# Egbert Castro, PhD

egbertcastro.com · egbcastro@gmail.com · linkedin.com/in/castroegbert · github.com/egbert · Los Angeles, CA

## RESEARCH INTERESTS

Machine learning for biology and scientific discovery; generative modeling for proteins and molecules; representation learning for structured biological data; latent-space optimization and controllable generation; graph neural networks and geometric deep learning.

## EDUCATION

**Yale University**                                                                                     New Haven, CT
*PhD, Computational Biology & Bioinformatics*                                             2018 – 2023
- Advisor: Prof. Smita Krishnaswamy
- Dissertation: *Representation Learning and Generative Modeling for Biomolecular Sequences and Structures*

**University of California, San Diego**                                                          La Jolla, CA
*BS, Pharmacological Chemistry; Minor in Mathematics*                                2014 – 2018

## TEACHING EXPERIENCE

**Yale University**                                                                                     New Haven, CT
*Teaching Assistant*
- **CPSC 4520b, Deep Learning Theory and Applications** (Instructor: Dr. Smita Krishnaswamy). Supported a course covering the principles and practice of neural networks and modern deep learning, with applications in computer vision, natural language processing, and biomedical problems.
- **MCDB 570, Biotechnology** (Instructor: Dr. Craig Crews). Supported instruction on cellular, molecular, and chemical techniques in biotechnology, including applications to therapeutics, industrial agents, and biological research.

## RESEARCH AND PROFESSIONAL EXPERIENCE

**Ascent Bio**                                                            Hybrid - Los Angeles, CA and New York, NY
*Co-founder, AI Research*                                                                            2023 – Present
- Led the zero-to-one effort to build a VC-backed agentic science platform, from research to production.
- Accepted into the Merck Digital Sciences Accelerator and IndieBio NYC Accelerator.
- Designed and built agentic systems for scientific and drug-discovery workflows that combined foundation models, scientific tools, and code execution across multi-step research tasks.
- Developed evaluation frameworks for correctness, tool-use reliability, and failure-mode analysis on realistic scientific problems.
- Contributed to molecule reasoning-model research using synthetic data generation, reproducible experimentation, and rapid iteration in a shared research codebase.

**Yale University – Krishnaswamy Lab**                                                    New Haven, CT
*Graduate Researcher, Computational Biology & Bioinformatics*                  2018 – 2023
- Conducted research in geometric deep learning and generative modeling for biological sequences, structures, and graphs, with applications spanning proteins, RNA, molecular graphs, and clinical prediction.
- Developed transformer-based protein generative models with regularized latent-space optimization for gradient-guided search in learned sequence representations; first-author work published in *Nature Machine Intelligence*.
- Designed representation-learning methods for RNA secondary-structure landscapes and molecular graph generation, integrating graph-based architectures, geometric scattering, and task-specific biological objectives.
- Established reproducible multi-GPU training pipelines, curated benchmark datasets, and defined biologically grounded evaluation protocols for comparative analysis of biomolecular machine-learning methods.

**Genentech**                                                           South San Francisco, CA
*Machine Learning Research Intern, Discovery Chemistry*                            Summer 2018
– Developed predictive models for ADMET properties on internal compound libraries using deep neural networks.
– Built automated machine-learning pipelines integrating RDKit-based featurization with model training and evaluation.
– Partnered with discovery scientists on model evaluation for medicinal chemistry use cases.

## PUBLICATIONS

† *denotes first author.*

### Peer-Reviewed Journal Articles and Conference Papers

† **Castro, E.**, Godavarthi, A., Rubinfien, J., Givechian, K., Bhaskar, D., Krishnaswamy, S. *Transformer-based protein generation with regularized latent space optimization*. **Nature Machine Intelligence**, 4(10):840–851, 2022.

• Bhaskar, D., Grady, J., **Castro, E.**, Perlmutter, M., Krishnaswamy, S. *Molecular graph generation via geometric scattering*. 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP), 1–6, 2022.

• Shung, D., Huang, J., **Castro, E.**, Tay, J.K., Simonov, M., Laine, L., Batra, R., Krishnaswamy, S. *Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit*. **Scientific Reports**, 11(1):8827, 2021.

† **Castro, E.**, Benz, A., Tong, A., Wolf, G., Krishnaswamy, S. *Uncovering the folding landscape of RNA secondary structure using deep graph embeddings*. 2020 IEEE International Conference on Big Data (Big Data), 4519–4528, 2020.

• Tew, B.Y., Hong, T.B., Otto-Duessel, M., Elix, C., **Castro, E.**, He, M., Wu, X., Pal, S.K., Kalkum, M., Jones, J.O. *Vitamin K epoxide reductase regulation of androgen receptor activity*. **Oncotarget**, 8(8):13818–13831, 2017.

### Patent Applications

† **Castro, E.**, Krishnaswamy, S., Godavarthi, A., Rubinfien, J. *Regularized deep learning based improvement of biomolecules*. US Patent Application 18/061,272, published September 14, 2023.

### Preprints

• Afrasiyabi, A., Kovalic, J., Liu, C., **Castro, E.**, Weinreb, A., Varol, E., Miller III, D.M., Hammarlund, M., Krishnaswamy, S. *CellSpliceNet: Interpretable multimodal modeling of alternative splicing across neurons in C. elegans*. **bioRxiv**, 2025.06.22.660966, 2025.

• Viswanath, S., Bhaskar, D., Johnson, D.R., Rocha, J.F., **Castro, E.**, Grady, J.D., Grigas, A.T., Perlmutter, M.A., O'Hern, C.S., Krishnaswamy, S. *ProtSCAPE: Mapping the landscape of protein conformations in molecular dynamics*. arXiv preprint arXiv:2410.20317, 2024.

### Conference Abstracts

† **Castro, E.**, Shung, D. *993 NEURAL NETWORK PREDICTS DROP IN HEMOGLOBIN REQUIRING TRANSFUSION FOR PATIENTS WITH ACUTE GASTROINTESTINAL BLEEDING ADMITTED TO THE ICU*. **Gastroenterology**, 158(6):S-196–S-197, 2020.

## PRESENTATIONS & TALKS

• **ProtSCAPE: Mapping the landscape of protein conformations in molecular dynamics**. SMT Conference, 2025.

• **ProGSNN**. Presented at the Learning Meaningful Representations of Life Workshop at ICLR, 2021, and at the NLM Training Grant Conference, 2022.

• **Molecular graph generation via geometric scattering**. IEEE International Workshop on Machine Learning for Signal Processing, 2022.

• **Guided Generative Protein Design using Regularized Transformers**. ISMB/ECCB, 2021.

• **Uncovering the Folding Landscape of RNA Secondary Structure with Deep Graph Embeddings**. Presented at the GRLB Workshop at ICML and at the IEEE International Conference on Big Data, 2020.

## TECHNICAL SKILLS

| | |
|---|---|
| **Methods** | transformers, generative models, latent-space optimization, representation learning, sequence modeling, graph neural networks, geometric deep learning |
| **Domains** | protein sequence modeling, molecular property prediction, RNA structure, multimodal biological data, AI for scientific discovery |
| **Tools** | Python, PyTorch, JAX, Hugging Face, scikit-learn, RDKit, Weights and Biases, multi-GPU training, benchmarking, reproducible experimentation |