
Education

- Dec 2022 **Ph.D Biomedical Engineering**, *University of Virginia*, Charlottesville, VA.
Title: *Explainable Deep Generative Models, Ancestral Fragments, and Murky Regions of the Protein Structure Universe: Datasets, Models, and Analyses of Fold Space*
Advisors: Philip E. Bourne and Cameron Mura
Chair: Kristen M. Naegle
Committee: Jason A. Papin, Aidong Zhang and Stephen Baek
- 2014 **B.S. Bioinformatics**, *University of California, Santa Cruz*, Santa Cruz, CA.
Thesis: *Correcting Frameshift Mutations in Transcriptomics Data*
Advisors: Mark Blaxter (Edinburgh), Dietlind Gerloff

Research Experience

- Apr-Aug 2022 **Machine Learning Intern**, *VantAI*, New York, NY.
- Jan 2018– **PhD student**, *University of Virginia*, Charlottesville, VA.
Work with Phil Bourne & Cameron Mura
Developed Prop3D, a python framework to compute atomic, residue and graph level physicochemical properties for all CATH domain in the cloud and share using HSDS
Developed DeepUrfold and DeepUrfold-explain, explainable deep generative models to explore remote relationships between CATH superfamilies at the Urfold level (e.g. discontinuous peptide fragments)
- Sep 2015– **Predoctoral Fellow**, *National Institutes of Health*, Bethesda, MD.
Aug 2018 Worked with Phil Bourne & Michael Grigg (NIAID),
Knocked out *T. gondii* surface proteins with CRISPR/Cas9 to understand function, evolution
- Sep 2015– **Graduate Student Intern**, *Harvard Medical School*, Boston, MA.
May 2016 Worked under Debora Marks & Chris Sander
Helped develop a python framework for coevolutionary sequence analysis
Added coevolutionary restraints into *Phenix* to improve low-resolution X-ray structures
- Jan-Aug 2015 **Postbac Fellow**, *National Center for Biotechnology Information*, Bethesda, MD.
Worked under Alexey Shaytan in Anna Panchenko's group
Wrote HistoneDB 2.0 Django webserver to classify histone sequences by variant with HMMs
- Apr-Dec 2014 **Student Assistant**, *Lawrence Berkeley National Labs*, Berkeley, CA.
- Sum 2010-11 Worked under Nigel Moriarty in Paul Adam's Lab
Validated and modelled glycoproteins by ligand fitting and built 2D carbohydrate builder
- Sep-Dec 2013 **Undergraduate Researcher**, *The University of Edinburgh*, Edinburgh, UK.
Worked under Martin Jones in Mark Blaxter's Lab for UCSC Senior Thesis
Wrote HSP-Tiler to fix frameshift mutations in RNAseq data using python, BLAST, HMMs
- May-Aug 2013 **Amgen Scholar**, *Washington University in St. Louis*, St. Louis, MO.
Worked under Garland Marshall
Wrote python script to add substituents to drugs and screened with OpenEye docking
- Sep 2010– **Undergraduate Researcher**, *University of California, Santa Cruz*, Santa Cruz, CA.
May 2013 Worked under Dietlind Gerloff
Predicted structures of Malarial surface proteins with MODELLER and HMMER
Created a public webserver to share structures

Publications

Published

Philip E. Bourne, **Eli J. Draizen**, and Cameron Mura. The curse of the ribbon. *PLoS Biology*, December 2022.

Daniele Parisi*, **Eli J. Draizen*** Gabriel J. Olguin-Orellana*, ..., and R. Gonzalo Parra.

Nurturing tomorrow's leaders: The iscb student council symposia in 2018. *F1000Research*, 2019.

Thomas A. Hopf, ..., **Eli J. Draizen**, ..., Chris Sander, and Debora S. Marks. The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics*, 2018.

Cameron Mura*, **Eli J. Draizen***, and Philip E. Bourne. Structural biology meets data science—does anything change? *Current Opinion in Structural Biology*, 2018.

Nigel W. Moriarty, **Eli J. Draizen**, and Paul D. Adams. A restraints editor for generation and customisation of geometry restraints. *Acta Cryst. D*, 73(2):123–130, Feb 2017.

Lei Xie, **Eli J. Draizen**, and Philip E. Bourne. Harnessing big data for systems pharmacology. *Annu Rev Pharmacol Toxicol.*, 57(1):245–262, 2017.

Eli J. Draizen, ..., David Landsman, and Anna R. Panchenko. HistoneDB 2.0: a histone database with variants – an integrated resource to explore histones and their variants. *Database*, 2016:baw014, 2016.

Preprints

Eli J. Draizen, Cameron Mura, and Philip E. Bourne. Explainable deep generative models, ancestral fragments, and murky regions of the protein structure universe. <https://doi.org/10.1101/2022.11.16.516787>, 2022.

Eli J. Draizen, Stella Veretnik, Cameron Mura, and Philip E. Bourne. Deep generative models of protein structure uncover distant relationships across a continuous fold space. <https://www.biorxiv.org/content/early/2022/08/01/2022.07.29.501943>, 2022.

