

A community effort to promote transparency and reproducibility in enhanced 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

List of members of the consortium and authors of the paper

- Massimiliano Bonomi, Institut Pasteur, CNRS UMR 3528, 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
- Wei Chen, University of Illinois, Urbana-Champaign, Illinois, USA
- 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, New South Wales, Australia
- Davide Donadio, University of California Davis, Davis, California, USA
- Viktor Drobot, Lomonosov Moscow State University, Moscow, Russia
- Bernd Ensing, University of Amsterdam, Amsterdam, The Netherlands
- Andrew Ferguson, University of Chicago, Chicago, Illinois, USA
- Marta Filizola, Icahn School of Medicine at Mount Sinai, New York, New York, USA
- James Fraser, University of California San Francisco, San Francisco, California, 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, Maryland, USA
- Toni Giorgino, CNR-IBF, National Research Council, Milan, Italy

- Gabriella T. Heller, University of Cambridge, Cambridge, UK
- Glen M. Hocky, New York University, New York, 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, Maryland, 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, Illinois, USA
- Cristina Pissoni, 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, Washington, 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, New York, USA
- David Quigley, University of Warwick, Coventry, UK
- Paolo Raiteri, Curtin University, Perth, Western Australia, 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, Maryland, 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, Illinois, USA
- Andrew White, University of Rochester, Rochester, New York, USA

The PLUMED consortium is an open community that unifies developers and contributors to PLUMED, an open-source library for enhanced-sampling, free-energy calculations and the analysis of molecular dynamics (MD) simulations. This commentary outlines the consortium's efforts to promote transparency and scientific reproducibility by disseminating protocols for enhanced-sampling molecular simulations.

MD simulations have become an important tool for characterizing the mechanisms underlying complex processes, interpreting experimental measurements, and predicting the behavior of molecular systems. This success has been achieved through more than 60 years of innovation. During this time, a wide community of researchers has developed accurate integrators of the equations of motion, effective thermostats, accurate physico-chemical models of small molecules, proteins, nucleic acids, liquids and materials, many techniques to accelerate sampling, sophisticated analysis tools, powerful, personal and dedicated computers and efficient, user-friendly software. These advances ensure that MD now provides a powerful and versatile computational microscope, which can be applied in fields ranging from physics and chemistry to biology and materials science.

Innovations in this field are still emerging at an increasing pace. Areas in which there are still open challenges and hence active development include: extending the timescales accessible in standard simulations¹, new methods for making sense of the high-dimensional data generated by MD², and new integrative approaches for improving the accuracy of force fields by incorporating quantum-chemical and experimental data³. Furthermore, although great efforts have been made to define good practices for preparing, executing, and analyzing MD simulations⁴, much work is still required to ensure that the community can 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 required. Numerous important initiatives have thus been launched to improve the way in which data and know-how is communicated. These initiatives include: SimTK^a, a platform for sharing software and data for the biomedical computation community; the Open Force Field Initiative^b, which has been developed in coordination with the Molecular Software Sciences Institute^c to improve the physico-chemical models used in MD; the Centre of Excellence for Computational Biomolecular Research^d, whose mission is to improve biomolecular software and spread best practices; the Living Journal of Computational Molecular Science^e, which publishes and regularly updates educational reviews and best practice papers; Materials Cloud^f, a platform for sharing resources in materials science; AiiDA^g, an infrastructure to disseminate simulation data and workflows; the E-CAM Centre of Excellence^h, which supports

^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/>

high-performance simulations in industry and academia; the NOMAD laboratoryⁱ, which maintains a repository of input/output files for total energy calculations in materials science; The Materials Project^j, which provides open access to information on known and predicted materials; Qresp^k, a tool for curating and exploring reproducible scientific papers; and nanoHUB^l, a platform to share simulation software and educational tools for nanoscience. A more exhaustive list of data repositories has been compiled by Scientific Data^m. Furthermore, these initiatives have been complemented by events focused on promoting openness, transparency and reproducibility in MD simulations, such as the recent workshop “*Sharing Data from Molecular Simulations*”, held in Stockholm in November 2018ⁿ.

