



MiMiC: Multiscale Modeling in Computational Chemistry

Viacheslav Bolnykh^{1*}, Jógvan Magnus Haugaard Olsen^{2*}, Simone Meloni³, Martin P. Bircher⁴, Emiliano Ippoliti⁵, Paolo Carloni^{5,6} and Ursula Rothlisberger^{1*}

¹ Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, ² Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT the Arctic University of Norway, Tromsø, Norway, ³ Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Ferrara, Ferrara, Italy, ⁴ Computational and Soft Matter Physics, University of Vienna, Vienna, Austria, ⁵ Computational Biomedicine, Institute for Advanced Simulation (IAS-5) and Institute of Neuroscience and Medicine (INM-9), Molecular Neuroscience and Neuroimaging, Institute of Neuroscience and Medicine (JARA INM-11), Forschungszentrum Jülich, Jülich, Germany, ⁶ Department of Physics and Universitätsklinikum Aachen, RWTH Aachen University, Aachen, Germany

Keywords: molecular dynamics, QM/MM, DFT, HPC, multiscale simulations, computational chemistry

1. INTRODUCTION

Hybrid quantum mechanics/molecular mechanics (QM/MM) approaches are commonly used methods for investigating a plethora of chemical, biochemical, and biophysical processes that require explicit treatment of the electronic degrees of freedom when the system is too big to be entirely treated by QM methods alone (Warshel and Levitt, 1976; Senn and Thiel, 2009; Adhikari et al., 2014; Campomanes et al., 2014, 2015; Brunk and Rothlisberger, 2015; Genna et al., 2016; Li et al., 2017; Cupellini et al., 2018; Loco et al., 2018; Morzan et al., 2018). It is often the method of choice for computational investigations of systems with more than a few thousand atoms (which is commonly the case for biological systems). In QM/MM, the system is split into two parts: a smaller part that is treated at the QM level of theory, whereas the remainder is described at the MM level, which is a computationally more expedient description. In this way, local electronic effects can be captured with the accuracy of a first-principles method, while at the same time explicitly including the effects of the environment at a reasonable computational cost. Current QM/MM implementations have roughly followed either of two strategies: (1) tight integration of QM and MM modules in a single software package or (2) loose coupling of separate QM and MM codes. Strategy (1) generally profits from computational efficiency due to the ability to pass data between the submodules directly (via function calls) but suffers from limited flexibility, since the available choice of methods is often restricted and extensions to different programs may require formidable programming efforts. In contrast, strategy (2), which is typically implemented resorting to data exchange between QM and MM codes *via* file input and output, enables high flexibility but penalizes efficiency because of increased communication overhead. However, with the field rapidly growing, new simulation paradigms and approaches might quickly emerge, clearly favoring strategy (2) over (1). In the following, we show that flexibility does not necessarily come at the expense of a high computation (or communication) overhead by presenting the recently developed MiMiC framework (Bolnykh et al., 2019; Olsen et al., 2019) that combines the capability of performing fast and efficient multiscale molecular dynamics (MD) simulations with facile support for flexible extensions. These objectives are achieved by applying (2) with an efficient method to exchange data among the coupled software packages. In practice, MiMiC implements a multiple program-multiple data (MPMD) paradigm through a message passing interface (MPI)-based communication library, which allows the entities collaborating within MiMiC to exchange data efficiently. Overall, MiMiC represents a highly modular and general multiscale simulation framework that enables the combination of multiple resolutions and methods for different parts of a system, while retaining high computational efficiency. Moreover, MiMiC was designed to have a flexible architecture enabling multiple resolutions, implementation of different

OPEN ACCESS

Edited by:

Giulia Palermo,
University of California, Riverside,
United States

Reviewed by:

Pablo Ricardo Arantes,
University of California, Riverside,
United States
Carme Rovira,
University of Barcelona, Spain

*Correspondence:

Viacheslav Bolnykh
viacheslav.bolnykh@epfl.ch
Jógvan Magnus Haugaard Olsen
jogvan.m.olsen@uit.no
Ursula Rothlisberger
ursula.rothlisberger@epfl.ch

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 23 December 2019

Accepted: 02 March 2020

Published: 20 March 2020

Citation:

