Alan N. Amin

alanamin@nyu.edu

EDUCATION AND POSITIONS

Faculty Fellow / Assistant Professor 2023-Now
Courant Institute, New York University
Host: Andrew Gordon Wilson

Postdoctoral scientist Summer 2023
Jura Bio

PhD candidate 2019-2023
Harvard University Systems Biology Program
Thesis: Nonparametric Methods for Building and Evaluating Models of Biological Sequences
Advisor: Debora S. Marks
Fellowship: NSERC PGS-D ($63000 over 3 years)

Undergraduate student 2015-2019
University of Toronto
Specialist in Biochemistry, Major in Mathematics
Advanced coursework: measure theory, operator theory, abstract algebra

RESEARCH INTERESTS

I develop mathematically principled machine learning methods for biological sequence modeling, from theoretical infrastructure (kernels, diffusion, testing) to applications that unlock massive datasets in immunology, human genetics, and therapeutic design.

PUBLICATIONS

(*Equal contribution)

Peer-reviewed

1. **Amin A N**, Marks D S*, Weinstein E N*. Kernels with Guaranteed Flexibility for Reliable Machine Learning on Biological Sequences. *JMLR*, 2025.
   →Student Paper Research Award at New England Statistics Symposium 2023

2. **Amin A N**, Gruver N, Wilson A G. Why Masking Diffusion Works: Condition on the Jump Schedule for Improved Discrete Diffusion, *NeurIPS*, 2025.

3. **Amin A N***, Potapczynski A*, Wilson A G. Training Flexible Models of Genetic Variant Effects from Functional Annotations using Accelerated Linear Algebra. *ICML*, 2025.
   →Oral and 2nd best paper award at AI4NA ICLR 2025 workshop. Oral at Machine Learning in Computational Biology meeting 2025

4. **Amin A N**, Gruver N*, Kuang Y*, Li L*, Elliott H, McCarter C, Raghu A, Greenside P, Wilson A G. Bayesian Optimization of Antibodies Informed by a Generative Model of Evolving Sequences. *ICLR*, 2025.
   →Spotlight at ICLR 2025. Spotlight and outstanding poster award at AIDrugX Neurips 2024 workshop

5. **Amin A N**, Wilson A G. Scalable and Flexible Causal Discovery with an Efficient Test for Adjacency. *ICML*, 2024

6. Glaser P, Paul S, Hummer A M, Deane C M, Marks D S, **Amin A N.** Kernel-Based Evaluation of Conditional Biological Sequence Models. *ICML*, 2024

7. **Amin A N**, Weinstein E N*, Marks D S*. A Kernelized Stein Discrepancy for Biological Sequences. *ICML*, 2023

8. Weinstein E N*, **Amin A N***, Frazer J, Marks D S. Non-identifiability and the blessings of misspecification in models of molecular fitness and phylogeny. *NeurIPS*, 2022
   →Oral at NeurIPS 2022

9. Weinstein E N, **Amin A N**, Grathwohl W, Kassler D, Disset J, Marks D S. Optimal design of stochastic DNA synthesis protocols based on generative sequence models, *AISTATS*, 2022

10. **Amin A N*,** Weinstein E N*, Marks D S. A generative nonparametric Bayesian model for whole genomes, *NeurIPS*, 2021.

11. **Amin A N**, Lin Y-H, Das S, Chan H S. "Theory for a Sequence-Specific "Fuzzy" Binding Mechanism Between a Pair of Intrinsically Disordered Proteins", *J Phys Chem B*, 2020.

12. Das S, **Amin A N**, Lin Y-H, Chan H S. "Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters." *Phys. Chem. Chem. Phys.* 2018.

13. Delplace V, Ortin-Martinex A, Tsai E L S, **Amin A N**, Wallace V and Shoichet M S. "Controlled Release Strategy Designed for Intravitreal Protein Delivery to the Retina." *J. Control. Release* 2018**.**

Preprints

14. Weinstein E N*, Gollub M G*, Slabodkin A*, Gardner C L, Dobbs K, Cui X-B, **Amin A N,** Church G M, Wood E B. Manufacturing-Aware Generative Model Architectures Enable Biological Sequence Design and Synthesis at Petascale. *Preprint*, 2024.
   →Top 4 paper award at MoML @ MIT 2024

Workshop papers

15. Chandra N A*, **Amin A N***, Wilson A G. A Unification of Discrete, Gaussian, and Simplicial Diffusion. *Fronteirs of Probabilistic Inference workshop at Neurips*, 2025.

