ufimtsev IS, Wang L-P, Martínez TJ: **TeraChem: a graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics**. *Wiley Interdiscip Rev Comput Mol Sci* 2021, **11**, e1494, https://doi.org/10.1002/wcms.1494.

38. Isborn CM, Götz AW, Clark MA, Walker RC, Martínez TJ: **Electronic absorption spectra from MM and ab initio QM/MM molecular dynamics: environmental effects on the absorption spectrum of photoactive yellow protein**. *J Chem Theory Comput* 2012, **8**:5092–5106, https://doi.org/10.1021/ct3006826.

39. Dziedzic J, Mao Y, Shao Y, Ponder J, Head-Gordon T, Head-Gordon M, Skylaris C-K: **TINKTEP: a fully self-consistent, mutually polarizable QM/MM approach based on the AMOEBA force field**. *J Chem Phys* 2016, **145**, 124106, https://doi.org/10.1063/1.4962909.

40. Loco D, Lagardère L, Caprasecca S, Lipparini F, Mennucci B, Piquemal J-P: **Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding**. *J Chem Theory Comput* 2017, **13**:4025–4033, https://doi.org/10.1021/acs.jctc.7b00572.

41. Cruzeiro VWD, Manathunga M, Merz KMJ, Götz AW: **Open-source multi-GPU-accelerated QM/MM simulations with AMBER and QUICK**. *J Chem Inf Model* 2021, **61**:2109–2115, https://doi.org/10.1021/acs.jcim.1c00169.

42. Götz AW, Clark MA, Walker RC: **An extensible interface for QM/MM molecular dynamics simulations with AMBER**. *J Comput Chem* 2014, **35**:95–108, https://doi.org/10.1002/jcc.23444.

43. Melo MCR, Bernardi RC, Rudack T, Scheurer M, Riplinger C, Phillips JC, Maia JDC, Rocha GB, Ribeiro JV, Stone JE, Neese F, Schulten K, Luthey-Schulten Z: **NAMD goes quantum: an**

integrative suite for hybrid simulations. *Nat Methods* 2018, **15**: 351–354, https://doi.org/10.1038/nmeth.4638.

44. Cruzeiro VWD, Wang Y, Pieri E, Hohenstein EG, Martínez TJ: **TeraChem protocol buffers (TCPB): accelerating QM and QM/MM simulations with a client–server model**. *J Chem Phys* 2023, **158**, 044801, https://doi.org/10.1063/5.0130886.

45. Woodcock HL, Miller BT, Hodoscek M, Okur A, Larkin JD, Ponder JW, Brooks BR: **MSCALE: a general utility for multiscale modeling**. *J Chem Theory Comput* 2011, **7**:1208–1219, https://doi.org/10.1021/ct100738h.

46. Ma C, Martin-Samos L, Fabris S, Laio A, Piccinin S: **QMMMW: a wrapper for QM/MM simulations with Quantum ESPRESSO and LAMMPS**. *Comp Phys Commun* 2015, **195**:191–198, https://doi.org/10.1016/j.cpc.2015.04.024.

47. Torras J, Deumens E, Trickey SB: **Software integration in multiscale simulations: the PUPIL system**. *J Comput Aided Mat Des* 2006, **13**:201–212, https://doi.org/10.1007/s10820-006-9011-3.

48. Torras J, Roberts BP, Seabra GM, Trickey SB: **Chapter one – PUPIL: a software integration system for multi-scale QM/MM-MD simulations and its application to biomolecular systems**. In *Combined quantum mechanical and molecular mechanical modelling of biomolecular interactions*. Edited by Karabencheva-Christova T, *Adv. Protein Chem. Struct. Biol.*, vol. 100; 2015:1–31, https://doi.org/10.1016/bs.apcsb.2015.06.002.

49. Řezáč J: **Cuby: an integrative framework for computational chemistry**. *J Comput Chem* 2016, **37**:1230–1237, https://doi.org/10.1002/jcc.24312.

50. Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Dułak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg EL, Kubal J, Kaasbjerg K, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW: **The atomic simulation environment—a Python library for working with atoms**. *J Phys Condens Matter* 2017, **29**, 273002, https://doi.org/10.1088/1361-648X/aa680e.

51. Weingart O, Nenov A, Altoè P, Rivalta I, Segarra-Martí J, Dokukina I, Garavelli M: **COBRAMM 2.0 — a software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations**. *J Mol Model* 2018, **24**, https://doi.org/10.1007/s00894-018-3769-6.

52. Metz S, Kästner J, Sokol AA, Keal TW, Sherwood P: **ChemShell—a modular software package for QM/MM simulations**. *Wiley Interdiscip Rev Comput Mol Sci* 2014, **4**:101–110, https://doi.org/10.1002/wcms.1163.

53. Lu Y, Farrow MR, Fayon P, Logsdail AJ, Sokol AA, Catlow CRA, Sherwood P, Keal TW: **Open-source, python-based redevelopment of the ChemShell multiscale QM/MM environment**. *J Chem Theory Comput* 2019, **15**:1317–1328, https://doi.org/10.1021/acs.jctc.8b01036.

54. Kratz EG, Walker AR, Lagardère L, Lipparini F, Piquemal J-P, Andrés Cisneros G: **LICHEM: a QM/MM program for simulations with multipolar and polarizable force fields**. *J Comput Chem* 2016, **37**:1019–1029, https://doi.org/10.1002/jcc.24295.

55. Gökcan H, Vázquez-Montelongo EA, Cisneros GA: **LICHEM 1.1: recent improvements and new capabilities**. *J Chem Theory Comput* 2019, **15**:3056–3065, https://doi.org/10.1021/acs.jctc.9b00028.

56. Zhang B, Altarawy D, Barnes T, Turney JM, Schaefer HFI: **Janus: an extensible open-source software package for adaptive QM/MM methods**. *J Chem Theory Comput* 2019, **15**:4362–4373, https://doi.org/10.1021/acs.jctc.9b00182.

57. Barnes TA, Marin-Rimoldi E, Ellis S, Crawford TD: **The MolSSI Driver Interface Project: a framework for standardized, on-the-fly interoperability between computational molecular sciences codes**. *Comp Phys Commun* 2021, **261**, 107688, https://doi.org/10.1016/j.cpc.2020.107688.

58. Martí S: **QMCube (QM3): an all-purpose suite for multiscale QM/MM calculations**. *J Comput Chem* 2021, **42**:447–457, https://doi.org/10.1002/jcc.26465.

59. Lin H, Zhang Y, Pezeshki S, Duster AW, Wang B, Wu X-P, Zheng S-W, Gagliardi L, Truhlar DG: **QMMM 2023: a program for combined quantum mechanical and molecular mechanical modeling and simulations**. *Comp Phys Commun* 2024, **295**, 108987, https://doi.org/10.1016/j.cpc.2023.108987.

60. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, Buch R, Fiorin G, Hénin J, Jiang W, McGreevy R, Melo MCR, Radak BK, Skeel RD, Singharoy A, Wang Y, Roux B, Aksimentiev A, Luthey-Schulten Z, Kalé LV, Schulten K, Chipot C, Tajkhorshid E: **Scalable molecular dynamics on CPU and GPU architectures with NAMD**. *J Chem Phys* 2020, **153**, https://doi.org/10.1063/5.0014475.

61. Páll S, Zhmurov A, Bauer P, Abraham M, Lundborg M, Gray A, Hess B, Lindahl E: **Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS**. *J Chem Phys* 2020, **153**, https://doi.org/10.1063/5.0018516.

