The PLUMED consortium: A community effort to promote openness, transparency and reproducibility in molecular simulations

The PLUMED consortium*

*To whom correspondence should be addressed: massimiliano.bonomi@pasteur.fr, bussi@sissa.it, carlo.camilloni@unimi.it, g.tribello@qub.ac.uk

Molecular dynamics (MD) simulations have become an important tool for characterizing mechanisms underlying complex processes, interpreting experimental measurements, and predicting the behavior of molecular systems. Success in this broad field has been achieved through more than 60 years of continuous innovation. During this time a wide community of researchers has developed accurate integrators of the equations of motion, effective thermostats, efficient methods for calculating electrostatic interactions, accurate physico-chemical descriptions of the system at quantum, classical and coarse-grained levels, a multitude of techniques to accelerate conformational sampling, sophisticated analysis tools, more powerful, personal and dedicated computers and new efficient, user-friendly software[1]. Together, these advances have ensured that MD now provides a powerful and versatile computational microscope, which can be applied in fields ranging from physics and chemistry to biology and material science. Furthermore, the basic techniques of MD can be employed at diverse spatial and temporal scales, ranging from chemical reactions, which require a quantum description of the system, to the dynamics of an entire cell, for which a hyper-coarse-grained representation is presently needed.

The fact that MD is now a well-established technique in many fields of computational science was recognized by the award of the 2013 Nobel Prize in Chemistry. Innovations in this field, however, are still emerging, and at an increasing pace. Areas in which there are still open challenges and hence active development by both the authors of this paper and others include extending the timescales accessible in standard simulations[2], new methods for dimensionality reduction, often based on machine-learning techniques, to make sense of the high-dimensional data generated by MD[3], and new integrative methods to improve the accuracy of molecular mechanics force fields by incorporating either quantum-chemical or experimental data[4]. Furthermore, despite great efforts have been made to define good practices for preparing, executing, and analyzing MD simulations[5], to make them reproducible, shareable and comparable across different MD codes and computational platforms, and ultimately to validate the results in a standardized way, there is still much work to do to ensure that the community is able to reap the maximal benefit from recent and ongoing developments.

The initiative that we present here stems from the realization that many of the challenges discussed above cannot be effectively overcome by individual researchers alone. Instead, a concerted effort by the entire simulation community is more likely to be successful. In recent years, numerous important initiatives have thus been launched to improve the way in which

simulation data and know-how is communicated. These initiatives include: SimTK[a], a platform for sharing software, data, and models for the biomedical computation community; the Open Force Field Initiative[b], which has been developed in coordination with the Molecular Software Sciences Institute[c] (MolSSI) to improve the computer models used in MD simulations; the Centre of Excellence for Computational Biomolecular Research[d] (BioExcel), whose mission is to improve biomolecular software and to spread best practices and expertise; the Living Journal of Computational Molecular Science[e], which publishes "living" educational reviews and best practice papers that are updated on a regular basis; Materials Cloud[f], a platform for seamlessly sharing resources in materials science, AiiDA[g], a flexible and scalable infrastructure to manage, preserve, and disseminate the simulations, data, and workflows of modern-day computational science, the E-CAM Centre of Excellence[h], which supports high-performance simulations in industry and academia, RosettaCommons[i], the central hub to contribute and share the ROSETTA[6] source code for computational modelling and analysis of protein structures, the NOMAD laboratory[j], which maintains the largest repository of input and output files for total energy calculations in materials science, The Materials Project[k], which provides open web-based access to computed information on known and predicted materials, Qresp[l], a tool for curating, discovering and exploring reproducible scientific papers, and nanoHUB[m], a platform to share simulation software and educational tools for nanoscience. A more exhaustive list of data repositories has been compiled by Scientific Data[n].

While these initiatives have laid crucial groundwork to increase the accessibility and reproducibility of MD, their effectiveness is hampered by the fact that it is difficult to use multiple MD codes within the same research project and to carry out development on multiple codes at the same time, as each one is designed with distinct goals and optimized for specific applications. While several methods have the potential to be useful across varying fields and platforms, their applicability is hindered by this lack of interoperability between MD codes. A strategy to resolve this problem was pioneered a decade ago with the creation of PLUMED[7], an open-source library that can be used in combination with many different MD codes (see **Box 1**). This library contains a range of different methods including enhanced-sampling algorithms, free-energy methods, and tools to analyze the vast amounts of data produced by MD simulations. Similar interoperable libraries, such as the collective variables module for molecular simulation

---

[a] https://simtk.org
[b] https://openforcefield.org
[c] https://molssi.org
[d] https://bioexcel.eu
[e] https://www.livecomsjournal.org
[f] https://www.materialscloud.org
[g] http://www.aiida.net
[h] https://www.e-cam2020.eu/
[i] https://www.rosettacommons.org
[j] https://nomad-coe.eu
[k] https://materialsproject.org
[l] http://qresp.org
[m] https://nanohub.org
[n] https://www.nature.com/sdata/policies/repositories

programs (Colvars)[8] and the software suite for advanced generalized ensemble simulations (SSAGES)[9], have since been developed.

