DEPARTMENT: EDUCATION

MolSSI Education: Empowering the Next Generation of Computational Molecular Scientists

Jessica A. Nash [ID], Mohammad Mostafanejad, and T. Daniel Crawford [ID], *Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA*

Ashley Ringer McDonald [ID], *California Polytechnic State University, San Luis Obispo, CA, 93401, USA*

The Molecular Sciences Software Institute (MolSSI) is a research and education center that supports software development in the computational molecular sciences (CMS). One of MolSSI's core objectives is to provide education and training for the next generation of computational researchers. MolSSI Education targets various career stages and skill levels through its live workshops, online resources, and software fellowship program. MolSSI Education focuses its efforts on four areas: programming and software development, high-performance computing, artificial intelligence, faculty and curriculum development, and the software fellowship program. This article delineates educational efforts at the MolSSI, overall goals, and resources that can be useful to researchers in the CMS.

Computational molecular science (CMS) is a broad term that describes the application of computational resources to chemical theory and systems. It is part of many fields, including chemistry, physics, materials science, and chemical engineering, and encompasses methods such as molecular mechanics, electronic structure, and machine learning (ML). Today, computational methods are widely used in industry and academia and are indispensable parts of many scientific breakthroughs.

Traditionally, computational practitioners have been the ones responsible for building the tools of the trade, namely, software. Over the past few decades, computational researchers have developed dozens of software packages that are used by thousands of scientists worldwide. Often, the software has been seen as a byproduct of scientific research. However, as scientific problems become more complex, it is necessary for scientists to focus on software engineering, a trend that is reflected in the emergence of the Research Software Engineer. For scientific software developers, deep knowledge of programming languages can be vital for creating performant and usable software. For researchers, knowing how to program can help improve research efficiency and increase the speed of data analysis. Thus, computational scientists often need both mastery of their scientific field and competency in computing skills such as programming and software development.

The Molecular Sciences Software Institute (MolSSI) aims to enable new science by supporting software development efforts in CMS. The MolSSI was founded in 2016 with a grant from the U.S. National Science Foundation (NSF) and is a collaborative, multi-institute center. In 2021, the NSF renewed MolSSI's funding for an additional five years. MolSSI provides software infrastructure, standards, community engagement, and training and education. MolSSI's open-source software projects address several needs in the community.[1] MolSSI's training programs include workshops, summer schools, and a fellowship program for graduate students and postdoctoral researchers. Since its establishment, MolSSI has funded 95 Software Fellow projects, hosted more than

25 community-led workshops, and reached more than 1500 students through educational events.

Our education program focuses on four areas: i) programming and software development, ii) high-performance computing (HPC), iii) faculty and curriculum development, and iv) a software fellowship program. By working in these areas, we hope to reach researchers across a broad range of skill levels and career stages. A full list of MolSSI Education resources can be accessed on the Education website.[2] For all of our training materials, we directly engage with the community. We deliver these resources through synchronous workshops and asynchronously through the Education website or YouTube channel.

MOLSSI EDUCATION COMPONENTS

Training in Programming and Software Development

The MolSSI's programming and software development resources are designed for beginner to intermediate programmers and introduce fundamental principles using examples relevant to computational molecular scientists. MolSSI's resources in this area focus on establishing programming skills and best practices for scientific software development using specific examples. Our curricula in programming and software development are created by software scientists and associates.

We lead our programming and software development workshops using a live-coding style. Live coding is a method popularized by the Software Carpentry organization and has been found to be successful in training novice programmers. In the live-coding approach to teaching, an instructor will share their screen and type code into their programming environment while explaining the thought process and reasoning behind their actions. We intersperse live coding with small challenges or exercises to allow students to apply concepts they have just learned.

MolSSI Education offers an introduction to Python programming in its flagship undergraduate workshop, Python Data and Scripting for Computational Molecular Scientists. In this workshop, students are introduced to Python syntax, working with text files, visualization, and running command-line programs. We offer this workshop synchronously once or twice a year. Registration is free and open to the public, though we typically partner with organizations that focus on undergraduate researchers such as the MERCURY Consortium.[3,4]

For students who have more experience programming or who are planning to work on software development projects, we offer the Python Package Best Practices workshop. The "Best Practices" covered in this workshop are topics recommended and often used in scientific and open-source software development, presented in a cohesive, hands-on format. Topics include version control, collaboration workflows, testing, documentation, and project structure. These are all practical skills widely used in software projects but currently rarely taught formally. Concepts are covered at a high level first, then, we demonstrate them with hands-on material. For example, when introducing software testing, we discuss the benefits and motivations of testing and also show specific examples of how one might test Python code. These workshops can be requested for groups or universities by contacting the Education team. We typically offer this workshop at least twice a year either to the public or in partnership with an academic research group.

MolSSI's other efforts in this area include a workshop on data visualization and a Python scripting workshop aimed at biochemists. In addition, MolSSI offers workshop materials on object-oriented programming and design patterns. A full list of resources is given in Table 1.

