HAN TANG

LinkedIn Profile: Han Tang
Github Page: Han Tang
Email: ht@di.ku.dk

PUBLICATIONS

UNAAGI: Atom-Level Diffusion for Generating Non-Canonical Amino Acid Substitutions
Submitted to ICLR 2026

- Developed **UNAAGI**, an E(3)-equivariant molecular diffusion model that generates full-atom side chains for protein residues, enabling unified modeling across canonical and non-canonical amino acids.
- Evaluated UNAAGI against Deep Mutational Scan (DMS) benchmarks, achieving competitive performance on ProteinGym and the **first meaningful correlation on NCAA benchmarks**, demonstrating cross-domain generalization.
- Positioned UNAAGI as a conceptual bridge between **structure-based drug design (SBDD)** and **protein mutational effect prediction**, leveraging chemical-interaction modeling techniques from SBDD to improve variant effect inference.

Contextual Molecule Representation Learning from Chemical Reaction Knowledge arXiv preprint, 2024

- Introduced **REMO**, a self-supervised learning framework for molecular graphs that reconstructs reaction centers rather than randomly masked fragments, capturing chemically meaningful context during pretraining.
- Demonstrated that reaction-aware reconstruction yields richer molecular representations and improves performance on downstream property prediction benchmarks.

ONGOING PROJECTS

Space-Time DDPM (ST-DDPM): Modeling Molecular Dynamics with Equivariant Diffusion

- Extended E(3)-equivariant molecular diffusion frameworks with an additional temporal dimension to simulate molecular dynamics.
- Validated the model on systems ranging from the Müller–Brown potential to dialanine peptides, showing comprehensive mode coverage in trajectory space via TICA and Ramachandran analyses.

Predicting Protein Ensembles using Bayesian OpenFold

- Fine-tuned OpenFold with a **Bayesian network–based variational inference** module, inspired by Hennig et al. (2024), to capture conformational ensembles rather than single static predictions.
- Exploring its capability to represent structural uncertainty and ensemble diversity in protein folding landscapes.

EDUCATION

University of Copenhagen, Copenhagen, Denmark *September 2023 - December 2026 (expected)*
Ph.D. in Machine Learning for Biology
Department of Computer Science (DIKU)

PhD Advisor: Wouter Boomsma

Research Topics:

Diffusion Models, Molecular Generative Modeling, Non-Canonical Amino Acids, Protein Ensembles, Protein Dynamics Simulation

Technological University of Dublin, Dublin, Ireland *September 2018 - Febraury 2020*
MSc. in Computing (Data Science)
School of Computing

Dublin Institute of Technology, Kevin St., Dublin, Ireland *September 2017 - June 2018*
Pre-master for MSc in computing.

Beijing University of Chemical Technology, Beijing, China *September 2013 - June 2017*

BSc. in Applied Chemistry

Relevant Subjects:
Inorganic Chemistry, Organic Chemistry, Physical Chemistry, Structural Chemistry
Statistics and Probability, Linear Algebra, Multivariable calculus

Dissertation:
Study Factors Influencing the Layer Heights for Layered Double Hydroxides Using Computing Approaches.

CAREER EXPERIENCE

Institute for AI Industry Research (AIR), Tsinghua University December 2021 - August 2023
Research Engineer

- Conducted research in **AI for Drug Discovery and Protein Structure Prediction**, focusing on representation learning for molecular graphs and 3D conformations.
- Developed pretraining and unsupervised learning methods for small molecules using 2D topologies and 3D structures, and applied these models to downstream property prediction tasks.
- Led a project on **reaction-aware molecular representations**, leveraging chemical reaction data to enhance molecular encoders for contextualized learning.
- Reproduced the core methods of RGN-2 prior to its open-source release, and contributed to the **AIR-Fold** protein prediction pipeline, which achieved top ranking on CAMEO for four consecutive weeks.

Technological University Dublin & AutoPlan September 2020 - March 2021
Research Assistant

- Collaborated with a startup to design similarity-ranking algorithms to support urban planning applications.

State Key Laboratory of Chemical Resource Engineering November 2015 - May 2017
Research Assistant

- Developed Python tools with GUI interfaces for chemical parameter computation.

LANGUAGES & SKILLS & INTERESTS

Programming	R, Python
ML Frameworks	PyTorch
Research Areas	Diffusion Models, GNNs, Molecular Modeling
Languages	English (Full professional proficiency) Mandarin Chinese (Native proficiency)
Interests	Football (Soccer), played in multiple weekend football amateur leagues.