Menuka Jaiswal, Saad Saleem, Yonghyeon Kweon, **Eli J Draizen**, Stella Veretnik, Cameron Mura, and Philip E. Bourne. Deep learning of protein structural classes: Any evidence for an 'unfold'? <https://arxiv.org/abs/2005.08443>, 2020.

Sean Mullane, Ruoyan Chen, Sri Vaishnavi Vemulapalli, **Eli J. Draizen**, Ke Wang, Cameron Mura, and Philip E. Bourne. Machine learning for classification of protein helix capping motifs. <https://arxiv.org/abs/1905.00455>, 2019.

In Preparation

Nigel W. Moriarty, **Eli J. Draizen**, and Paul D. Adams. CarboLoad: Modelling, validation, and prediction of carbohydrates in phenix.

Eli J. Draizen, Edward Y. Liaw, Jonathan Magasin, D. Steve Hall, Felecia D. Kemp, and Dietlind L. Gerloff. Tracking the evolutionary link between distant Apicomplexan surface protein families by their distinct disulfide patterns.

Talks

Invited Talks

December 11, 2020 **NeurIPS-LMRL**, *Deep Generative Models of Protein Domain Structures Uncover Distant Relationships: Evidence for an Unfold*, Virtual, <https://doi.org/10.5281/zenodo.4299037>.

July 15, 2020 **ISMB-3DSig**, *Deep Learning of Protein Structural Classes: Any Evidence for an 'Unfold'?*, Virtual, <https://doi.org/10.5281/zenodo.3909755>.

July 21, 2017 **ISMB-ISCB Student Council Symposium**, *Determining the functions of the Apicomplexan SRS/6-Cys protein family: A Structural and Evolutionary Understanding of Pathogen Invasion*, Prague, CZ.

May 14, 2015 **ICSB RSG-DC Student Research Symposium**, *Classification of Histone Variants*, College Park, MD.

Contributed and Seminar Talks

April 1, 2022 **Using HDF5 and HSDS for Open Science Research at Scale Workshop**, *Large-scale Biological Data Engineering with HDF5 + the Highly Scalable Data Service*, University of Notre Dame (Virtual), <https://doi.org/10.5281/zenodo.6406982>.

- June 25, 2021 **GBMES Student Summer Seminar Series**, *Deep Generative Models of Protein Domain Structures Can Uncover Distant Relationships: Evidence for an 'Urfold'*, University of Virginia, Charlottesville, VA.
- July 31, 2019 **Systems Biology Journal Club**, *Protein Interaction Prediction Using Deep Learning and Protein Structural Data*, Dept. of Biomedical Engineering, University of Virginia, Charlottesville, VA
- July 6, 2018 **ISMB-ISCB Student Council Symposium**, *Deep Learning Approaches to Predict Protein-Protein Interactions*, Chicago, IL.
- April 19, 2018 **Computational Biology Branch Seminar**, *Deep Learning Approaches to Predict Protein-Protein Binding Sites*, NLM/NCBI, National Institutes of Health, Bethesda, MD.
- May 11, 2017 **Laboratory of Parasitic Diseases Division Seminar**, *Determining the Function of the Apicomplexan SRS/6-Cys Protein Family: A Structural and Evolutionary Understanding of Pathogen Invasion*, NIAID, National Institutes of Health, Bethesda, MD.