Recently, PLUMED has incorporated additional functionalities that surpass those that were originally conceived by its core developers. As a result of a series of technical innovations and the development of an infrastructure which allows external contributions to be easily incorporated into the code (see **Box 2**), PLUMED has evolved into a flexible open-source library[10]. As of today, PLUMED can be used with both state-of-the-art classical and *ab initio* MD codes, such as ACEMD[11], Amber[12], DL_POLY[13], GROMACS[14], LAMMPS[15], NAMD[16], OpenMM[17], ABIN[o], CP2K[18], i-PI[19], PINY-MD[20] and Quantum Espresso[21]. Furthermore, PLUMED has also been used to enhance the capabilities of analysis tools such as VMD[22], and platforms such as HTMD[23] and OpenPathSampling[24]. When new techniques are implemented in PLUMED, they can thus be rapidly disseminated to a large and diverse user base. As a matter of fact, PLUMED has made all these techniques accessible and easy to use for everyone. The impact across multiple communities is further accelerated by the fact that PLUMED uses a simple and unified syntax for all the MD codes and analysis tools with which it can be used. The interoperability that PLUMED provides ultimately allows for cross-validation between different MD software and cross-fertilization of ideas between different fields of computational chemistry, biophysics and materials science.

The success of the open-source PLUMED library is due not only to the contributions of the core developers but also to the efforts of an entire community of users and developers. To acknowledge this working model, here we announce the foundation of the PLUMED consortium, whose mission is to transform how scientists communicate MD protocols towards increasing transparency and maximizing impact of new research. The primary way in which the consortium achieves this goal is through the continued, community-driven development of the PLUMED open-source library, which welcomes the addition of new code from all members of the consortium as well as other researchers. Additionally, the consortium will also collectively strive to promote good practices in enhanced-sampling MD simulations to encourage the highest possible standards of scientific reproducibility. Finally, the consortium will be actively engaged in introducing and training the newest members in the community in the methods that are disseminated via the PLUMED library.

To increase clarity and transparency of MD simulations, we have created a public repository, called PLUMED-NEST[p], which all members of the consortium as well as other researchers are encouraged to use to share their PLUMED input files along with all the data required to replicate the calculations presented in their peer-reviewed papers and preprints (see **Box 3**)**.** This repository will not only promote scientific reproducibility, but it will also serve as an instrument for novice users to learn the basic and advanced techniques implemented in PLUMED through a series of real-life applications in the fields of computational chemistry, physics, and biology. In this regard, the repository will also help us to develop resources for the education and training of

[o] https://github.com/PHOTOX/ABIN
[p] https://www.plumed-nest.org

the new generations of scientists in the community - an objective that is currently met and that will continue to be met by organizing user meetings, in-person and online tutorials.

The PLUMED consortium will be an open community composed of current and past developers, contributors, and all those researchers whose work builds in part on PLUMED and at the same time drives the development and dissemination of PLUMED. The list of the members of the consortium will regularly be updated on the PLUMED website[q] and the current PLUMED developers will act as coordinators for this community. We strongly believe that this novel organization, which unifies core developers and contributors to the code, is the most representative way of reflecting the community-driven effort that is the heart and soul of open-source software development, which is a crucial part of any methodological advancement[25]. Therefore, we welcome all future members who share our vision.

Box 1: Overview of the PLUMED library

PLUMED is an open-source C++ library that can be interfaced with many state-of-the-art MD codes. Its basic functionalities are: *i)* to retrieve the current atom positions from the main code; *ii)* to calculate additional quantities, such as coarse-grained representations of the system (collective variables) and external biasing potentials; *iii)* to communicate external forces back to the MD code and hence to modify the dynamics. PLUMED provides all the functionalities needed to perform MD simulations using well-established enhanced-sampling methods, such as umbrella sampling[26], metadynamics[27], and steered MD[28] as well as more recent approaches. These techniques can be used in combination with a large toolbox of collective variables that describe complex processes in physics, chemistry, material science, and biology. PLUMED can also calculate these descriptors *a posteriori* on pre-calculated trajectories via the *driver* utility and can perform various other analyses such as dimensionality reduction or reweighting.

