Jason **Yang**

jyang4@caltech.edu | jsunn-y.github.io | jsunn-y | @jsunn_y

OBJECTIVE: Dedicated and highly motivated PhD candidate developing new machine learning-assisted workflows for protein engineering.

Education

California Institute of Technology
Pasadena, CA

PH.D. IN CHEMICAL ENGINEERING — *2021 - 2026*

GPA: 4.0/4.0 | Advisors: Frances Arnold, Yisong Yue

Yale University
New Haven, CT

B.S. IN CHEMICAL ENGINEERING | B.S. IN APPLIED MATH — *2017- 2021*

GPA: 3.93/4.00

New Trier High School
Winnetka, IL

HIGH SCHOOL DIPLOMA — *2013- 2017*

GPA: 5.33/5.33

Honors & Awards

2024 - 2026 Google PhD Fellowship
2021 - 2024 NSF Graduate Research Fellowship
2020 - 2021 Barry Goldwater Scholarship
2019 Tau Beta Pi Engineering Honor Society
2018 First-Year Summer Research Fellowship, Yale University
2017 Regeneron Science Talent Search Top Scholar
2017 National Merit Scholarship Finalist
2017 Congressional Award for Youth Gold Medal
2017 Science Olympiad National Medalist

Experience

Graduate Research at Caltech
Pasadena, CA

ADVISORS: PROF FRANCES ARNOLD, PROF YISONG YUE — *2021 - present*

- Machine learning-assisted workflows to accelerate protein engineering
- Diffusion models for protein sequence generation and optimization
- Protein language models for enzyme generation
- Enzyme retrieval and annotation using bioinformatics and multimodal models
- Uncertainty quantification and Bayesian optimization with deep kernels for protein optimization
- Representation learning using physics-informed features and graph neural networks

Summer PhD Intern at Profluent
Berkeley, CA

ADVISORS: DR. JEFFREY RUFFOLO, AADYOT BHATNAGAR, DR. ALI MADANI — *2024*

- Developed a method for conditional generation of enzymes using protein language models

Undergraduate Research at Yale

New Haven, CT

Advisors: Prof Menachem Elimelech, Prof Ying Li, Prof Smita Krishnaswamy *2018 - 2021*

- Implemented machine learning models to predict the permeability and selectivity of polymeric gas separation membranes
- Evaluated the parameters controlling water-organic separations in nanoporous graphene membranes using molecular dynamics simulations
- Investigated cation-specific selectivity in membranes by synthesizing polyelectrolytes and incorporating them into novel nanofiltration membranes using layer-by-layer assembly
- Studied the mechanism of ecotoxicology and environmental impact of MnO2 nanomaterials using dynamic light scattering and experimental assays

Undergraduate DoE National Lab Summer Intern

Golden, CO

National Renewable Energy Lab (NREL) | Advisor: Dr. Heather Mayes *2020*

- Implemented machine learning models using features from molecular dynamics simulations to accurately predict mechanical properties of bio-based polymers

Undergraduate Pharma Technical Development Summer Intern

South San Francisco, CA

Genentech | Advisor: Steve Russell *2019*

- Automated an ICP-MS trace-metals assay for greater throughput by developing software and hardware

High School Summer Research Intern

Evanston, IL

Northwestern | Advisor: Prof Teri Odom *2016*

- Developed a cost-effective nanoscale patterning method of polyolefin thin films for enhanced hydrophobicity

