

RESEARCH INTERESTS	<p>I am interested how machine learning models can generate hypotheses to probe and understand biology, in particular</p> <ul style="list-style-type: none"> • Generative models of protein sequence and structure applied to protein engineering, in particular motif scaffolding • Navigating joints, conditionals, and marginals of $p(\text{sequence, structure, function})$
PUBLICATIONS	<p>Lu T, Liu M, Chen Y, Kim J, Huang PS. Assessing generative model coverage of protein structures with SHAPES. <i>Cell Systems</i>. 2025;16(8):101347.</p> <p>Chu AE, Lu T, Huang PS. Sparks of function by de novo protein design. <i>Nature Biotechnology</i>. 2024 Feb;42(2):203-15.</p> <p>Menard J, Diep P, Sheikh F, Escobar A, Dykstra C, Sajtovich VA, Ahmadi A, Bodyreva E, Boucinha A, Chandrasekharan S, Duan J, Emond C, Lu T, McLean I, Morse L, Serra D, Stancescu A, Suresh S, Ingalls BP. Synthetic biology education and pedagogy: a review of evolving practices in a growing discipline. In <i>Frontiers in Education</i> 2024 Oct 7 Vol. 9, p. 1441720.</p> <p>Powers AS, Yu HH, Suriana P, Koodli RV, Lu T, Paggi JM, Dror RO. Geometric deep learning for structure-based ligand design. <i>ACS Central Science</i>. 2023 Nov 17;9(12):2257-67.</p> <p>Strokach A, Lu T, and Kim PM. ELASPIC2 (EL2): Combining Contextualized Language Models and Graph Neural Networks to Predict Effects of Mutations. <i>Journal of Molecular Biology</i> (2021): 166810.</p>
CONFERENCES & WORKSHOPS	<p>Lu T, Shuai R, Kouba P, Li Z, Chen Y, Shirali A, Kim J, Huang PS. Conditional protein structure generation with protpardelle-1c. <i>bioRxiv</i>. 2025 Aug 18:2025-08. (Accepted at MLSB 2025)</p> <p>Shuai RW, Lu T, Bhatti S, Kouba P, Huang PS. Ensemble-conditioned protein sequence design with Caliby. <i>bioRxiv</i>. 2025 Oct 2:2025-09. (Accepted at MLSB 2025)</p> <p>Boucinha A, Kell B, Sheikh F, Diep P, Yeung A, Escobar A, Emond CA, Pierce C, Siddhartha K, Chang L, Sadatmousavi P, Stephens S, Lu T, Sajtovich VA. A framework towards transdisciplinary synthetic biology curricula for heterogeneous undergraduate cohorts. <i>Canadian Engineering Education Association Conference</i> (2021).</p>
PREPRINTS	<p>Chen Y, Lu T, Zhao C, Wayment-Steele HK, Huang P. SLAE: Strictly Local All-atom Environment for Protein Representation. <i>bioRxiv</i>. 2025:2025-10.</p> <p>Miao Y, Pacesa M, Georgeon S, Schmidt J, Lu T, Huang P, Correia B. Leveraging protein representations to explore uncharted fold spaces with generative models. <i>bioRxiv</i>. 2025:2025-10.</p> <p>Powers AS, Lu T, Koodli RV, Xu M, Gu S, Karelina M, Dror RO. MedSAGE: Bridging Generative AI and Medicinal Chemistry for Structure-Based Design of Small Molecule Drugs. <i>bioRxiv</i>. 2025:2025-05.</p> <p>Lu T, Lu AX, and Moses AM. Random Embeddings and Linear Regression can Predict Protein Function. <i>arXiv preprint arXiv:2104.14661</i> (2021).</p> <p>Lu T, and Silva A. dynUGENE: an R package for uncertainty-aware gene regulatory network inference, simulation, and visualization. <i>bioRxiv</i> (2021).</p>
TEACHING	<p>BIOE 301E: Computational Protein Modeling Laboratory – Head TA 2023 & 2025</p>
EXPERIENCE	<p>Stanford University - Protein Design Lab</p> <p>Ph.D. Candidate 09/2022 – Present</p> <p>Advisor: Possu Huang</p> <ul style="list-style-type: none"> • Designing <i>de novo</i> lanthanide-binding proteins with conditional diffusion models of protein structures • Designed and experimentally characterized novel LOV domain dimers for light-controllable protein binding using customized AlphaFold2 hallucination <p>ProteinQure</p> <p>Junior Machine Learning Scientist 05/2021 – 08/2022</p>

