DAVID GRAFF

(954) 648-9654 | 286 Cambridge St. Boston, MA 02114 | davidegraff62@gmail.com | github.com/davidegraff

EDUCATION

Harvard University 2023
Ph.D., Theoretical Chemistry
Research Advisors: Connor Coley, Eugene Shakhnovich
Thesis: Accelerating Exploration in Virtual Chemical Libraries

Princeton University 2018
A.B. *magna cum laude*, Chemistry
Certificate in Applications of Computing
Research Advisor: Robert Knowles
Thesis: Development of Two Novel Reactions for Alkene Hydroamination Enabled by Proton-Coupled Electron Transfer

RESEARCH

Postdoctoral Associate Jun 2023–
Advisor: Connor Coley
MIT Department of Chemical Engineering

- Researching novel algorithms to accelerate optimization in discrete, non-enumerated design spaces
- Adapting techniques in continuous optimization to discrete settings via hybrid continuous-discrete optimization techniques
- Exploring model-free optimization approaches to identify potent molecules in discrete molecular libraries

Graduate Research Assistant Jan 2020–May 2023
Advisors: Connor Coley and Eugene Shakhnovich
MIT Department of Chemical Engineering / Harvard Department of Chemistry and Chemical Biology, Cambridge, MA

- Researched algorithms and software for accelerated molecular discovery in large, enumerated compound spaces using machine learning, structure-based drug design, cheminformatics, and distributed computing
- Lead developer of Chemprop, a Python package for molecular property prediction using message passing neural networks (https://github.com/chemprop/chemprop)
- Investigated performance of pretrained chemical representations generated by large language models (LLMs) on QSPR benchmark tasks using ROGI-XD, a scalar metric to quantify QSPR surface that is invariant to representation dimensionality (https://github.com/coleygroup/rogi-xd)
- Established design space pruning (DSP), an extension to model-guided active learning for enumerated design spaces that reduces optimization overhead costs 50% with no impact on optimization performance
- Developed MolPAL, an algorithm for accelerated high-throughput virtual screening that uses active learning to identify computational hits from molecular docking 50–100x faster than random selection (https://github.com/coleygroup/molpal)
- Authored pyscreener, a Python library to enable programmatic calculation of molecular docking scores directly from SMILES strings (https://github.com/coleygroup/pyscreener)

Machine Learning Research Intern Jun 2022–Aug 2022
Roivant Discovery, Boston, MA

- Integrated active learning software (MolPAL) into structure-based drug design workflows, enabling both automated and efficient search of large (10M–100M) compound libraries with molecular docking
- Developed and implemented strategy for seamless, horizontal scaling of discovery programs to any number of protein targets via Google Cloud Platform (w/ Shivam Patel)

- Extended model-guided optimization workflow to molecular dynamics-based virtual screening through explicit usage of active site and docking-level information
- Utilized combination of reaction-based enumeration, QSAR/ADMET modelling, computational docking, and molecular dynamics to identify novel inhibitors for kinase and protease targets

Undergraduate Research Assistant Feb 2017–May 2018
Advisor: Robert Knowles | Collaborators: Casey Roos, Qilei Zhu
Princeton University Department of Chemistry, Princeton, NJ

- Elaborated substrate scope of novel intermolecular hydroamination reaction of olefins with sulfonamides catalyzed by proton-coupled electron transfer
- Assisted in reaction screening for enantioselective variant of olefin hydroamination reaction catalyzed by proton-coupled electron transfer (PCET)
- Developed a computational model for hydroamination enantioselectivity using physical organic descriptors of chiral phosphate base catalyst

Harvard University BCMP Summer Scholar Jun 2016–Aug 2016
Advisor: Eric Fischer | Collaborator: Radek Nowak
Harvard Medical School / Dana Farber Cancer Institute, Boston, MA

- Developed compound profiling workflow to pipeline parse raw binding fluorescence data into dose-response curves visualized through an interactive web interface
- Investigated crystallization conditions for apo crystal structure of DDB1ΔB-CRBN E3 ubiquitin ligase

