

Tanmoy Sanyal

San Francisco Bay Area, CA

tanmoy.7989@gmail.com • (805) 637-0375 • tanmoy7989.github.io • [LinkedIn](#) • [GitHub](#)

- Computational biology/chemistry/applied-ML generalist with 10+ years of experience developing physics-based and ML methods for protein design and structural biology, spanning academia, biotech, and pharma.
- Effective communicator skilled at translating complex technical concepts across disciplines into actionable insights for stakeholders and collaborators.

Education

University of California, Santa Barbara 2013–2018
Ph.D., Chemical Engineering (graduate emphasis in high performance computing)

Indian Institute of Technology, Kharagpur 2008–2013
Integrated M.Tech & B.Tech (Hons.), Chemical Engineering

Experience

Amgen Inc. — South San Francisco, CA
Senior Scientist (2023 – 2024), Principal Scientist (2024 – Present)

- **Design of multi-specific antibodies:** Extended antibody avidity models to bi-specific formats and heterogeneous receptor cell-surface morphologies through (AI co-pilot paired) pen-paper statistical mechanistic methods. These models are competitive with sequence → avidity ML approaches.
- **Data-efficient ML for property prediction:** Developed train/test split algorithms for small, imbalanced datasets preventing data leakage and confounding features; **rescued 4+ legacy datasets** to build ML screens (protein language model embeddings + classification heads) for **antibody hotspots**, **target protein expression**, and **nanobody thermo-stability**.
- **Neural ODEs for systems-biology parameter estimation:** Demonstrated parameter estimation in **large ODE networks (~50–60 equations)** of QSP models using Neural ODEs; built end-to-end pipelines for rate estimation in Bi-specific T-Cell Engager (BiTE) binding kinetics.
- **In-silico display library design:** Linear-programming algorithms to optimize degenerate codon libraries using protein language models; **fastest request-to-delivery (~3 days)** for two 1M+ diverse antibody libraries, enabling cross-reactivity engineering across 2 therapeutic areas.
- **MLOps:** Built Amgen's first cloud (Databricks) - native MLOps framework for supervised biophysical property prediction from protein sequences; now in use across **8+ antibody design projects**. Mentored junior data scientists in onboarding this framework.

Novo Nordisk Research Center Seattle Inc. — Seattle, WA 2022–2023
Protein Design Scientist

- **Peptide modifications for half-life extension:** Computational screens for scanning PEGylation sites on peptides for half life extension; published and open-sourced with a benchmark on GLP-1 analogs for anti-obesity therapeutics.
- **Biologics design:** Mini-binder design using RFDiffusion → ProteinMPNN → AlphaFold-2; active-learning pipeline (Gaussian process regression) achieving 20x affinity improvement over 3 design–experiment iterations.
- **Cloud infrastructure:** Re-tooled open-source AlphaFold-2 inference with AWS for scalable (~100k sequences) in-house structure prediction; now standard in designed-binder screening.

University of California, San Francisco — Sali Lab Jan 2019–Jan 2022
Postdoctoral Scholar

- **Nanobody biophysics:** Protein–protein docking score for nanobody–antigen interactions using crosslink

and escape mutation data; identified **vulnerable and resilient epitopes across SARS-CoV-2 variants**; revealed **framework mutations modulating distal CDR3 conformations** via atomistic MD simulations.

- **Integrative structure modeling**: Graph-sampling algorithms extracting domain boundaries in large protein complexes from crosslink data; published the **first coarse-grained structure** of the SMC5/6-NSE-2,5,6 pentameric complex.
- **Whole-cell digital twin**: Bayesian network models combining pharmacokinetic, Brownian dynamics, and enzyme-pathway models of the pancreatic beta-cell; software lead for the Pancreatic Beta-Cell Consortium (UCSF/USC).

University of California, Santa Barbara — Shell Lab

Ph.D. Researcher

2013–2018

- **Coarse-grained (CG) molecular models of liquids**: Local-density-dependent force field for thermodynamically accurate CG simulations of liquid–liquid phase separation of hydrophobes in water.
- **Variational inference for protein folding**: CG backbone force fields for **template-free folding of 200+ residue domains** via information-theoretic ensembling of MD data; applied to self-assembly in amyloidogenic peptides.

Computational Skills

- **ML on biological data**: PyTorch-focused modeling of protein function from sequence/structure, and systems-biology temporal profiles; learning from sparse/imbalanced data.
- **Mechanistic modeling**: Rapid prototyping of binding, avidity, selectivity, and stability via statistical-mechanics-based tractable likelihood models.
- **Compute**: Cloud (AWS Batch) and on-prem (Slurm) GPU training and orchestration.
- **Comp. chem/bio stack**: RFDiffusion → ProteinMPNN → AlphaFold-2, Rosetta, OpenMM, LAMMPS, GROMACS, IMP.
- **ML stack**: Scikit-learn, PyTorch, Lightning, TensorFlow-Probability, PyMC3, MLFlow, Hydra, Dask, Lux/Flux (Julia).

Open-Source Contributions

- **nnprotscan** — Computational scanning of chemical modifications in peptides.
- **nbspike** — Integrative epitope prediction for nanobody–antigen interactions.
- **PDB-Dev deposition** — Integrative structure of the SMC-5/6-NSE-2,5,6 pentamer complex.
- **Local density potentials** — LAMMPS plugin for local-density CG molecular mechanics.

Selected Publications

- N. Thomas, **T. Sanyal**, P. Greisen Jr. & K. Diebler. Structure-Based Computational Scanning of Chemical Modification Sites in Biologics. *ACS Omega*, 9(34), 36787–36794, 2024.
- F.D. Mast, P.C. Fridy, et al., **T. Sanyal** et al. Highly synergistic nanobody combinations targeting SARS-CoV-2 resistant to escape. *eLife*, 10: e73027, 2021.
- B. Raveh, L. Sun, K.L. White, **T. Sanyal**, et al. Bayesian metamodeling of complex biological systems across varying representations. *PNAS*, 118(35) e2104559118, 2021.
- Y. Yu, S. Li, Z. Ser, **T. Sanyal**, et al. Integrative analysis of the unique structural and functional features of the Smc5/6 complex. *PNAS*, 118(19) e2026844118, 2021.
- **T. Sanyal**, J. Mittal and M. Scott Shell. A hybrid, bottom-up, structurally-accurate, Go-like coarse-grained protein model. *J. Chem. Phys.*, 151(4): 044111, 2019.
- **T. Sanyal** and M. Scott Shell. Coarse-grained models using local-density potentials optimized with the relative entropy. *J. Chem. Phys.*, 145, 034109, 2016.