Bolnykh V, Olsen JMH, Meloni S,
Bircher MP, Ippoliti E, Carloni P and
Rothlisberger U (2020) MiMiC:
Multiscale Modeling in Computational
Chemistry. *Front. Mol. Biosci.* 7:45.
doi: 10.3389/fmolb.2020.00045

types of coupling (e.g., QM/QM, QM/QM/MM, etc.), and to straightforwardly incorporate emerging—and future—methods and software packages in the field of computational chemistry. This flexibility is of utmost importance in the light of the rapid development of computational methods enabling researchers to tackle complex scientific problems with more and more degrees of freedom that require the incorporation of multiple space and time resolution scales on the one hand, and the rapid advent of new computational approaches on the other hand.

2. MIMIC ARCHITECTURE

2.1. Model

MiMiC implements a generalized version of the fully Hamiltonian electrostatic embedding scheme introduced in Laio et al. (2002). The key quantity is the electrostatic QM/MM coupling energy term:

$$E^{\text{QM/MM}} = \sum_i^{N^{\text{MM}}} q_i^{\text{MM}} \int d\mathbf{r} \rho^{\text{QM}}(\mathbf{r}) \frac{r_{c,i}^4 - |\mathbf{R}_i - \mathbf{r}|^4}{r_{c,i}^5 - |\mathbf{R}_i - \mathbf{r}|^5} \quad (1)$$

where N^{MM} is the total number of MM atoms, q_i^{MM} and $r_{c,i}$ are the partial charge and the covalent radius of the i -th MM atom, respectively, while $\mathbf{R}_i$ is its coordinate and $\rho^{\text{QM}}(\mathbf{r})$ is the electron density in point $\mathbf{r}$. This form of the electrostatic QM/MM coupling term modifies the Coulomb interaction at short range, thus avoiding electron spill-out (Laio et al., 2002). It is worth remarking that the QM/MM term is responsible for the polarization of the electronic density due to MM atoms and, thus, models the effects of the environment on the properties of the chemically active subdomain.

The straightforward implementation of such a term is rather costly to compute, in particular for systems with large MM regions. Therefore, a hierarchical electrostatic embedding approach (Laio et al., 2002) is used in order to mitigate the high computational cost of a direct evaluation. Within this hierarchical scheme the QM/MM electrostatic interactions are divided into two groups depending on the distance (commonly referred to as the cutoff distance) of MM atoms from the QM subsystem. In the vicinity of the QM part the interaction is computed using Equation (1), whereas more distant atoms are coupled via a multipole expansion of the electrostatic potential of the QM charge distribution. We have extended the original scheme with an open-ended multipole expansion allowing the user to choose the order at which the expansion is truncated. This allows (i) higher accuracy in the calculation of the electrostatic QM/MM interactions, at a negligibly higher computational cost and (ii) reduction of the cutoff distance, thus further lowering the computational cost (Olsen et al., 2019).

An official release of MiMiC will be published under the open-source GPLv3+ license in 2020.

2.2. Implementation

MiMiC is a loosely-coupled MPMD multiscale simulation framework. Within this approach, both QM and MM codes run simultaneously with computational resources being allocated