PUBLICATIONS

1. Graff, D.E.; Pyzer-Knapp, E.O.; Shakhnovich, E.I.; Coley, C.W. Evaluating the Roughness of Structure-Property Relationships using Pretrained Molecular Representations. *arXiv:2305.08238* **2023**
2. Aldeghi, M.; Graff, D.E.; Frey, N.C.; Morrone, J.A.; Pyzer-Knapp, E.O.; Jordan, K.E.; Coley, C.W. Roughness of Molecular Property Landscapes and Its Impact on Modellability. *J. Chem. Inf. Model.* **2022**, *62* (19), 4660
3. Graff, D.E.; Aldeghi, M.; Morrone, J.A.; Jordan, K.E.; Pyzer-Knapp, E.O.; Coley, C.W. Self-Focusing Virtual Screening with Active Design Space Pruning. *J. Chem. Inf. Model.* **2022**, *62* (16), 3854
   ACS Editors' Choice
4. Graff, D.E.; Coley, C.W. pyscreener: A Python Wrapper for Computational Docking Software. *J. Open Source Softw.* **2022**, *7* (71), 3950
5. Graff, D.E.; Shakhnovich, E.I., Coley, C.W. Accelerating High-Throughput Virtual Screening Through μolecular Pool-Based Active Learning. *Chem. Sci.* **2021**, *12*, 7866
   Featured in State of AI Report 2021
6. Roos, C.B.; Demaerel, J.; Graff, D.E.; Knowles, R.R. Enantioselective Hydroamination of Alkenes with Sulfonamides Enabled by Proton-Coupled Electron Transfer. *J. Am. Chem. Soc.* **2020**, *142* (13), 5974
7. Zhu, Q.; Graff, D.E.; Knowles, R.R. Intermolecular Anti-Markovnikov Hydroamination of Unactivated Alkenes with Sulfonamides Enabled by Proton-Coupled Electron Transfer. *J. Am. Chem. Soc.* **2018**, *140* (2), 741.

PRESENTATIONS

Oral

1. Graff, D.E. and Coley, C.W. MolPAL: Software for Sample Efficient High-Throughput Virtual Screening NeurIPS AI4Mat, *New Orleans, LA*, December 2022
   Spotlight Talk
2. Graff, D.E.; et al. Chemprop new features and updates. MIT MLPDS Consortium Meeting, *Boston, MA*, November 2022

3. Graff, D.E. and Coley, C.W. Open-source software tools to accelerate docking-based molecular discovery. Multidisciplinary and Data-Driven Software Development, ACS National Meeting, *Chicago, IL*, August 2022
4. Graff, D.E.; Shakhnovich, E.I.; Coley, C.W Accelerating HTVS through molecular pool-based active learning. The Future of AI in Chemistry, ACS National Meeting, *Virtual*, April 2021
5. Graff, D.E.; Nowak, R. P.; Fischer, E. S. Apo crystal structure of DDB1ΔB. Harvard Medical School BSSP Research Symposium, *Boston, MA*, August 2016

Poster

1. Graff, D.E.; Aldeghi, M.; Coley, C.W. The questionable advantages of pretrained chemical models in QSPR. OpenEye CUP XXII, *Santa Fe, NM*, March 2023
2. Graff, D.E.; Aldeghi, M.; Morrone, J.A.; Jordan, K.E.; Pyzer-Knapp, E.O.; Coley, C.W. Self-focusing virtual screening with active design space pruning. MIT-IBM Watson AI Lab open house, *Boston, MA*, May 2022
3. Graff, D.E.; Zhu, Q.; Knowles, R. A Tale of Two Sulfonamides. Princeton Chemistry Summer Research Symposium, *Princeton, NJ*, August 2017

TEACHING

Harvard University
Cambridge, MA

Teaching Fellow, Foundational Chemistry and Biology (LPS A)	Fall 2020
Teaching Fellow, Experimental Inorganic Chemistry (CHEM 145)	Fall 2019
Teaching Fellow, Biochemistry (CHEM 27)	Spring 2019

Princeton University
Princeton, NJ

McGraw Center Head Tutor, Organic Chemistry (CHM 303/304/304B)	Fall 2017–Spring 2018
Undergraduate Teaching Assistant, Organic Chemistry II (CHM 304)	Spring 2018
Undergraduate Teaching Assistant, Organic Chemistry II (CHM 304B)	Spring 2017, Spring 2018
Undergraduate Teaching Assistant, Organic Chemistry I (CHM 303)	Fall 2016, Fall 2017
Peer Tutor, Algorithms and Data Structures (COS 226)	Spring 2018
Peer Tutor, Introduction to Computer Science (COS 126)	Fall 2016–Spring 2018

AWARDS

Godfrey Summer Research Award, *Princeton University* 2013

SKILLS

Technical

Machine Learning: graph neural networks, active learning, optimization, uncertainty quantification, generative models, large language models, pretrained, self-supervised learning
Chemistry: QSPR modelling, molecular docking, structure-based drug design

Software

Languages: Python (fluent), Bash, C, Java, OCaml
Frameworks: PyTorch, Tensorflow, RDKit, Ray, Dask, NumPy, SciPy, Scikit-Learn, Pandas, Matplotlib/Seaborn
Software: SLURM, Docker, AutoDock Vina, DOCK6, PyMol
Practices: HPC, CI/CD, automated deployment, packaging, containerization, testing, version control, microservices