Nathan C. Frey, PhD

nfrey213@gmail.com • (636) 634-7386 • www.linkedin.com/in/ncfrey/ • https://ncfrey.github.io/

Experience

Prescient Design, Genentech, New York, NY

Team Lead, Foundation Models for Drug Discovery, and Principal Machine Learning Scientist 2024 - Present

- Led a multidisciplinary project team of 15 scientists and engineers to conduct collaborative research on biological foundation models and antibody lead identification and optimization; managed a $1M+ R&D budget; established and executed on objectives and key results enabling the introduction of AI scientists as team leads on active drug discovery portfolio projects, scaling from 0% to >25% of the asset portfolio.
- Set research and product directions, technical roadmaps (1-2 year), and long-term (3-5 year) AI strategy for Roche and Genentech, impacting 5 major computational teams, 100+ scientists and engineers, and $50M+ annual investment in compute resources and FTE.
- Established and led the Foundation Models for Drug Discovery team, a Roche-wide effort to develop and apply 1B+ parameter large language models, novel architectures, and agentic workflows, spanning portfolio strategy, target assessment, lead molecule identification, lead optimization, pre-clinical research, and clinical trial design, impacting 10+ large (biologics) and small molecule programs.
- Filed 4 provisional patents spanning novel approaches to machine learning for drug discovery, enabling >15,000x reduction in compute costs for developing foundation models via efficient training techniques, 10-100x binding affinity optimization improvements compared to state-of-the-art experimental lab techniques, *de novo* binder design with state-of-the-art binding rates (>70%), and integrated ML systems for autonomous multi-property optimization of therapeutic molecules with 80% reduction in R&D material costs.
- Led a research group of 3 machine learning scientists and 3 graduate student interns/contractors to conduct high-impact research in ML and biology and transform the drug discovery process with AI and ML.
- Won the first ever Best Paper Award (top <0.1% of submissions) for machine learning and biology at a major machine learning conference, the ICLR 2024 Outstanding Paper Award for innovative work on generative modeling and biology; additionally received 2 main conference oral Spotlight awards (ICLR and NeurIPS, top <1.5% of submissions) and 6 conference workshop Spotlight awards (NeurIPS, ICLR, ICML, top 10% of submissions).

Senior Machine Learning Scientist 2022 - 2024

- Established $2M+ annual collaboration between NVIDIA and Genentech, resulting in acceleration of model training (>2x) and inference (>3-5x) of production models in usage across >25% of the asset portfolio.
- Led NVIDIA-Genentech joint project team on Protein Language Models, resulting in new state-of-the-art biological foundation models with 1,000x acceleration of molecular property prediction compared to traditional computational methods.
- Invented a state-of-the-art ML method for antibody binder design, yielding >95% novel (*de novo*) antibodies and 30-75% binding rates to therapeutically relevant targets.
- Co-led a working group on Iterative Design and Active Learning to establish best practices for ML-driven drug design across 3 teams at Genentech and 20+ computational scientists.
- Communicated impactful findings in 2 peer-reviewed ML conference proceedings and 10 external invited talks at international forums for machine learning and biology research.

Massachusetts Institute of Technology, Cambridge, MA 2021 - 2022

Postdoctoral Associate, Lincoln Laboratory Supercomputing Center and AI Technology

- Published the first research paper investigating scaling phenomena in neural networks for chemistry, molecular design, and molecular dynamics; built 1B+ parameter foundation models for drug discovery and showed the existence of empirical scaling laws similar to those observed in large language models.

- Developed benchmarks and tools for efficiently scaling graph neural networks and large language models for chemistry and drug design, performing 3,400+ neural network training runs on up to 424 GPUs (53 nodes) on the MIT SuperCloud cluster.
- Discovered efficient training techniques that were implemented across MIT's supercomputing resources, reducing power consumption for deep learning training by 20%.

University of Pennsylvania, Philadelphia, PA 2016 - 2021
National Defense Science & Engineering Graduate Fellow
- Performed multiscale modeling of electronic and energy materials with density functional theory, molecular dynamics, and finite element analysis; proposed simple phenomenological models supported by first-principles calculations to explain material properties.
- Developed a semi-supervised machine learning framework to predict synthesizability of novel materials and identified over 100 promising theoretical materials for synthesis.
- Deployed deep transfer learning and machine learning to predict optimal materials for information processing.
- Communicated findings in 6 first-author and 2 co-first-author publications in leading journals, and 5 national conference presentations.

