Inhee Park, PhD iparkirk@gmail.com | (408) 826-1692 | San Jose, CA | web | linkedin | LLM

ML Data Scientist in Biotech + PhD in Computational Chemical Physics + MS in Computer Science.

Professional Experience

Sr ML Data Scientist | Protein Evolution | San Jose, CA (Remote) | 2024–2024

- **NLP Model Development:** Designed and fine-tuned a domain-specific BERT model with custom masking strategies; integrated into protein sequence design workflows, predicting fitness for target-specific variants.
- **Interpretable ML for Imbalanced Data:** Shifted from pre-trained deep embeddings (ESM2) to interpretable, feature-engineered classifiers, improving F1-scores (0.75+ from 0.60) on highly imbalanced datasets while applying SHAP for model interpretability.

Computational Protein / ML Engineer | Aether Biomachines | Menlo Park, CA | 2021–2023

- **Recommender Systems Development:** Designed and deployed a chemistry-aware content-based recommender system, reducing chemical-to-enzyme mapping time by 95% and accelerating project kickoff.
- **Generative ML for Protein Design:** Built a GPT-2-based sequence generator integrating chemical reaction pathways and motif embeddings; led end-to-end development including dataset curation, BPE tokenization, model training, and tool integration.
- **Active Learning Optimization:** Developed an active learning pipeline to refine protein reactivity classification, improving performance at the decision boundary and enabling adaptive retraining through performance-based iteration.
- **ML-Driven Deployment Engineering:** Containerized 3D modeling workflows and built reproducible ML pipelines for catalytic site prediction, streamlining integration with downstream protein engineering tools.

Postdoctoral Fellow | Novartis Institutes for BioMedical Research | San Diego, CA | 2012–2015

- **Computational Modeling for Biologics Prediction:** Developed a GPU-accelerated predictive model for biologics, complementing mass spectrometry by saving resources, enhancing stability and mutation assessments, and driving innovation in computational bioanalytics critical to biotech research.
- **Multi-Omics Integration for Antibody Reconstruction:** Built and deployed an internal pipeline combining statistical models with genomics, proteomics, and informatics data, enabling high-confidence antibody sequence reconstruction and supporting rapid design with IP compliance.

Postdoctoral Fellow | Beckman Research Institute at City of Hope| Duarte, CA | 2010–2012

- Developed high-dimensional biological sampling protocols using physics-inspired motion models to improve simulation-informed predictive modeling of protein-ligand interactions.

Skills

- **Programming:** Python, Pandas, R, Bash, SQL
- **ML/DL Framework:** Scikit-learn, PyTorch, Hugging Face Transformer
- **Computing Environment:** Git, Linux, Docker, AWS

EDUCATION

- **MS Computer Science,** San Jose State University, CA 2018 — 2020
  - Thesis: "Multi-Agent Deep Reinforcement Learning for Walker Systems"
- **PhD** Computational Chemical Physics, The Ohio State University, OH 2006 — 2010
  - Dissertation: "Computational Simulations of Protein-Ligand Molecular Recognition via Enhanced Samplings, Free Energy Calculations and Applications to Structure-Based Drug Design"
- **BS** Chemistry & Physics, SookMyung Women's University, Seoul, Korea

SELECTED PUBLICATIONS

- Inhee Park and Teng-Sheng Moh, "Multi-Agent Deep Reinforcement Learning for Walker Systems", 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021, pp. 490-495

  #deep reinforcement learning (drl), #ai, #multi-agent

- Inhee Park, J. Venable, C. Steckler, S. Cellitti, S.A. Lesley, G. Spraggon and A. Brock, "Estimation of Hydrogen Exchange Protection Factors from MD Simulation Based on Amide Hydrogen Bonding Analysis". J Chem Inf Model. 55(9):1914-25. 2015

  #monte carlo, #mass spectrometry, #proteomics, #molecular simulation, #computational structural biology

- Inhee Park and A. Brock, "Genomics/Proteomics/Informatics Complementary Approach to High Confident Reconstruction of Antibody V-region Sequence" ACS West Coast Analytical Chemistry Symposium, Apr 2015, San Diego

  #bioinformatics, #genomics, #proteomics, #mass spectrometry, #dynamic programming

- Inhee Park, V. Gangupomu, J. Wagner, A. Jain and N. Vaidehi, "Structure Refinement of Protein Low Resolution Models Using the GNEIMO Constrained Dynamics Method" J Phys Chem B. 116(8):2365-75. 2012

  #homology modeling, #sampling, #molecular simulation, #free energy, #computational structural biology

- Inhee Park and C. Li, "Dynamic ligand-induced-fit simulation via enhanced conformational samplings and ensemble dockings: a survivin example" J Phys Chem B. 114(15):5144-53. 2010

  #induced-fit-docking, #molecular simulation, #free energy, #enhanced sampling

- Inhee Park and C. Li, "Characterization of molecular recognition of STAT3 SH2 domain inhibitors through molecular simulation" J Mol Recognition. 24(2):254-65. 2010

  #docking, #molecular simulation, #quantum chemistry, #free energy, #cheminformatics