Publications

1. **Yang, J.***; Chu, W.*; Khalil, D.; Astudillo, R.; Wittmann, B.J.; Arnold, F. H.; Yue, Y.; Steering Generative Models with Experimental Data for Protein Fitness Optimization. *Under Review*.
2. **Yang, J.**; Li, F.-Z.; Long, Y.; Arnold, F. H. Illuminating the Universe of Enzyme Catalysis in the Era of Artificial Intelligence. *Under Review*.
3. **Yang, J.***; Lal, R. G.*; Bowden, J. C.; Astudillo, R.; Hameedi, M. A.; Kaur, S.; Hill, M.; Yue, Y.; Arnold, F. H. Active Learning-Assisted Directed Evolution. Nat Commun 2025, 16 (1), 714.
4. **Yang, J.**; Mora, A.; Liu, S.; Wittmann, B. J.; Anandkumar, A.; Arnold, F. H.; Yue, Y. CARE: A Benchmark Suite for the Classification and Retrieval of Enzymes. NeurIPS Datasets and Benchmarks 2024.
5. Li, F.-Z.; **Yang, J.**; Johnston, K. E.; Guersoy, E.; Yue, Y.; Arnold, F. H. Evaluation of Machine Learning-Assisted Directed Evolution Across Diverse Combinatorial Landscapes. bioRxiv 2024.
6. **Yang, J.**; Bhatnagar, A.; Ruffolo, J. A.; Madani, A. Conditional Enzyme Generation Using Protein Language Models with Adapters. arXiv 2024.
7. Johnston, K. E.; Almhjell, P. J.; Watkins-Dulaney, E. J.; Liu, G.; Porter, N. J.; **Yang, J.**; Arnold, F. H. A Combinatorially Complete Epistatic Fitness Landscape in an Enzyme Active Site. Proc. Natl. Acad. Sci. 2024, 121 (32), e2400439121.
8. **Yang, J.**; Li, F.-Z.; Arnold, F. H. Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering. ACS Cent. Sci. 2024, 10 (2), 226–241.

9. **Yang, J.**; Ducharme, J.; Johnston, K. E.; Li, F.-Z.; Yue, Y.; Arnold, F. H. DeCOIL: Optimization of Degenerate Codon Libraries for Machine Learning-Assisted Protein Engineering. ACS Synth. Biol. 2023, acssynbio.3c00301.
10. He, J.; **Yang, J.**; McCutcheon, J. R.; Li, Y. Molecular Insights into the Structure-Property Relationships of 3D Printed Polyamide Reverse-Osmosis Membrane for Desalination. Journal of Membrane Science 2022, 658, 120731.
11. **Yang, J.***; Tao, L.*; He, J.*; McCutcheon, J. R.; Li, Y. Machine Learning Enables Interpretable Discovery of Innovative Polymers for Gas Separation Membranes. Sci. Adv. 2022, 8 (29), eabn9545.
12. DuChanois, R. M.; Heiranian, M.; **Yang, J.**; Porter, C. J.; Li, Q.; Zhang, X.; Verduzco, R.; Elimelech, M. Designing Polymeric Membranes with Coordination Chemistry for High-Precision Ion Separations. Sci. Adv. 2022, 8 (9), eabm9436.
13. **Yang, J.**; Shen, Z.; He, J.; Li, Y. Efficient Separation of Small Organic Contaminants in Water Using Functionalized Nanoporous Graphene Membranes: Insights from Molecular Dynamics Simulations. Journal of Membrane Science 2021, 630, 119331.
14. Zucker, I.; Hashmi, S. M.; **Yang, J.**; He, Y.; Pfefferle, L. D.; Elimelech, M. Shape-Dependent Interactions of Manganese Oxide Nanomaterials with Lipid Bilayer Vesicles. Langmuir 2019, 35 (43), 13958–13966.

Presentations

Aug 2025	**BioML Symposium**, Oral Presentation	*Prague, CZ*
Apr 2025	**GEM Workshop at ICLR**, Poster	*Singapore, SG*
Apr 2025	**Stanford AI for Science Workshop**, Poster	*Stanford, CA*
Mar 2025	**Caltech AI in Biology Symposium**, Poster	*Pasadena, CA*
Dec 2024	**NeurIPS Datasets & Benchmarks Main Conference**, Poster	*Vancouver, BC*
Dec 2024	**Machine Learning for Structural Biology Workshop at NeurIPS**, Poster	*Vancouver, BC*
Sep 2024	**Debbie Marks Lab Reading Group**, Oral Presentation	*Virtual*
Jul 2024	**Machine Learning for Protein Engineering Seminar Series**, Oral Presentation	*Virtual*
Jul 2024	**Caltech ML for Chemistry Seminar Series**, Oral Presentation	*Pasadena, CA*
Jun 2024	**Simons Institute Workshop on AI for Science**, Oral Presentation	*Berkeley, CA*
Oct 2023	**Caltech CCE Seminar Day**, Oral Presentation Winner	*Pasadena, CA*