DeepChem, Stanford, CA 2020 - 2022
Developer
- Developed open-source deep learning libraries for computational drug discovery.
- Expanded and standardized the MoleculeNet benchmarking suite for molecular machine learning.
- Built molecular generative models with normalizing flows for multi-objective optimization of druglike molecules.
- Refined methods for featurizing protein-ligand complexes and predicting binding energies.

Lawrence Berkeley National Laboratory, Berkeley, CA 2019 - 2020
Affiliate Scientist, Materials Project
- Built a high-throughput, automated computational workflow for studying novel materials with density functional theory.
- Identified 18 potential quantum materials with high-throughput virtual screening.
- Integrated workflow with Materials Project database, one of the largest online repositories of materials simulation data with over 560,000 users.

Education

University of Pennsylvania, Philadelphia, PA 2021
PhD Materials Science & Engineering
Thesis: "Creation and control of quantum states in layered materials"
Advisor: Vivek B. Shenoy, Director of NSF Center for Engineering Mechanobiology

Boston University, Boston, MA 2016
MA Physics

University of Missouri, Columbia, MO 2014
BSc Physics and Mathematics, *Summa Cum Laude* & Phi Beta Kappa
BSc Honors Thesis: "Automatic oligomeric state analysis of SecYEG in atomic force microscopy images"

Select Awards

- NeurIPS MLSB Oral Presentation Spotlight, *Tokenized and Continuous Embedding Compressions of Protein Sequence and Structure*, 2024
- NeurIPS AIDrugX Oral Presentation Spotlight, *LLMs are Highly-Constrained Biophysical Sequence Optimizers*, 2024
- ICML AccMLBio Oral Presentation Spotlight, *Cramming Protein Language Model Training in 24 GPU Hours*, 2024
- ICLR Outstanding Paper Award, *Protein Discovery with Discrete Walk-Jump Sampling*, 2024
- ICLR Conference Oral Presentation Spotlight, top 1.2% of all submissions, *Protein Discovery with Discrete Walk-Jump Sampling*, 2024
- NeurIPS GenBio Workshop Oral Presentation Spotlight, *Protein Discovery with Discrete Walk-Jump Sampling*, 2023
- Machine Learning for Computational Biology Oral Presentation Spotlight, *Protein Discovery with Discrete Walk-Jump Sampling*, 2023
- NeurIPS Conference Spotlight, *Protein Design with Guided Discrete Diffusion*, 2023
- ICLR Physics for Machine Learning Spotlight, *Protein Discovery with Discrete Walk-Jump Sampling*, 2023
- S.J. Stein Prize for Best PhD Dissertation on Electronic Materials, University of Pennsylvania, 2021
- Certificate of Advanced Scientific Computing, Penn Institute for Computational Science, 2021
- Merck Quantitative Biosciences Career Exploration Fellowship, University of Pennsylvania, 2020
- Geoffrey Belton Memorial Fellowship, University of Pennsylvania, 2019
- National Defense Science & Engineering Graduate Fellowship, Department of Defense, 2016
- Materials Science & Engineering Fellowship, University of Pennsylvania, 2016
- Dean's Fellowship, Boston University, 2016
- Dean's Award, Boston University, 2014
- Student Undergraduate Lab Internship, Brookhaven National Laboratory, 2014
- Howard Hughes Medical Institute Fellowship, University of Missouri, 2013
- Physics Leaders Meeting Undergraduate Research Prize, University of Missouri, 2012
- Eagle Scout, Boy Scouts of America, 2010