16. Baron E*, **Amin A N***, Weitzman R, Marks D S, Wilson A G. A Diffusion Model to Shrink Proteins While Maintaining their Function. *GenBio ICML*, 2025.
   →Spotlight at GenBio ICML 2025 workshop, Best paper at ExAIT ICML 2025 Workshop, Pitch award at GenBio ICLR 2025 workshop. Best paper award at MoML @ MIT 2024

17. Ober S W, McCarter C, Raghu A, Li Y L, **Amin A N**, Wilson A G, Elliott H. Is Sequence Information All You Need for Bayesian Optimization of Antibodies? *NeurIPS AI4Science workshop*, 2025

18. Berenberg D, Gruver N, **Amin A N**, Groth P M, Chen L, Srivastava H R, Notin P, Marks D S, Wilson A G, Cho K, Bonneau R. Residue-level text conditioning for protein language model mutation effect prediction. *ICLR GEM workshop*, 2025.

19. Shaw A, Shin J-E, Thadani N N, **Amin A N**, Marks D S. Designing Proteins using Sparse Data. *Learning Meaningful Representations of life workshop at NeurIPS,* 2022.

ACADEMIC AWARDS

MoML @ MIT best paper award	$1526	2025
ICML Exploration in AI today best paper	$500	2025
ICLR AI for Nucleic acids 2nd best paper	$1200	2025
ICLR Generative biology workshop pitch	$1000	2025
Student Paper Research Award at New England Statistics Symposium	$300	2023
NSERC Postgraduate Scholarships Doctoral program	$63'000	2022
Princeton International Internship	$Lodging	2018
NSERC Undergraduate Student Research Award	$4500	2017
Various undergraduate scholarships	$4890	2015-2019

PRESENTATIONS

Invited talks

ML for Protein engineering early career seminar	2025
Gatsby Institute, Machine Learning seminar	2025
Fable Therapeutics, seminar	2025
Oxford university, Deane lab meeting	2025
NYU Institute for Systems Genetics, nanoseminar	2025
ML for Protein engineering seminar	2025
Columbia University, AlQuraishi lab meeting	2024
NYU AI school	2024
NYU Centre for Data Science, seminar	2024
New York Genome Center, lab meeting	2023
Cold Spring Harbour, seminar	2023
Gatsby Institute, Machine Learning seminar	2023
Harvard QBio institute, lab meeting	2023
Harvard Systems Biology department, pizza talk	2023
MIT Readstat reading group	2023
Harvard Systems biology program, mini symposium	2023
Harvard Statistics Department, Stat 300 Seminar series	2023
Broad institute, Models, Inference and Algorithms Talks	2021

Conference talks

ICML Exploration in AI workshop oral 2025
ICLR AI for Nucleic Acids workshop oral 2025
NeurIPS oral 2022
NeurIPS Learning Meaningful Representations of Life Workshop oral 2021

TEACHING

Sole Instructor

CSC102 Data Structures Four semesters 2024-2025
New York University

Teaching Assistant

CSC102 Data Structures Fall 2023
New York University
BCMP230 Principles and practice of drug development Spring 2022
Harvard University
MAT224 Linear Algebra Spring 2019
University of Toronto
MAT135-136 Calculus Academic year 2018
University of Toronto
MAT137 Calculus Academic year 2017
University of Toronto

SERVICE

Reviewer at ICLR, ICML, NeurIPS, AAAI, Workshops 2022-Now
Reviewer at eLife 2024
Harvard Graduate Student Union steward 2020-2023
Top reviewer at NeurIPS — moderated deep-dive session 2022
Area chair at LMRL workshop at NeurIPS 2022
Top reviewer at ICML — chaired session 2022
Systems Biology PhD applicant mentor 2021
Systems Biology Department Equitable mentorship working group 2020

UNDERGRADUATE RESEARCH

Predicting interactions from sequences of disordered proteins using physics models 2017-2019
University of Toronto
Advisor: Hue Sun Chan

Measuring mechanics of nuclear membrane-less organelles using microfluidics Summer 2018
Princeton University
Advisor: Clifford Brangwynne

Designing new hydrogels with desired mechanical properties for drug delivery Summer 2017

University of Toronto
Advisor: Molly Shoichet

Loading tRNAs with alternative amino acids for synthetic biology 2016-2017
University of Toronto
Advisor: Ronald Kluger