Poster Presentations

- 2022 **NuerIPS-MLSB**, *Explainable Deep Generative Models, Ancestral Fragments, and Murky Regions of the Protein Structure Universe*, New Orleans, LA.
Eli J. Draizen, Cameron Mura, Philip E. Bourne
- 2020 **NeurIPS-LMRL**, *Deep Generative Models of Protein Domain Structures Uncover Distant Relationships: Evidence for an Urfold*, Virtual, <https://doi.org/10.5281/zenodo.4299037>.
Eli J. Draizen, Menuka Jaiswal, Saad Saleem, Yonghyeon Kweon, Stella Veretnik, Cameron Mura, Philip E. Bourne
- 2020 **ISMB-3DSig**, *A Data-engineering Pipeline for Deep Learning in Structural Bioinformatics*, Virtual, <https://doi.org/10.5281/zenodo.3914449>.
Eli J. Draizen, Cameron Mura, Philip E. Bourne
- 2019 **7th CAPRI**, *Template-based protein interaction and binding affinity predictions for deep learning*, Hinxton, UK, <https://doi.org/10.5281/zenodo.2641017>.
Eli J. Draizen, Zheng Zhao, Cameron Mura, Philip E. Bourne
- 2018 **ISMB-ISCB Student Council Symposium**, *Deep Learning Approaches to Predict Protein-Protein Interactions*, Chicago, IL.
Eli J. Draizen, Alexander Goncarenco, Cameron Mura, Anna R. Panchenko, Philip E. Bourne
- 2018 **NIH Graduate Student Research Symposium**, *Identifying Structural Motifs involved in Protein-Protein Interactions using Deep Neural Networks*, Bethesda, MD.
Eli J. Draizen, Alexander Goncarenco, Anna R. Panchenko, Philip E. Bourne
- 2017 **ISMB-ISCB Student Council Symposium**, *Determining the functions of the Apicomplexan SRS/6-Cys protein family: A Structural and Evolutionary Understanding of Pathogen Invasion*, Prague, CZ.
Eli J. Draizen, Beth Gregg, Philippe Youkharibache, Dietlind L. Gerloff, Michael E. Grigg, Philip E Bourne
- 2016 **International Workshop on Bioinformatics and Systems Biology**, *Predicting circadian gene expression in *Nuerospora crassa**, Tokyo, JP.
 Alan Pacheco, Yuqing Zhang, **Eli Draizen**, Matthew Sachs, Deborah Bel-Pedersen, James Galagan
- 2013 **Amgen Scholars Closing Symposium**, *Developing Epigenetic Therapeutics for the Eradication of Latent HIV*, Washington Univ. in St. Louis, St. Louis, MO.
Eli J. Draizen, Matthew K. Nguyen, Isaac B. Henson, Chris M. W. Ho, Garland R. Marshall
- 2012 **Research Review Day**, *The Evolutionary Link Between Coccidian SRS Proteins and the Malarial '6-Cys'-Domain Family: New Clues from Modelled Structures*, Santa Cruz, CA.
 Dietlind L Gerloff, Edward Liaw, **Eli Draizen**, David S Hall, Felicia Kemp, Richard Carter
- 2011 **Undergraduate Research Symposium**, *Bioinformatics Data Preparation Enables Effective Searches For Malarial Antigen Homologs*, Santa Cruz, CA.
Eli Draizen, Felicia Kemp, Edward Liaw, and Dietlind L. Gerloff

Honors and Awards

- 2019 **Presidential Fellowship in Data Science**, *University of Virginia*, Charlottesville, VA.
- 2015 **NIH Predoctoral Fellowship**, *Intramural Research Training Award (IRTA)*, Graduate Partnership Program (GPP), National Institutes of Health, Bethesda, MD.

Teaching Experience

- Sep-Dec 2020 **Biomedical Data Science (BME4550)**, *Teaching Assistant*, University of Virginia, Charlottesville, VA.
- 2012-2014 **Programming for Biologists (BME160)**, *Grader/Tutor*, Univ. of California, Santa Cruz.

Trainees

- 2020-2022 Abigail Newbury (REU student, Math), Skylar Brodowski (Undergrad, Chemistry)
- 2020 Mark Bray (REU Student), Saad Saleem (MSDS Capstone Student), Yonghyeon Kweon (MSDS Capstone Student), Menuka Jaiswal (MSDS Capstone Student)
- 2019 Gabriel Lipkowitz (Summer Student), Sri Vaishnavi V. (MSDS Capstone Student), Ruoyan Chen (MSDS Capstone Student), Sean Mullane (MSDS Capstone Student)

Service and Outreach

- July 6, 2018 **14th Annual ISCB Student Council Symposium**, *Co-Chair*, Chicago, IL.
- July 12, 2017 **ISCB RSG-DC Summer Workshop**, *Organizer*, College Park, MD.

Skills

- Languages Python, R, Perl, C/C++, Cython, SQL, Javascript, Java, PHP, Obj-C
- Frameworks PyTorch, PyTorch-Lightning, Keras, Tensorflow, Scikit-Learn, pandas, numpy, numba, matplotlib, Seaborn, HDF/HSDS
- Web Flask, Django, SQLAlchemy, Bootstrap, Jekyll, D3
- Cloud/HPC Kubernetes, AWS, Mesos, Slurm, SGE, LSF, Docker, Singularity, Toil, Snakemake, CWL
- Structure AlphaFold2, Geometricus, foldseek, ATOM3D, HADDOCK, CNS, PyMol, MODELLER, APBS, Phenix, Chimera, AutoDock
- Sequence ESM, ProtTrans, MMSeqs2, HMMER, HHpred, Blast, Biopython, EVCouplings, MUSCLE, UCLUST, FastTree, RAxML
- Databases AlphaFold DB, CATH, EPPIC, ECOD, SCOPe, NCBI, PDB, Pfam, SIFTS, VEuPathDB

Relevant Coursework (* Graduate)

- Comp. Sci. Data Structures, Algorithms, Database Systems, Machine Learning, Graph Neural Networks*
- Math/Stats Vector Calculus, Discrete Math, Linear Algebra, Differential Equations, Probability Theory, Classical/Bayesian Inference, Linear Statistical Modeling, Statistical Analysis of Network Data*, Biomedical Measurements (FFTs, Statistics)*
- Science General Chemistry, Organic Chemistry, Biochemistry, Genetics, Cell Biology, Intro Physics: Mechanics/E&M, Advanced Macromolecular Structure/Function*, Enzyme Mechanisms*, Physiology*
- Bioinformatics Comp Bio Tools, Bioinformatics Models/Algorithms*, Computational Genomics*, Computational Systems Biology*, Dynamics and Evolution of Biological Networks*