Tanmoy Sanyal

Scientist, Protein Design || Structural biology

About

San Francisco Bay Area, CA

tanmoy.7989@gmail.com

(805) 637-0375

tanmoy7989.github.io

in/tanmoy-sanyal

tanmoy7989

Protein Design Workflows

Rosetta

RFDiffusion + ProteinMPNN + AlphaFold

Protein Language Models (ESM2, Amplify)

Molecular modeling

openMM

LAMMPS

GROMACS

IMP

Modeller

UCSF ChimeraX

PyMOL

ML/ MLOps

PyTorch

PyTorch-lightning

Tensorflow-probability

PyMC3

MLFlow

AWS Batch

Languages

Python (+Cython) (~10000+ lines)

Fortran-90 (~1000+ lines)

C++ (~4000+ lines)

Bash

Effective github-copilot pair programmer

Open Source

nnprotscan

Computational scanning of chemical modifications in peptides

Education

University of California Santa Barbara *2013 - 2018*

Santa Babara, CA, USA

Ph.D Chemical Engineering

Graduate emphasis in Computational Science & Engineering (High Performance Computing)

Indian Institute of Technology Kharagpur *2008 - 2013*

Kharagpur, WB, India

Masters and Bachelors of Technology (Hons.) integrated dual degree Chemical Engineering

Employment History

Principal protein design scientist *April 2024 - Present*

Amgen Inc., South San Francisco, CA, USA

Multispecific molecules, MLOps

Senior protein design scientist *Oct 2023 - March 2024*

Amgen Inc., South San Francisco, CA, USA

Combinatorial protein library design

Protein design scientist *Jan 2022 - Oct 2023*

Novo Nordisk Research Center Seattle Inc., Seattle, WA, USA

Peptide modifications for half-life protraction

Postdoctoral scholar *Jan 2019 - Jan 2022*

University of California San Francisco, San Francisco, CA, USA

Structure determination from chemical crosslinks, nanobody biophysics, whole cell models

Experience

Amgen *Nov 2023 - Present*

- **Multispecifics design platform**: Developed cloud-native highly parallelized ODE solvers for pharmacokinetic modeling of bispecific molecules. Deployed ODE solutions through ML regression models providing 50x speedup.
- **Statistical mechanics of multispecific molecules**: Translated statistical mechanical models of superselectivity originally developed for multivalent ligands, to multispecifics. Currently undergoing experimental validation and benchmarking on bispecific candidates.
- **MLOps architect for protein design**: Created Amgen's first, cloud-native, end-to-end MLOps pipeline dedicated towards supervised biophysical property prediction from protein sequences. Pipeline currently used across 8+ antibody design projects. Also mentored junior data scientists in developing MLOps workflows for protein thermostability prediction from sequence.
- **In-silico display library design automation**: Developed linear-programming based algorithms to optimize degenerate codon libraries using protein language models. Set an internal record for the fastest request to delivery time (~ 3 days) for 1M+ diverse antibody library. Algorithm succesfully applied to engineer cross-reactivity in molecules across 2 TAs

Novo Nordisk Research Center Seattle *2022 - 2023*

- **Peptide modifications for half-life extension**: Developed a computational method for systematically scanning PEG-ylation sites on peptides for half life protraction. Published and open-sourced the method, illustrating a benchmark study on GLP-1 analogs that are particularly important for anti-obesity therapeutic development.

nbspike
Integrative epitope prediction

IMP-raindrops
Crosslink guided domain discovery

PDB-Dev deposition
Structure of SMC5/6

LAMMPS
Local density potential

Research Interests

Multi-specific molecules

Nanobodies

Optimizing LLM inference pipleines

Integrative structure modeling

Non-canonical amino acids

Variational methods for Bayesian inference

- **Biologics design**: Designed mini-binders for a rare disease TA using Rosetta-ddg, and AlphaFold / RFDiffusion / ProteinMPNN workflow. Developed a preliminary active-learning pipeline based on gaussian process regression, that provided an overall affinity improvement of 20x across 3 iterations of in-silico design and mRNA display.
- **Scientific software development**: Worked closely with AWS to re-tool open-source AlphaFold-2 and deploy a faster cloud-native in-house version for robust and scalable (~100k sequences) structure prediction; reguarly applied now in variant design within the company.

University of California San Francisco, Andrej Sali lab

2019 - 2022

- **Nanobody biophsyics and vaccine design**: Developed a protein-protein docking score for nanobody interaction with pathogenic proteins such as the SARS-CoV-2 Spike protein, using chemical crosslinking and escape mutation data. Applied this score to discover vulnerable and resilient epitopes on the Spike protein across variants of concern. Also performed MD simulation studies of point mutants in anti-SARS-CoV-2 nanobodies to elucidate the unusual effects of framework mutations on distal CDR3 residues.
- **Protein structure modeling using chemical crosslinks**: Developed graph sampling algorithms to extract domain boundaries from chemical cross-linking data. Applications included SMC and DNA helicase protein families. Published the first coarse-grained structure of the full SMC5/6 protein complex using crosslinking data and homology models.
- **Whole cell models**: Developed Bayesian networks to combine continuum scale (pharmacokinetic), particle scale (Brownian dynamics) and network scale (enzyme pathways) models of the pancreatic beta-cell into a proof-of-concept multimodal digitial twin. Served as software lead within the Pancreatic Beta-Cell consortium (collaborating across UCSF and USC)

University of California Santa Barbara, M. Scott Shell lab

2013 - 2018

- **Coarse-graining algorithms for bulk liquids**: Introduced the local density potential for structurally and thermodynamically accurate coarse-grained MD simulations of liquid-liquid phase separation, which are relatively rare in the chemical engineering literature
- **Coarse-graining algorithms for protein folding using variational inference techniques**: Developed protein backbone models for (i) template-free folding of 200+ residue coarse-grained protein domains using MD simulations, and (ii) studying self-assembly in amyloidogenic peptides
- **High Performance Computing**: As system administrator, set up 144 core Linux (Rocks 6.2) cluster and wrote utilities for automating resource sharing between lab members.

Volunteering

Peer-review

2019 - Present

Structure, Journal of Physical Chemistry, Proteins: Structure,Function and Bioinformatics, Review of Scientific Instruments, Rapid Reviews Covid-19, Life

Selected Publications

N. Thomas, **T. Sanyal**, P. Greisen Jr. & K. Diebler, *Structure-Based Computational Scanning of Chemical Modification Sites in Biologics*, 2024, ACS Omega, 9 (34), 36787-36794

F.D. Mast, P.C. Fridy, N.E. Karen, J. Wang, E.Y. Jacobs, J.P. Olivier, **T. Sanyal**, et al, *Highly synergistic combinations of nanobodies that target SARS-CoV-2 and are resistant to escape*, 2021, eLife, 10: e73027

B. Raveh, L. Sun, K.L. White, **T. Sanyal**, et al, *Bayesian metamodeling of complex biological systems across varying representations*, 2021, PNAS, 118 (35) e2104559118

Y. Yu, S. Li, Z. Ser, **T. Sanyal**, et al, *Integrative analysis reveals unique structural and functional features of the Smc5/6 complex*, 2021, Proc. Natl. Acad. Sci. USA, 118 (19) e2026844118

T. Sanyal, J. Mittal and M. Scott Shell, *A hybrid, bottom up, structurally-accurate, Go-like coarse-grained protein model*, 2019, J. Chem. Phys., 151 (4): 044111

T. Sanyal, and M. Scott Shell, *Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation*, 2016, J. Chem. Phys., 145, 034109