62. Kowalski K, Bair R, Bauman NP, Boschen JS, Bylaska EJ, Daily J, de Jong WA, Dunning T, Govind N, Harrison RJ, Keçeli M, Keipert K, Krishnamoorthy S, Kumar S, Mutlu E, Palmer B, Panyala A, Peng B, Richard RM, Straatsma TP, Sushko P, Valeev EF, Valiev M, van Dam HJJ, Waldrop JM, Williams-Young DB, Yang C, Zalewski M, Windus TL: **From NWChem to NWChemEx: evolving with the computational chemistry landscape**. *Chem Rev* 2021, **121**: 4962–4998, https://doi.org/10.1021/acs.chemrev.0c00998.

63. Das S, Motamarri P, Subramanian V, Rogers DM, Gavini V: **DFT-FE 1.0: a massively parallel hybrid CPU-GPU density functional theory code using finite-element discretization**. *Comp Phys Commun* 2022, **280**, 108473, https://doi.org/10.1016/j.cpc.2022.108473.

64. ** Schade R, Kenter T, Elgabarty H, Lass M, Kühne TD, Plessl C: **Breaking the exascale barrier for the electronic structure problem in ab-initio molecular dynamics**. *Int J High Perform Comput Appl* 2023, **37**:530–538, https://doi.org/10.1177/10943420231177631.
An impressive demonstration of the current state-of-art of GPU-accelerated tight-binding-based molecular dynamics simulations with an application to the SARS-CoV-2 spike protein anchored in a lipid bilayer in aqueous solution, a system containing 83 million atoms.

65. Carnimeo I, Affinito F, Baroni S, Baseggio O, Bellentani L, Bertossa R, Delugas PD, Ruffino FF, Orlandini S, Spiga F, Giannozzi P: **Quantum ESPRESSO: one further step toward the exascale**. *J Chem Theory Comput* 2023, **19**:6992–7006, https://doi.org/10.1021/acs.jctc.3c00249.

66. Zahariev F, Xu P, Westheimer BM, Webb S, Galvez Vallejo J, Tiwari A, Sundriyal V, Sosonkina M, Shen J, Schoendorff G, Schlinsog M, Sattasathuchana T, Ruedenberg K, Roskop LB, Rendell AP, Poole D, Piecuch P, Pham BQ, Mironov V, Mato J, Leonard S, Leang SS, Ivanic J, Hayes J, Harville T, Gururangan K, Guidez E, Gerasimov IS, Friedl C, Ferreras KN, Elliott G, Datta D, Cruz DDA, Carrington L, Bertoni C, Barca GMJ, Alkan M, Gordon MS: **The general atomic and molecular electronic structure system (GAMESS): novel methods on novel architectures**. *J Chem Theory Comput* 2023, **19**: 7031–7055, https://doi.org/10.1021/acs.jctc.3c00379.

67. * Gavini V, Baroni S, Blum V, Bowler DR, Buccheri A, Chelikowsky JR, Das S, Dawson W, Delugas P, Dogan M, Draxl C, Galli G, Genovese L, Giannozzi P, Giantomassi M, Gonze X, Govoni M, Gygi F, Gulans A, Herbert JM, Kokott S, Kühne TD, Liou K-H, Miyazaki T, Motamarri P, Nakata A, Pask JE, Plessl C, Ratcliff LE, Richard RM, Rossi M, Schade R, Scheffler M, Schütt O, Suryanarayana P, Torrent M, Truflandier L, Windus TL, Xu Q, Yu VW-Z, Perez D: **Roadmap on electronic structure codes in the exascale era**. *Model Simul Mat Sci Eng* 2023, **31**, 063301, https://doi.org/10.1088/1361-651x/acdf06.
Instructive overview of the current state-of-the-art of 14 popular electronic structure codes and the challenges and opportunities that the community faces when entering the exascale computing era.

68. ***Kick-off of 10 centres of excellence in HPC to support the transition towards exascale***. https://eurohpc-ju.europa.eu/kick-10-centres-excellence-hpc-support-transition-towards-exascale-2023-01-26_en [Date accessed: 2024-02-11].