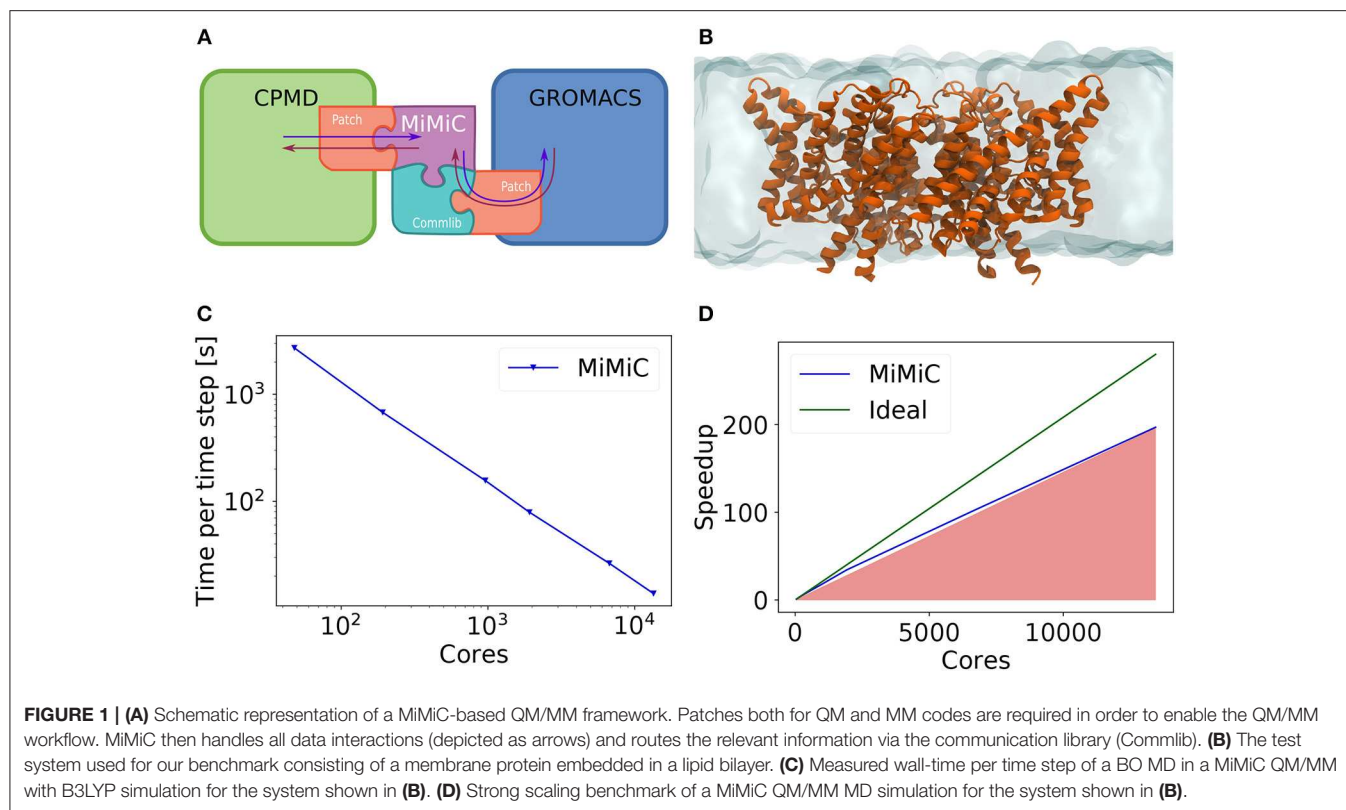
separately to either entity. Moreover, while enabling efficient communication, such an approach avoids tight integration of MiMiC into either code, which would incur a high implementation and maintenance effort. This enables the construction of a highly modular and efficient multiscale simulation framework capable of coupling virtually any set of simulation codes with the potential for extending it further to enable the support of alternative levels of theory such as a different QM method, coarse-grained approaches, or approaches based on artificial intelligence (Behler and Parrinello, 2007; Christensen et al., 2019; Singraber et al., 2019). In the present implementation, CPMD 4.3 (Hutter et al., 2018) computes the QM contributions, while GROMACS 2019 (Spoel et al., 2005; Abraham et al., 2015, 2019) computes the classical interactions within the MM subsystem as well as all bonded and Lennard-Jones interactions crossing the QM/MM interface. The electrostatic QM/MM interactions are computed by MiMiC. Finally, CPMD integrates the equations of motion.

The structure of a QM/MM implementation using the MiMiC framework is shown in **Figure 1A**. The use of a plane wave-based code to handle the QM subsystem ensures highly efficient scaling performance, while GROMACS guarantees expedient MM computations.

The workflow of a QM/MM MD simulation using MiMiC follows closely the workflow of a typical MD simulation in CPMD. At the beginning of each time step, MiMiC collects atomic coordinates from CPMD and dispatches them to GROMACS, which then computes MM forces and energies. While this is done, CPMD computes QM contributions and MiMiC computes the electrostatic QM/MM interaction terms. MiMiC adds up all force contributions and provides them to CPMD, which uses them to propagate atomic positions according to the selected ensemble and imposing the necessary constraints.