PLUMED can be interfaced with the host code by using a single well-documented API that enables the PLUMED functionalities to be imported. The API is accessible from multiple languages (C, C++, FORTRAN, and Python), and is thus compatible with the majority of the codes used in the community. The PLUMED license (L-GPL) allows it to be interfaced also with proprietary software. Several MD codes, including AMBER, LAMMPS, DL_POLY, i-PI, and OpenMM now have the interface to PLUMED natively implemented.

The current structure of PLUMED is modular, *i.e.* new functionalities can be added without changing the core of the code. New collective variables and new biasing or analysis methods can thus either be distributed with the main core or loaded as separate dynamic libraries. The ease with which PLUMED can be extended is evidenced by the number of forks of the official repository[r] that have been created on GitHub by independent groups and by the growing number of additional modules contributed to PLUMED (see **Box 2**).

[q] https://www.plumed.org
[r] https://github.com/plumed/plumed2

Box 2: Overview of the current PLUMED modules

Currently, six modules have been contributed to the official PLUMED distribution by groups external to the core developers of the code. All these modules are independent from the PLUMED core and need to be activated when configuring PLUMED. The largest contributed module (VES) implements the Variationally Enhanced Sampling[29] approach and includes a number of basis functions, target distributions, and optimization algorithms. The DRR module implements the Extended-System Adaptive Biasing Force method for enhanced-sampling and free-energy calculations[30]. The EDS module implements the Experiment Directed Simulation approach to adaptively construct linear restraints that ensure that the biased collective variables sample a new target mean value[31]. More recently, three additional modules have been contributed to PLUMED and implement the enhanced-sampling method called logarithmic mean-force dynamics[32], the permutation invariant vector[33], a general-purpose representation of the structure of materials that can be used to analyze and simulate transitions between ordered or disordered forms, as in crystal nucleation, and a method to find multiple diverse reaction pathways of ligand unbinding, available in the MAZE module[34]. Finally, the PLUMED-ISDB[35] module has been contributed by a subset of PLUMED developers and enables different types of experimental data, such as measurements from nuclear magnetic resonance spectroscopy, small-angle X-ray scattering, and cryo-electron microscopy, to be integrated into MD simulations. All the collective variables and enhanced-sampling approaches contributed via these modules can be seamlessly combined with all the other functionalities implemented in PLUMED. Each module is accompanied by documentation and tutorials that are integrated in the online PLUMED documentation.

Box 3: PLUMED-NEST, the public repository of the PLUMED consortium

A public repository has been created to collect contributions from members of the PLUMED consortium and from other PLUMED users. Dubbed PLUMED-NEST, it hosts the PLUMED input files and the links to all the data required to reproduce the results of enhanced-sampling simulations or analysis that have been carried out with PLUMED. Currently, PLUMED-NEST already contains more than 40 examples of applications and method developments in the areas of computational biology, chemistry, and material science.

All PLUMED input files deposited in PLUMED-NEST are automatically tested to verify whether they are still functional with the current and development versions of the code. Furthermore, the keywords in input files appear as links to the documentation in the manual so that the users can easily access more information about what is being computed. It is also possible to add contextual, tutorial-like information to the input files and thus to provide additional details about what is being computed by a particular file. Finally, we envision incorporating commenting functionalities on the input files included in the repository thus providing direct feedback to the authors of the original papers.

PLUMED-NEST will serve multiple functions. It will *i)* promote scientific reproducibility by offering users a place to share with the community all the information that is required to repeat a simulation or analysis reported in a published paper; *ii)* serve as a repository of real-life examples for educational purposes, and thus complement the tutorials that are already available on the PLUMED website; *iii)* enable the PLUMED developers to identify which functionalities are most used and thus guide them in improving the PLUMED core code and documentation.