While these initiatives have laid crucial groundwork that increases reproducibility for MD simulations, their effectiveness is hampered by the lack of interoperability between MD codes. As matter of fact, methods that have the potential to be useful across different fields cannot be easily incorporated across multiple MD softwares, as each one of these codes is optimized for specific applications and might be written in a different programming language. A strategy to resolve this problem was pioneered with the creation of PLUMED⁵, an open-source library that provides enhanced-sampling algorithms, free-energy methods, and tools to analyze the vast amounts of data produced by MD simulations (see **Box 1**). Similar interoperable libraries, such as the collective variables module for molecular simulation programs (Colvars)⁶ and the software suite for advanced generalized ensemble simulations (SSAGES)⁷, have since been developed.

Recently, PLUMED has incorporated functionalities that surpass those originally conceived by its creators. The development of an infrastructure that allows external contributions to be easily incorporated into the code (see **Box 2**) has transformed PLUMED into a flexible open-source library⁸. As of today, PLUMED can be used with both classical and *ab initio* MD codes, such as ACEMD⁹, Amber¹⁰, DL_POLY¹¹, GROMACS¹², LAMMPS¹³, NAMD¹⁴, OpenMM¹⁵, ABIN^o, CP2K¹⁶, i-PI¹⁷, PINY-MD¹⁸ and Quantum Espresso¹⁹. Furthermore, PLUMED can now be used to enhance the capabilities of analysis tools such as VMD²⁰, and platforms such as HTMD²¹ and OpenPathSampling²². When new techniques are implemented in PLUMED, they can be rapidly disseminated to a large and diverse user base, thus making them accessible and easy for everyone to use. This impact across multiple communities is further accelerated by the fact that PLUMED uses a unified syntax for all the programs with which it can be used. The interoperability enabled by PLUMED ultimately allows for cross-validation between various MD softwares and cross-fertilization of ideas between computational chemistry, biophysics and materials science.

The success of PLUMED is not just due to the contributions of few core developers but also to the efforts of a wider community of users and developers. To acknowledge this model, we are announcing the foundation of the PLUMED consortium, whose mission is to transform how scientists communicate MD protocols so as to increase reproducibility and maximize the impact of new research. The primary way in which the consortium achieves this goal is through the

ⁱ <https://nomad-coe.eu>

^j <https://materialsproject.org>

^k <http://qresp.org>

^l <https://nanohub.org>

^m <https://www.nature.com/sdata/policies/repositories>

ⁿ <https://zenodo.org/record/2652703>

^o <https://github.com/PHOTOX/ABIN>

continued, community-driven development of PLUMED. Additionally, the consortium will collectively strive to promote good practices for selecting the simulation parameters to be used in enhanced-sampling MD simulations, for reporting simulation protocols and for quantifying the statistical uncertainty in calculated observables. To facilitate the adoption of these practices and to encourage the highest possible standards of scientific reproducibility, consortium members will share the input files and protocols for the simulations in their published works.

To realize the mission of the consortium, we have created the PLUMED-NEST^P repository, which all members as well as other researchers are encouraged to use to share their PLUMED input files and all other data required to replicate the calculations presented in their papers (see **Box 3**). This repository will not only promote scientific reproducibility, it will also serve as an instrument that novices can use to study real-life applications of these techniques in the fields of computational chemistry, physics, and biology. In this regard, the repository will help us develop resources for the training of the new generations of scientists in the community - an objective that will continue to be met by organizing user meetings and both in-person and online tutorials.

The PLUMED consortium is an open community composed of developers, contributors, and all those researchers whose work builds in part on PLUMED and at the same time drives its development. The list of the members of the consortium will be updated regularly^Q and the PLUMED core developers will act as coordinators. We strongly believe that this novel organization best represents the community-driven effort that is the heart and soul of open-source software development, which is a crucial part of any methodological advancement²³. Furthermore, this de-centralized organization with the support of a wide community will be essential for the long-term sustainment of PLUMED. We, therefore, welcome all future members who share our vision.

Contributions

All authors contributed to writing the manuscript.

Competing Financial Interests Statement

G. M. H. is currently consulting on a US Department of Energy grant to Parallel Works, Inc.

^P <https://www.plumed-nest.org>

^Q <https://www.plumed.org>

References

1. Camilloni, C. & Pietrucci, F. Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Adv Phys X* **3**, 1477531 (2018).
2. Mittal, S. & Shukla, D. Recruiting machine learning methods for molecular simulations of proteins. *Mol Simulat* **44**, 891-904 (2018).
3. Bottaro, S. & Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355-360 (2018).
4. Grossfield, A. et al. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations. *Living J Comp Mol Sci* **1**, 5067 (2018).
5. Bonomi, M. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* **180**, 1961-1972 (2009).
6. Fiorin, G., Klein, M.L. & Henin, J. Using collective variables to drive molecular dynamics simulations. *Mol Phys* **111**, 3345-3362 (2013).
7. Sidky, H. et al. SSAGES: Software Suite for Advanced General Ensemble Simulations. *J Chem Phys* **148**, 044104 (2018).
8. 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).
9. 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).
10. Case, D.A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688 (2005).
11. 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).
12. Abraham, M.J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
13. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J Comput Phys* **117**, 1-19 (1995).
14. Phillips, J.C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
15. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Comput Biol* **13**, e1005659 (2017).
16. 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).