The calculation of the QM/MM interactions of Equation (1) can be parallelized by distributing MM atoms and points of the mesh discretizing the QM domain of integration. Extreme scalability is achieved parallelizing over both degrees of freedom through a multi-layered hybrid distributed- and shared-memory parallelization strategy. At the top layer, all MPI tasks are divided into groups, each receiving a subset of MM atoms. Then, at a lower level, the mesh discretizing the QM subspace is split into a set of 2D slabs along the X dimension. Each of the MPI tasks belonging to each group receives a subset of these slabs to compute the corresponding part of the integral in Equation (1) (and other analogous terms). Finally, at the lowest level, the shared-memory simultaneous multi-threading (SMT) approach (based on OpenMP) is employed in order to further extend the scalability limit. At this level, each of the slabs is divided into a set of 1D "pencils," which are then attributed to the threads associated with a particular MPI task.

Using this multi-layered parallelization scheme, we have demonstrated efficient scalability using over ten thousand cores in a single QM/MM MD simulation while maintaining an overall parallel efficiency above 75% for a system containing a large Cl^-/H^+ antiporter protein embedded in a lipid membrane bilayer (**Figure 1B**) solvated in water. In this system, 19 atoms out of a total of 150,925 atoms were treated at the QM



level. The size of the whole system was $126.9 \times 126.8 \times 99.3 \text{ \AA}^3$, and the size of the cubic QM box was $17.7 \times 17.7 \times 17.7 \text{ \AA}^3$. We used a plane wave cutoff of 90 Ry, which corresponds to a real-space mesh with 240 points along each dimension. Benchmarks were performed using Troullier–Martins pseudopotentials (Troullier and Martins, 1991). The average wall time of a single MD time step is around 13 s (Bolnykh et al., 2019) when computationally demanding hybrid exchange–correlation functionals, such as B3LYP (Becke, 1988, 1993; Lee et al., 1988), are employed. This enables nanosecond-scale QM/MM MD simulations to be performed, which in turn allows one to obtain converged free energy calculations of biological systems if enough computational resources are available. Some representative scaling benchmark results are shown in **Figures 1C,D**. We expect similar extreme scalability for systems characterized by QM domains of similar size.

3. CONCLUSION

We have given a short introduction to the recently developed MiMiC framework as a highly flexible and extremely powerful multiscale modeling software solution capable of delivering unprecedented levels of scaling performance. The efficiency of the framework is ensured by using a well-established and extensively validated electrostatic embedding scheme while flexibility and modularity is achieved via an efficient loosely coupled MPMD architecture. Finally, extreme scalability is attained through a multi-layered parallelization strategy.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

UR acknowledges funding from the Swiss National Science foundation via the NCCR MUST and individual grant No. 200020_185092. This project/research has received funding from the 230 European Union's Horizon 2020 Framework Programme for Research and Innovation 231 under Specific Grant Agreement 720270 (Human Brain Project SGA2). JO acknowledges financial support from the Research Council of Norway through its Centres of Excellence scheme (Project ID: 262695). MB acknowledges a postdoctoral fellowship from the Swiss National Science Foundation (project 184500). PC acknowledges the funding by the Deutsche Forschungsgemeinschaft via FOR 2518 DynIon project P6 and by the Centre of Excellence for Computational Biomolecular Research, BioExcel CoE, funded by the European Union (contracts: H2020-INFRAEDI-02-2018-823830, H2020-EINFRA-2015-1-675728).

ACKNOWLEDGMENTS

The authors thank Dr. Maria Gabriella Chiarello for providing the structure of the protein used in benchmarks.