- Designed an embedding-based kernel for Gaussian process regression for non-canonical helical and cyclic peptide property prediction. Experimentally validated EC50 and IC50 prediction on various GPCRs and other clinically-relevant targets.
- Modeled 3D atom density preferences with 3D-CNNs for a non-canonical amino acid energy function.
- Built sequence, structure, and surface-based property predictors for nanobody design: immunogenicity, solubility, non-specificity, aggregation propensity, target-agnostic binding score.
- Built combinatorial libraries of nanobodies with genetic algorithm using neural network priors on p(sequence) and p(structure)

University of Toronto

Research Student

09/2020 – 04/2021

Advisor: Alan Moses

- Modeled gene regulatory network dynamics using neural networks mixed with ODEs.
- Designed gene regulatory networks *de novo* using automatic differentiation through ODE solvers.
- Established baseline models and datasets protein language model pre-training tasks.

University of Toronto

Research Student

08/2019 – 09/2020

Advisors: Philip Kim, Pedro Alberto Valiente Flores

- Designed novel protein folds with generative models of protein structures (Transformer, GAN).
- Designed oncoprotein inhibitors using Rosetta and molecular dynamics simulations.
- Implemented code to search the PDB for protein surfaces that mimic DNA.

iGEM Toronto

Co-President

11/2020 – 08/2022

Drylab Lead

04/2020 – 11/2020

Drylab Member

04/2019 – 04/2020

Advisor: Radhakrishnan Mahadevan

- Contributed ML-guided designed plastic-degrading enzyme to [BioBrick](#).
- Implemented active learning methods for an iterative drylab-wetlab feedback loop.
- Designed a plastic-degrading enzyme using model-based optimization.
- Quantified benefits of learned protein sequence embeddings on protein function prediction.
- Analyzed PET catalysis dynamics with molecular dynamics simulations.

Canadian Synthetic Biology Education Research Group

Lead Instructor

10/2021 – 06/2022

Machine Learning Instructor

09/2019 – 10/2021

Advisors: Patrick Diep, Brayden Kell

- Created code resource on machine learning for [protein design](#) and [systems biology](#).
- Presented seven hours of content, covering both classical (Docking, Rosetta, MD) and recent methods (Sequence-to-function models, generative models, representation learning, active learning).
- Led meetings on curriculum development based on pedagogical principles.

EDUCATION

Stanford University

Ph.D. Bioengineering Year 4

09/2022 – Present

University of Toronto

B.Sc. Bioinformatics and Computational Biology, Computer Science
cGPA: 3.94/4.00

09/2018 – 06/2021

McGill University

Biological, Biomedical and Life Sciences
cGPA: 4.00/4.00

09/2017 – 05/2018

SKILLS

Programming Python, PyTorch, NumPy, Bash

Tools PyMOL, VMD, GROMACS, RosettaScripts, Unix, \LaTeX

TALKS	<ol style="list-style-type: none"> 1. De Novo Design of Metal-Binding Proteins for Critical Mineral Recovery, <i>Synthetic Biology for Sustainability Symposium</i>, Stanford, CA, 05/25 2. Assessing generative model coverage of protein structures with SHAPES, <i>BioML @ Berkeley Seminar</i>, Berkeley, CA, 04/25 3. Design of light-driven protein oligomerization, <i>RosettaCON 2023</i>, Leavenworth, WA, 08/23 4. Accelerating Plastic Recycling with PETase, <i>iGEM Grand Jamboree</i>, Boston, MA, 10/2019. 5. Recurrent Neural Networks for Protein Design, <i>Ontario iGEM Conference</i>, Guelph, ON, 07/2019.
SERVICE	<ul style="list-style-type: none"> • Co-organizer of Machine Learning for Protein Engineering seminar series 09/2022 – 09/2024 • Reviewer for Nature Machine Intelligence, Bioinformatics, ACS Synthetic Biology, NeurIPS Machine Learning in Structural Biology Workshop (2023, 2024), MLSB 2025.
AWARDS & GRANTS	<ul style="list-style-type: none"> • Stanford Synthetic Biology for Sustainability Grant (\$200,000) 2025 • Bioengineering TA Award - BIOE 301E 2023 • Stanford Graduate Fellowship 2022 - 2025 • Skule Conference Grant (\$2000) 2022 • iGEM Impact Grant (\$3000) 2022 • Skule Endowment Fund (\$5000) 2021 • iGEM Impact Grant (\$3000) 2021 • Institute of Medical Science Grant (\$2000) 2021 • Dean's Student Initiative Fund (\$1000) 2021 • NSERC Undergraduate Research Award (\$6000) 2020