References

1. Allen, M.P. & Tildesley, D.J. Computer simulation of liquids, Edn. Second edition. (Oxford University Press, Oxford, United Kingdom; 2017).
2. Camilloni, C. & Pietrucci, F. Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Adv Phys X* **3**, 1477531 (2018).
3. Mittal, S. & Shukla, D. Recruiting machine learning methods for molecular simulations of proteins. *Mol Simulat* **44**, 891-904 (2018).
4. Bottaro, S. & Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355-360 (2018).
5. Grossfield, A. et al. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations. *Living J Comp Mol Sci* **1**, 5067 (2018).
6. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487**, 545-574 (2011).
7. Bonomi, M. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* **180**, 1961-1972 (2009).
8. Fiorin, G., Klein, M.L. & Henin, J. Using collective variables to drive molecular dynamics simulations. *Mol Phys* **111**, 3345-3362 (2013).
9. Sidky, H. et al. SSAGES: Software Suite for Advanced General Ensemble Simulations. *J Chem Phys* **148**, 044104 (2018).
10. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* **185**, 604-613 (2014).
11. Harvey, M.J., Giupponi, G. & De Fabritiis, G. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **5**, 1632-1639 (2009).
12. Case, D.A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688 (2005).
13. Todorov, I.T., Smith, W., Trachenko, K. & Dove, M.T. DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *J Mater Chem* **16**, 1911-1918 (2006).
14. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
15. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J Comput Phys* **117**, 1-19 (1995).
16. Phillips, J.C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).

17. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Comput Biol* **13**, e1005659 (2017).
18. Hutter, J., Iannuzzi, M., Schiffmann, F. & VandeVondele, J. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip Rev Comput Mol Sci* **4**, 15-25 (2013).
19. Ceriotti, M., More, J. & Manolopoulos, D.E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Comput Phys Commun* **185**, 1019-1026 (2014).
20. Tuckerman, M.E., Yarne, D.A., Samuelson, S.O., Hughes, A.L. & Martyna, G.J. Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Comput Phys Commun* **128**, 333-376 (2000).
21. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J Phys-Condens Mat* **21**, 395502 (2009).
22. Giorgino, T. PLUMED-GUI: An environment for the interactive development of molecular dynamics analysis and biasing scripts. *Comput Phys Commun* **185**, 1109-1114 (2014).
23. Doerr, S., Harvey, M.J., Noe, F. & De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **12**, 1845-1852 (2016).
24. Swenson, D., Prinz, J.-H., Noé, F., Chodera, J.D. & Bolhuis, P.G. OpenPathSampling: A Python framework for path sampling simulations. I. Basics. *J Chem Theory Comput* **15**, 813-836 (2019).
25. Giving software its due. *Nat Methods* **16**, 207-207 (2019).
26. Kästner, J. Umbrella sampling. *Wiley Interdiscip Rev Comput Mol Sci* **1**, 932-942 (2011).
27. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip Rev Comput Mol Sci* **1**, 826-843 (2011).
28. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys* **119**, 3559-3566 (2003).
29. Valsson, O. & Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys Rev Lett* **113**, 090601 (2014).
30. Chen, H., Fu, H., Shao, X., Chipot, C. & Cai, W. ELF: An Extended-Lagrangian Free Energy Calculation Module for Multiple Molecular Dynamics Engines. *J Chem Inf Model* **58**, 1315-1318 (2018).
31. White, A.D. & Voth, G.A. Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J Chem Theory Comput* **10**, 3023-3030 (2014).
32. Morishita, T., Yonezawa, Y. & Ito, A.M. Free Energy Reconstruction from Logarithmic Mean-Force Dynamics Using Multiple Nonequilibrium Trajectories. *J Chem Theory Comput* **13**, 3106-3119 (2017).
33. Pipolo, S. et al. Navigating at Will on the Water Phase Diagram. *Phys Rev Lett* **119**, 245701 (2017).
34. Rydzewski, J. maze: Heterogeneous Ligand Unbinding along Transient Protein Tunnels. *arXiv*, 1904.03929 (2019).
35. Bonomi, M. & Camilloni, C. Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* **33**, 3999-4000 (2017).