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001
- Abraham, M. J., van der Spoel, D., Lindahl, E., Hess, B., and the GROMACS Development Team (2019). *GROMACS User Manual version 2019*.
- Adhireksan, Z., Davey, G. E., Campomanes, P., Groessl, M., Clavel, C. M., Yu, H., et al. (2014). Ligand substitutions between ruthenium–cymene compounds can control protein versus DNA targeting and anticancer activity. *Nat. Commun.* 5:3462. doi: 10.1038/ncomms4462
- Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098–3100. doi: 10.1103/physreva.38.3098
- Becke, A. D. (1993). A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* 98, 1372–1377. doi: 10.1063/1.464304
- Behler, J., and Parrinello, M. (2007). Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* 98:146401. doi: 10.1103/PhysRevLett.98.146401
- Bolnykh, V., Olsen, J. M. H., Meloni, S., Ippoliti, E., Bircher, M. P., Carloni, P., et al. (2019). Extreme scalability of DFT-based QM/MM MD simulations using MiMiC. *J. Chem. Theory Comput.* 15, 5601–5613. doi: 10.1021/acs.jctc.9b00424
- Brunk, E., and Rothlisberger, U. (2015). Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.* 115, 6217–6263. doi: 10.1021/cr500628b
- Campomanes, P., Neri, M., Horta, B. A. C., Röhrig, U. F., Vanni, S., Tavernelli, I., et al. (2014). Origin of the spectral shifts among the early intermediates of the rhodopsin photocycle. *J. Am. Chem. Soc.* 136, 3842–3851. doi: 10.1021/ja411303v
- Campomanes, P., Rothlisberger, U., Alfonso-Prieto, M., and Rovira, C. (2015). The molecular mechanism of the catalase-like activity in horseradish peroxidase. *J. Am. Chem. Soc.* 137, 11170–11178. doi: 10.1021/jacs.5b06796
- Christensen, A. S., Faber, F. A., and von Lilienfeld, O. A. (2019). Operators in quantum machine learning: response properties in chemical space. *J. Chem. Phys.* 150:064105. doi: 10.1063/1.5053562
- Cupellini, L., Caprasecca, S., Guido, C. A., Müh, F., Renger, T., and Mennucci, B. (2018). Coupling to charge transfer states is the key to modulate the optical bands for efficient light harvesting in purple bacteria. *J. Phys. Chem. Lett.* 9, 6892–6899. doi: 10.1021/acs.jpclett.8b03233
- Genna, V., Vidossich, P., Ippoliti, E., Carloni, P., and De Vivo, M. (2016). A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc.* 138, 14592–14598. doi: 10.1021/jacs.6b05475
- Hutter, J., Alavi, A., Deutsch, T., Bernasconi, M., Goedecker, S., Marx, D., et al. (2018). *CPMD Copyright IBM Corp 1990-2019, MPI für Festkörperforschung Stuttgart 1997-2012*.
- Laio, A., VandeVondele, J., and Rothlisberger, U. (2002). A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116, 6941–6947. doi: 10.1063/1.1462041
- Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37, 785–789. doi: 10.1103/PhysRevB.37.785
- Li, J., Lyu, W., Rossetti, G., Konijnenberg, A., Natalello, A., Ippoliti, E., et al. (2017). Proton dynamics in protein mass spectrometry. *J. Phys. Chem. Lett.* 8, 1105–1112. doi: 10.1021/acs.jpclett.7b00127
- Loco, D., Buda, F., Lugtenburg, J., and Mennucci, B. (2018). The dynamic origin of color tuning in proteins revealed by a carotenoid pigment. *J. Phys. Chem. Lett.* 9, 2404–2410. doi: 10.1021/acs.jpclett.8b00763
- Morzan, U. N., de Armiño, D. J. A., Foglia, N. O., Ramírez, F., Lebrero, M. C. G., Scherlis, D. A., et al. (2018). Spectroscopy in complex environments from QM–MM simulations. *Chem. Rev.* 118, 4071–4113. doi: 10.1021/acs.chemrev.8b00026
- Olsen, J. M. H., Bolnykh, V., Meloni, S., Ippoliti, E., Bircher, M. P., Carloni, P., et al. (2019). MiMiC: a novel framework for multiscale modeling in computational chemistry. *J. Chem. Theory Comput.* 15, 3810–3823. doi: 10.1021/acs.jctc.9b00093
- Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chem.* 48, 1198–1229. doi: 10.1002/anie.200802019
- Singraber, A., Morawietz, T., Behler, J., and Dellago, C. (2019). Parallel multistream training of high-dimensional neural network potentials. *J. Chem. Theory Comput.* 15, 3075–3092. doi: 10.1021/acs.jctc.8b01092
- Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- Troullier, N., and Martins, J. L. (1991). Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* 43, 1993–2006. doi: 10.1103/physrevb.43.1993
- Warshel, A., and Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249. doi: 10.1016/0022-2836(76)90311-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bolnykh, Olsen, Meloni, Bircher, Ippoliti, Carloni and Rothlisberger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.