Conference Proceedings

1. Ismail, Aya Abdelsalam, …**Frey NC****. *Concept Bottleneck Language Models for Protein Design*. **International Conference on Learning Representations** (ICLR, 2025).
2. **Frey NC*,** Berenberg D*, Zadorozhny K, Kleinhenz J, Lafrance-Vanasse J, Hotzel I, Wu Y, Ra S, Bonneau R, Cho K, Loukas A. *Protein Discovery with Discrete Walk-Jump Sampling***. International Conference on Learning Representations** (ICLR, 2024).
3. Gruver N*, Stanton S*, **Frey NC**, Rudner TG, Hotzel I, Lafrance-Vanasse J, Rajpal A, Cho K, Wilson AG. *Protein Design with Guided Discrete Diffusion*. **Conference on Neural Information Processing Systems** (NeurIPS, 2023).
4. **Frey NC**, Li B, McDonald J, Zhao D, Jones M, Bestor D, Tiwari D, Gadepally V, Samsi S. *Benchmarking resource usage for efficient distributed deep learning*. **IEEE High Performance Extreme Computing Conference** (HPEC, 2022).

* Denotes equal contribution

** Denotes senior authorship

Nathan C. Frey, PhD

Select Journal Publications

1. Lu AX, Yan W, Yang KK, Gligorijevic V, Cho K, Abbeel P, Bonneau R, **Frey NC****. *Tokenized and Continuous Embedding Compressions of Protein Sequence and Structure*. **Cell Patterns** (2025).
2. **Frey NC**, Soklaski R, Axelrod S, Samsi S, Gomez-Bombarelli R, Coley C, Gadepally V. *Neural scaling of deep chemical models*. **Nature Machine Intelligence** (2023).
3. Price CC, Singh A, **Frey NC**, Shenoy VB. *Efficient catalyst screening using graph neural networks to predict strain effects on adsorption energy*. **Science Advances** (2022).
4. Krenn M, Ai Q, Barthel S, Carson N, Frei A, **Frey NC**, Friederich P, Gaudin T, Gayle AA, Jablonka KM, Lameiro RF, …Aspuru-Guzik A. *SELFIES and the future of molecular string representations*. **Cell Patterns** (2022).
5. Aldeghi M, Graff DE, **Frey NC**, Morrone JA, Pyzer-Knapp EO, Jordan KE, Coley CW. *Roughness of molecular property landscapes and its impact on modellability*. **Journal of Chemical Information and Modeling** (2022).
6. **Frey NC**, MK Horton, JM Munro, SM Griffin, KA Persson, VB Shenoy. *High-throughput search for magnetic and topological order in transition metal oxides*; **Science Advances** (2020).
7. **Frey NC**, D Akinwande, D Jariwala, VB Shenoy. *Machine Learning-Enabled Design of Point Defects in 2D Materials for Quantum and Neuromorphic Information Processing*; **ACS Nano** (2020).
8. **Frey NC**, J Wang, GIV Bellido, B Anasori, Y Gogotsi, VB Shenoy. *Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning*; **ACS Nano** (2019).

Select Conference Workshop Papers

1. Lu, A.X., Yan, W., Yang, K.K., Gligorijevic, V., Cho, K., Abbeel, P., Bonneau, R. and **Frey, NC****, *Tokenized and Continuous Embedding Compressions of Protein Sequence and Structure*. bioRxiv, pp.2024-08. NeurIPS 2024 MLSB Workshop (2024).
2. Chen A, Stanton S, Alberstein R, Watkins A, Bonneau R, Gligorijevic V, Cho K, **Frey NC****, *LLMs are Highly-Constrained Biophysical Sequence Optimizers*. NeurIPS 2024 AIDrugX Workshop (2024).
3. **Frey NC**, Joren T, Ismail A, Goodman A, Bonneau R, Cho K, Gligorijevic V. *Cramming Protein Language Model Training in 24 GPU Hours*. bioRxiv. 2024:2024-05. ICML 2024 AccMLBio Workshop (2024).
4. Adebayo J, Stanton S, Kelow S, Bonneau R, Gligorijevic V, Cho K, Ra S, **Frey NC****. *Identifying regularization schemes that make feature attributions faithful*. NeurIPS 2023 Workshop on New Frontiers of AI for Drug Discovery and Development (2023).
5. Lee JH, Yadollahpour P, Watkins A, **Frey NC**, Leaver-Fay A, Ra S, Cho K, Gligorijević V, Regev A, Bonneau R. *Equifold: Protein structure prediction with a novel coarse-grained structure representation*, Machine Learning for Structural Biology Workshop, NeurIPS (2022).
6. **Frey NC**, Gadepally V, Ramsundar B. Fastflows: *Flow-based models for molecular graph generation*. ELLIS Machine Learning for Molecule Discovery Workshop (2022).
7. **Frey NC**, S Samsi, J McDonald, L Li, CW Coley, V Gadepally**.** *Scalable Geometric Deep Learning on Molecular Graphs*; NeurIPS 2021 AI for Science Workshop (2021).
8. **Frey NC**, S Samsi, B Ramsundar, CW Coley, V Gadepally. *Bringing Atomistic Deep Learning to Prime Time*; NeurIPS 2021 AI for Science Workshop (2021).