69. ***Exascale computing Project***. https://www.exascaleproject.org/ [Date accessed: 2024-02-11].

70. CPMD: **Copyright 1990-2023 by IBM Corp. and copyright 1994-2001 by Max Planck Institute**. *Stuttgart* 2023. http://www.cpmd.org/; 2023 [Date accessed: 2024-02-11].

71. ***JUWELS: Jülich Wizard for European leadership science***. https://www.fz-juelich.de/en/ias/jsc/systems/supercomputers/juwels [Date accessed: 2024-02-11].

72. Chiariello MG, Bolnykh V, Ippoliti E, Meloni S, Olsen JMH, Beck T, Rothlisberger U, Fahlke C, Carloni P: **Molecular basis of CLC antiporter inhibition by fluoride**. *J Am Chem Soc* 2020, **142**:7254–7258, https://doi.org/10.1021/jacs.9b13588.

73. Chiariello MG, Alfonso-Prieto M, Ippoliti E, Fahlke C, Carloni P: **Mechanisms underlying proton release in CLC-type F–/H+ antiporters**. *J Phys Chem Lett* 2021, **12**:4415–4420, https://doi.org/10.1021/acs.jpclett.1c00361.

74. ** Schackert FK, Biedermann J, Abdolvand S, Minniberger S, Song C, Plested AJR, Carloni P, Sun H: **Mechanism of calcium permeation in a glutamate receptor ion channel**. *J Chem Inf Model* 2023, **63**:1293–1300, https://doi.org/10.1021/acs.jcim.2c01494.

Combined experimental and computational study using X-ray crystallography, multisite classical MD, and multiscale QM/MM simulations to reveal the mechanism of Ca^{2+} permeation in AMPAR neurotransmitter-activated cation channels.

75. Capelli R, Lyu W, Bolnykh V, Meloni S, Olsen JMH, Rothlisberger U, Parrinello M, Carloni P: **Accuracy of molecular simulation-based predictions of k_{off} values: a metadynamics study**. *J Phys Chem Lett* 2020, **11**:6373–6381, https://doi.org/10.1021/acs.jpclett.0c00999.

76. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Schütt KT, Tkatchenko A, Müller K-R: **Machine learning force fields**. *Chem Rev* 2021, **121**:10142–10186, https://doi.org/10.1021/acs.chemrev.0c01111.

77. Kocer E, Ko TW, Behler J: **Neural network potentials: a concise overview of methods**. *Annu Rev Phys Chem* 2022, **73**:163–186, https://doi.org/10.1146/annurev-physchem-082720-034254.

78. * Mouvet F, Villard J, Bolnykh V, Rothlisberger U: **Recent advances in first-principles based molecular dynamics**. *Acc Chem Res* 2022, **55**:221–230, https://doi.org/10.1021/acs.accounts.1c00503.

Reports on recent progress in accelerating first-principles-based QM/MM MD simulations up to several orders of magnitude through the use of multiple time step and machine learning techniques.

79. * Galvelis R, Varela-Rial A, Doerr S, Fino R, Eastman P, Markland TE, Chodera JD, De Fabritiis G: **NNP/MM: accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics**. *J Chem Inf Model* 2023, **63**:5701–5708, https://doi.org/10.1021/acs.jcim.3c00773.

Application of a combined neural network potential/molecular mechanics (NNP/MM) approach to different ligand-protein systems with substantial computational speedups.

80. ** Rizzi A, Carloni P, Parrinello M: **Free energies at QM accuracy from force fields via multimap targeted estimation**. *Proc Natl Acad Sci* 2023, **120**, e2304308120, https://doi.org/10.1073/pnas.2304308120.

Astute multimap targeted MD approach applied to the calculation of ligand binding affinities with QM accuracy, based on force field reference potentials augmented with the help of normalizing flow neural networks.