17. 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).
18. 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).
19. 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).
20. Giorgino, T. PLUMED-GUI: An environment for the interactive development of molecular dynamics analysis and biasing scripts. *Comput Phys Commun* **185**, 1109-1114 (2014).
21. 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).
22. 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).
23. Giving software its due. *Nat Methods* **16**, 207-207 (2019).
24. Valsson, O. & Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys Rev Lett* **113**, 090601 (2014).
25. 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).
26. 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).
27. 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).
28. Pipolo, S. et al. Navigating at Will on the Water Phase Diagram. *Phys Rev Lett* **119**, 245701 (2017).
29. Rydzewski, J. maze: Heterogeneous Ligand Unbinding along Transient Protein Tunnels. *arXiv*, 1904.03929 (2019).
30. Bonomi, M. & Camilloni, C. Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* **33**, 3999-4000 (2017).

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 or CVs) and external biasing potentials; *iii*) to communicate external forces back to the MD code and to hence modify the dynamics. PLUMED provides all the functionalities needed to perform MD simulations using well-established enhanced-sampling methods, such as umbrella sampling, metadynamics, and steered MD. These techniques can be used in combination with a large toolbox of CVs that describe complex processes in physics, chemistry, material science, and biology. These CVs include distances between pairs of atoms, torsional angles, secondary structure content, radius of gyration, contact maps, energy of the system, various metrics to measure the distance from reference conformations, and functionalities to employ user-defined CVs based on artificial neural networks. PLUMED can also calculate these descriptors *a posteriori* on pre-calculated trajectories via a postprocessing utility called *driver* and can perform other analyses such as dimensionality reduction or calculating equilibrium properties of CVs not directly biased in the simulation.

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 also be interfaced with proprietary software. Several MD codes, including AMBER, LAMMPS, DL_POLY, i-PI, and OpenMM now have the interface to PLUMED natively implemented. Furthermore, the PLUMED developers currently provide and maintain a set of patches for popular MD packages, such as GROMACS and NAMD.

The current structure of PLUMED is modular, *i.e.* new functionalities can be added without changing the core of the code. New CVs 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**).

Box 2: Overview of the current PLUMED modules

All members of the consortium as well as other researchers can contribute an external module to PLUMED. As described in the PLUMED developer manual, contributed modules must pass quality control checks before being incorporated into the official PLUMED distribution and should have an open source license. Currently, six modules have been contributed 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 Sampling²⁴ 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

^r <https://github.com/plumed/plumed2>

calculations²⁵. The EDS module implements the Experiment Directed Simulation approach to adaptively construct linear restraints that ensure that the biased CVs sample a new target mean value²⁶. The logmfd module can be used to perform enhanced-sampling using logarithmic mean-force dynamics²⁷. The piv module contains an implementation of the permutation invariant vector²⁸, which is a general purpose representation of the structure of materials that can be used to analyze and simulate the transitions between ordered and disordered forms that take place in processes such as crystal nucleation. The MAZE module²⁹ can be used to find multiple diverse reaction pathways for ligand unbinding. Furthermore, the PLUMED-ISDB³⁰ 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 CVs 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 users. Dubbed PLUMED-NEST, it hosts the PLUMED input files and links to all the data required to reproduce the results of enhanced-sampling simulations or analyses that have been carried out with PLUMED. PLUMED-NEST already contains more than 50 examples of applications and method developments in the areas of computational biology, chemistry, and material science. The repository can be searched based on the type of contribution (method development, application in chemistry, biology and material science), author, project name, and specific keywords. In the future, it will also be possible to search the PLUMED input files based on the functionalities and keywords used.

All input files deposited in PLUMED-NEST are automatically tested to ensure that both the current and the development versions of PLUMED can successfully parse the input files provided. 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 to provide 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.