Teaching & Mentorship

Mentorship

Pasadena, CA

Caltech

Spring 2023 - Present

- Mentored five undergraduate students for research during the summer and school year
- Mentored three graduate rotation students
- Mentored four visiting masters students
- Prepared onboarding guides and collections of resources

Graduate Teaching Assistant

Pasadena, CA

Caltech

Fall 2023 - Present

- BE/ChE163: Introduction to Biomolecular Engineering (instructor Steve Mayo)
- ChE165: Graduate Level Chemical Thermodynamics (instructor Zhen-Gang Wang)
- Duties: Delivering selected lectures, writing new problem set questions, office hours, grading

Undergraduate Teaching Assistant

New Haven, CT

Yale University

Fall 2019 - Spring 2021

- CENG 150: Intro to Engineering Analysis (instructor Michael Loewenberg)
- CHEM 333b: Physical Chemistry with Applications in the Physical Sciences II (instructor Patrick Vaccaro)

Service & Outreach

Academic and Professional Service

Reviewing

- ACS Central Science, Briefings in Bioinformatics, ACS Catalysis, ACS Synthetic Biology, MLSB NeurIPS Workshop, GEM ICLR Workshop

ICLR Volunteer Chair *2025*

- Recruited and assigned duties for volunteers at the ICLR conference to promote diversity and inclusion.

Outreach

Co-organizer for the ML for Protein Engineering Seminar Series *Fall 2024 - Present*

- Invited speakers, scheduled seminars, and managed social media for a community of nearly 3000 members worldwide

Science Olympiad Volunteer *Spring 2024 - Present*

- Organized lab tours and science activities for middle schoolers and high schoolers at the Socal State Science Olympiad tournament

Coursework

Graduate Studies

Pasadena, CA

Caltech

2021 -2026

- **Chemical Engineering Coursework:** Kinetics, Thermodynamics, Statistical Mechanics, Transport
- **Other Coursework:** Machine Learning, Advanced Topics in Machine Learning, Deep Learning for Biology, Bioinformatics, Biomolecular Engineering, Enzymology and Biochemistry, Computational Tools for Metagenomics

Undergraduate Studies

New Haven, CT

Yale University

2017-2021

- **Chemical Engineering Coursework:** Organic Chemistry, Physical Chemistry, Materials Science, Programming for Engineers, Fluid Mechanics, Thermodynamics, Transport Phenomena, Chemical Kinetics and Reactors, Separations and Purifications
- **Applied Math Coursework:** Vector Calculus, Linear Algebra, Differential Equations, Data Analysis, Probability Theory, Applied Numerical Methods for Differential Equations, Statistical Methods for Science, Machine Learning, Discrete Math
- **Elective Courses:** Environmental Physiochemical Processes, Computational Chemistry, Inorganic Chemistry, Solid State Physics, Biochemistry and Cell Biology
- **Other:** Tokyo, Fiction and Consciousness, Graphic Design, Intro to Psychology, Healthcare Economics and Public Policy, European Intellectual History, Advanced Culture and Conversation in French

Skills

Machine Learning

transformers, diffusion models, generative models, Graph NNs, deep kernels, contrastive learning, self-supervision, VAEs, GANs, CNNs, RNNs

Python

pytorch, transformers, tensorflow, gpytorch, gotorch, scikit, xgboost, mumpy, pandas, seaborn, matplotlib, bokeh, holoviews, biopython

Lab Skills

methods in directed evolution of proteins, cloning, transformation, screening, HPLC, TLC, TOC, ICP-MS, ion chromatography, FTIR, UV-VIS, NMR, SEM, AFM, dynamic light scattering, spin coating, ellipsometry, organic synthesis, nanoscale surface modification of polymer materials

Other

graphic design, piano, advanced Mandarin Chinese and French, Chinese calligraphy