Training in HPC and Artificial Intelligence (AI)

MolSSI Education's newest initiative is in high performance computing (HPC) and artificial intelligence (AI). Our HPC and AI Education Programs consist of five major divisions: online educational resources,[2] certified university curricula, industrial training programs, instructor-led hands-on workshops, and community guidelines and best practices.[5] MolSSI Education launched this initiative in 2021 in recognition of the key focus areas of the Exascale Computing Project, top national priorities, and strategic plans.

We base our HPC online educational resources on open-source industry standards, vendors' expert recommendations, and community guidelines and best practices. Table 1 provides a high-level view of our online educational resources in HPC designed for a variety of user backgrounds and skill levels. For example, the fundamentals of the program would be most beneficial to beginner and intermediate-level users, while the homogeneous parallel programming resources are designed for intermediate and advanced users. The heterogeneous parallel programming section, on the other hand, provides online resources for all background levels.

A key component of the MolSSI Education program in HPC is industry partnerships. In collaboration with NVIDIA and Intel, we offer a series of certified instructor-led hands-on workshops. Our collaboration with NVIDIA through the Certified Instructor and University Ambassador Programs allows the members of the

TABLE 1. Online resources from the MolSSI Education Program. A superscript a (ᵃ) indicates that resources are available on the MolSSI Education website at education.molssi.org/resources. A superscript b (ᵇ) indicates resources are under development.

Topic	Resources
Programming and Software Development	- Python Data and Scripting[a] - Python Data and Scripting for Biochemistry [a] - Scientific Data Visualization using Python[a] - Best Practices in Software Development[a] - Object Oriented Programming and Design Patterns[a]
Molecular Science	- Quantum Mechanics Tools[a] - Molecular Mechanics Tools[a]
Fundamentals of HPC	- Principles of Scientific HPC[b] - RAJA and Kokkos models for abstraction and portability[b] - Slurm and Moab workload management systems and schedulers[b]
Homogeneous Parallel Programming[b]	- Message Passing Interface[b] - Shared-Memory Parallel Programming with OpenMP[b]
Heterogeneous Parallel Programming[b]	- Fundamentals of Heterogeneous Parallel Programming with CUDA C/C++[a] - A Systematic Approach to CUDA C/C++ Parallel Programming[b] - Applications of Heterogeneous Programming in Computational Molecular Sciences[b] - OneAPI: A Unified Approach to Heterogeneous Parallel Programming[b] - Modern Platforms for Heterogeneous Parallel Programming[b]

community to have free access to certified training programs that otherwise would involve a registration fee. These programs are divided into four major specialization areas: Data science, deep learning, accelerated computing, and conversational AI. Intel, on the other hand, enabled us to offer a variety of training modules within two main focus areas: Essentials of data-parallel C++ (DPC++) and basics of OpenMP offload. For our certified instructor-led hands-on workshops, university courses, and industrial training programs, we provide the registered users with access to a cloud virtual machine armed with a variety of accelerator architectures, such as CPUs, GPUs, and FPGAs. The users can access all training resources using either the command line or JupyterLab's user-friendly notebooks.

The last component of MolSSI's HPC and AI and Education initiative involves establishing best practices for data management. Due to the dire need of the scientific software community for improving the quality of the scientific data management plans according to FAIR principles[6] and to ensure the reproducibility, interoperability, and replicability of the computational research products, we have developed a public platform[5] for publishing community guidelines and best practices for all domains of CMS. We have found this platform upon our years of experience in serving and the open-source software communities, hundreds of interviews and surveys

gathered from our discovery project, expert recommendations, and best practices documents provided by the major vendors, and our collaborations with the -funded projects, such as the XPERT Network.[7]

Faculty and Curriculum Development

MolSSI's efforts in faculty and curriculum development focus on helping faculty members to incorporate resources into their classes and promoting the use of programming in the chemistry curriculum. Although MolSSI Education resources were originally designed to enable students to more effectively participate in research experiences, there have been a notable number of faculties using the MolSSI Education materials as a starting point to develop curricular resources to incorporate programming into their courses. Examples of these types of curricular innovations are highlighted in's recent ACS Symposium Series Book.[8] The examples in this book incorporate programming in a variety of different classes, ranging from general chemistry to graduate-level courses. Programming is used to analyze data, make visualizations demonstrate physical phenomena that are otherwise hard to describe, solve chemistry problems numerically, and more.

The MolSSI has partnered with faculty professional development groups, such as Enhancing Science Course

by Integrating Python (ESCIP)[a] and Psi4Education[b] to work with faculty to develop curricular resources. ESCIP is a group of faculty from the Cottrell Scholars Program, sponsored by Research Corporation for Science Advancement, who develop and share a curriculum that utilizes Python programming for chemistry, physics, and math courses. Psi4Education is the education and outreach program of the quantum chemistry software package Psi4 that uses Psi4's Python interface, to create lab activities for use across all levels of the chemistry curriculum. MolSSI supports these curricular development efforts by meeting with these faculty development groups to advise them on strategies and best practices for teaching programming. MolSSI also sponsors symposia and workshops for faculty at conferences, such as the Biennial Conference on Chemical Education, and hosts instructor training workshops to help faculty upskill so they can better teach best programming practices to their students.