List of members of the consortium and authors of the paper

- Massimiliano Bonomi, Institut Pasteur - CNRS, Paris, France
- Giovanni Bussi, International School for Advanced Studies, Trieste, Italy
- Carlo Camilloni, University of Milan, Milan, Italy
- Gareth Tribello, Queen's University Belfast, Belfast, UK
- Pavel Banáš, Palacky University, Olomouc, Czech Republic
- Alessandro Barducci, Centre de Biochimie Structurale de Montpellier, Montpellier, France
- Mattia Bernetti, International School for Advanced Studies, Trieste, Italy
- Peter G. Bolhuis, University of Amsterdam, Amsterdam, The Netherlands
- Sandro Bottaro, Italian Institute of Technology, Genova, Italy
- Davide Branduardi, Schrödinger Inc., Cambridge, UK
- Riccardo Capelli, Forschungszentrum Jülich, Jülich, Germany
- Paolo Carloni, Forschungszentrum Jülich, Jülich, Germany
- Michele Ceriotti, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland
- Andrea Cesari, International School for Advanced Studies, Trieste, Italy
- Haochuan Chen, Nankai University, Tianjin, China
- Francesco Colizzi, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
- Sandip De, BASF, Ludwigshafen, Germany
- Marco DeLaPierre, Pawsey Supercomputing Centre, Kensington, Australia
- Davide Donadio, University of California Davis, Davis, USA
- Viktor Drobot, Lomonosov Moscow State University, Moscow, Russia
- Bernd Ensing, University of Amsterdam, Amsterdam, The Netherlands
- Marta Filizola, Icahn School of Medicine at Mount Sinai, New York, USA
- James Fraser, University of California San Francisco, San Francisco, USA
- Haohao Fu, Nankai University, Tianjin, China
- Piero Gasparotto, University College London, London, UK
- Francesco Luigi Gervasio, University College London, London, UK
- Federico Giberti, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland
- Alejandro Gil-Ley, National Institutes of Health, Bethesda, USA
- Toni Giorgino, CNR-IBF, National Research Council, Milan, Italy
- Gabriella T. Heller, University of Cambridge, Cambridge, UK
- Glen Hocky, New York University, New York, USA
- Marcella Iannuzzi, University of Zurich, Zurich, Switzerland
- Michele Invernizzi, ETH Zurich and Università della Svizzera italiana, Lugano, Switzerland
- Kim Jelfs, Imperial College London, London, UK
- Alexander Jussupow, Technische Universität München, Munich, Germany
- Evgeny Kirilin, Lomonosov Moscow State University, Moscow, Russia
- Alessandro Laio, International School for Advanced Studies, Trieste, Italy
- Vittorio Limongelli, Università della Svizzera italiana, Lugano, Switzerland and University of Naples "Federico II", Naples, Italy

- Kresten Lindorff-Larsen, University of Copenhagen, Copenhagen, Denmark
- Thomas Löhr, University of Cambridge, Cambridge, UK
- Fabrizio Marinelli, National Institutes of Health, Bethesda, USA
- Layla Martin-Samos, CNR-IOM Democritos, Trieste, Italy
- Matteo Masetti, University of Bologna, Bologna, Italy
- Ralf Meyer, Graz University of Technology, Graz, Austria
- Angelos Michaelides, University College London, London, UK
- Carla Molteni, King's College London, London, UK
- Tetsuya Morishita, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
- Marco Nava, University of Illinois, Urbana-Champaign, USA
- Cristina Paissoni, University of Milan, Milan, Italy
- Elena Papaleo, Danish Cancer Society Research Center, Copenhagen, Denmark
- Michele Parrinello, ETH Zurich and Università della Svizzera italiana, Lugano, Switzerland; Italian Institute of Technology, Genova, Italy
- Jim Pfaendtner, University of Washington, Seattle, USA
- Pablo Piaggi, ETH Zurich and Università della Svizzera italiana, Lugano, Switzerland
- GiovanniMaria Piccini, ETH Zurich and Università della Svizzera italiana, Lugano, Switzerland
- Adriana Pietropaolo, University of Catanzaro, Catanzaro, Italy
- Fabio Pietrucci, Sorbonne Université, Paris, France
- Silvio Pipolo, Université de Lille, Lille, France
- Davide Provasi, Icahn School of Medicine at Mount Sinai, New York, USA
- David Quigley, University of Warwick, Coventry, UK
- Paolo Raiteri, Curtin University, Perth, Australia
- Stefano Raniolo, Università della Svizzera italiana, Lugano, Switzerland
- Jakub Rydzewski, Nicolaus Copernicus University, Torun, Poland
- Matteo Salvalaglio, University College London, London, UK
- Gabriele Cesare Sosso, University of Warwick, Coventry, UK
- Vojtech Spiwok, University of Chemistry and Technology, Prague, Czech Republic
- Jiří Šponer, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic
- David W. H. Swenson, École Normale Supérieure de Lyon, Lyon, France
- Pratyush Tiwary, University of Maryland, College Park, USA
- Omar Valsson, Max Planck Institute for Polymer Research, Mainz, Germany
- Michele Vendruscolo, University of Cambridge, Cambridge, UK
- Gregory A. Voth, University of Chicago, Chicago, USA
- Andrew White, University of Rochester, Rochester, USA