Invited Talks

- *Scalable Active Learning for Therapeutic Antibody Design*, PEGS Europe, Barcelona, Spain, November 2024.

- *Cramming Protein Language Model Training in 24 GPU Hours*, Oral Spotlight Talk, AccMLBio Workshop, ICML, Vienna, Austria, July 2024.
- *Protein Discovery with Discrete Walk Jump Sampling*, Outstanding Paper and Oral Spotlight Talk, ICLR 2024, Vienna, Austria, May 2024.
- *"Lab-in-the-loop" drug discovery and design*, PepTalk, San Diego, CA, January 2024.
- *Protein Discovery with Discrete Walk Jump Sampling*, Spotlight Talk, GenBio Workshop, NeurIPS, New Orleans, LA, December 2023.
- *Protein Discovery with Discrete Walk Jump Sampling*, Machine Learning for Computational Biology Workshop, Seattle, WA, December 2023.
- *"Lab-in-the-loop" drug discovery and design*, PEGS Europe, Lisbon, Portugal, November 2023.
- *Protein Discovery with Discrete Walk Jump Sampling*, ML for Protein Engineering Seminar Series, September 2023.
- *Walk-Jump Sampler: the Mallet of Prescient Design*, Prescient Design/LMDD Fall Workshop, South San Francisco, CA, September 2023.
- *"Lab-in-the-loop" drug discovery and design*, Accelerate Conference, University of Toronto, Toronto, CA, August 2023.
- *"Lab-in-the-loop" drug discovery and design*, LightningAI, New York, NY, April 2023.
- *Learning protein family manifolds with smoothed energy-based models*, Spotlight talk, Physics4ML Workshop, International Conference on Learning Representations (ICLR), April 2023.
- *Neural Scaling of Deep Chemical Models*, Fordham University, New York, NY, November 2022.
- *Neural Scaling of Deep Chemical Models*, Pre-training Workshop, International Conference on Machine Learning (ICML), July 2022.
- *Scalable Geometric Deep Learning on Molecular Graphs*, Molecular Modeling and Drug Discovery Symposium, Valence Discovery and Mila - Quebec AI Institute, February 2022.
- *Integrating High-Throughput Computing and Machine-Learning Enabled Atomistic Design*, Air Force Research Lab, November 2021.
- *Machine Learning in Materials Science and Chemistry*, US Patent and Trademark Office, Alexandria, VA, February 2020.

Contributed Talks and Poster Presentations

- *Molecular Discovery and Design with Machine Learning* (poster), Genentech Group Leader Offsite, Napa, CA, June 2023.
- *Machine Learning-Enabled Design of Point Defects in 2D Materials,* American Physical Society March Meeting, March 2021.
- *Flow-Based Models for Active Molecular Graph Generation* (virtual poster), Machine Learning for Molecules Workshop, NeurIPS, December 2020.
- *High-throughput Search for Magnetic and Topological Order in Transition Metal Oxides*, American Physical Society March Meeting, March 2020.
- *Engineering Zero-Dimensional Quantum Confinement in Transition Metal Dichalcogenide Heterostructures*, Society of Engineering Science Technical Meeting, St. Louis, MO, October 2019.
- *Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning*, Materials Research Society Spring Meeting, Phoenix, AZ, April 2019.

Professional Activities & Service

- Peer reviewer for NeurIPS, ICML, ICLR, TMLR, Nature Biotechnology, Nature Communications
- Area chair for Generative and Experimental Perspectives for Biomolecular Design ICLR Workshop