In addition, MolSSI is involved in the University Ambassador Program with the NVIDIA. Through this partnership, MolSSI has access to teaching kits designed for university courses, which include syllabi, lecture notes, curricular resources, and programming activities that cover major topics in HPC, deep learning, and robotics. MolSSI can support faculty in implementing these courses at their institutions and help them customize the contents of each course in the domain to meet their program goals and requirements.

MolSSI Software Fellowship Program

MolSSI provides direct support to software development efforts of early-career researchers through its Software Fellowship program. Software Fellowships are highly selective awards that fund graduate students and postdocs who develop software infrastructure for CMS. MolSSI prioritizes projects of broad interest with potential high impact to the communities. As of 2022, the Software Fellowships are year-long awards, with calls for applications starting in February and awards starting in July.

Each cohort of software fellows receives special training in a week-long Software Fellow Bootcamp. The Bootcamp curriculum includes topics in software development best practices, software design and distribution, and special topics related to the projects and interests of the software fellow cohort. The Bootcamp material draws heavily from the Education resources featured in

this paper, particularly the Python Package Best Practices and Software Design workshops.

Throughout the fellowship, the software fellows receive one-on-one mentoring from a software scientist who provides guidance for the software fellow's project.

CONCLUSION

The MolSSI Education Team is engaged with, and welcomes comments and collaborations from members of the CMS and scientific software development communities. MolSSI disseminates all new updates and the latest releases to the educational resources through its social media accounts, newsletter, and the organizational website.[2,9] All of our educational resources, including the hands-on tutorials, instructor-led workshop materials, and self-paced online courses are free and available under open-source licenses. Members of the scientific community can provide direct feedback, comments, or contributions to all educational resources through opening discussions, issues, and pull requests on the Education repositories.[10] We continuously measure the impact of our online educational resources through tracking the user analytics gathered from the hosting websites and repositories.

ACKNOWLEDGMENTS

REFERENCES

1. MolSSI Software Projects. [Online]. Available: https://molssi.org/software-projects/
2. MolSSI Education. [Online]. Available: https://education.molssi.org/
3. MERCURY Consortium—molecular education and research consortium in undergraduate computational chemistry. [Online]. Available: https://mercuryconsortium.org/
4. A. R. Mcdonald *et al.*, "Building capacity for undergraduate education and training in computational molecular science: A collaboration between the MERCURY consortium and the Molecular Sciences Software Institute," *Int. J. Quantum Chem.*, vol. 120, no. 20, Jul. 4, 2020, doi: 10.1002/qua.26359.
5. MolSSI guidelines, checklist, and best practices. [Online]. Available: https://molssi.github.io/molssi-guidelines/
6. M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, Art. no. 160018, 2016, doi: 10.1038/sdata.2016.18.

[a]https://escip.github.io/
[b]https://psicode.org/posts/psi4education/

7. P. Barakhshan *et al.*, "Exchanging best practices and tools for supporting computational and data-intensive research," The Xpert Network, 2021, doi: 10.48550/ARXIV.2102.09373.
8. A. R. McDonald and J. Nash, "Teaching Programming Across the Chemistry Curriculum," 2021. doi: 10.1021/BK-2021-1387.
9. The Molecular Sciences Software Institute. [Online]. Available: https://molssi.org/
10. MolSSI Education GitHub. [Online]. Available: https://github.com/molssi-education
11. ESCIP | Enhancing Science Courses by Integrating Python (ESCIP) is funded by the Research Corporation for Scientific Advancement through the Cottrell Scholars Collaborative program. [Online]. Available: https://escip.github.io/
12. *Psi4Education: Computational Labs Using Free Software*. [Online]. Available: https://psicode.org/posts/psi4education/

JESSICA A. NASH is with the Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA. She is a software scientist and the education lead at MolSSI, Blacksburg, VA, 24060. Contact her at janash@vt.edu.

MOHAMMAD MOSTAFANEJAD is with the Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA. He is the lead software scientist in HPC education, AI, and ML at the MolSSI, Blacksburg, VA, 24060. Contact him at smostafanejad@vt.edu.

T. DANIEL CRAWFORD is the director of the MolSSI, Blacksburg, VA, 24060, USA, as well as the university distinguished professor and ethyl chair of the Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061. Contact him at crawdad@vt.edu.

ASHLEY RINGER MCDONALD is an associate professor with the Department of Chemistry and Biochemistry, California Polytechnic State University, San Luis Obispo, CA, 93401, USA. She serves on the board of directors of the MolSSI, Blacksburg, VA, 24060, USA, as the co-director for education, training, and faculty development. Contact her at armcdona@calpoly.edu.