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

ML Data Scientist in Biotech + PhD in Structure-base Drug Design + MS in Computer Science.

Professional Experience

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

- **NLP for Sequence Design:** Developed a domain-specific BERT model with varying masking percentages, enabling the generation of high-fitness sequences tailored to specific queries and accelerating protein design workflows.
- **Data-Centric Classifiers with Ultra High-Throughput Data:** Transitioned from model-centric deep learning with mutational ESM2 embeddings (0.65) to data-centric workflows focused on feature engineering to address data imbalance, achieving an improved F1-score of 0.75+ and leveraging explainable AI for library design optimization.
- **Enhanced Enzyme-Substrate Sampling:** Leveraged molecular dynamics simulations to improve protein stability (robustness) and activity (specificity), utilizing high-temperature MD snapshots and inverse virtual screening for conformational space exploration.

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

- **Chemistry-Aware Enzyme Recommender System:** Develpoed a content-based recommender system to retrieve enzymes/proteins based on chemical reaction pattern similarity. This innovation reduced retrieval times from one month to just one hour, significantly accelerating workflows and enabling the rapid initiation of new projects.
- **Generative AI with Chemistry-Biology Knowledge Association:** Designed and trained a GPT-2-based sequence generator with a BPE tokenizer (to reduce perplexity), integrating reaction pathways and active site motifs. This approach generated optimized sequences and accelerated discovery without structural data.
- **Active Learning for Imbalanced Data:** Developed a framework targeting low-confidence samples near the decision boundary, improving binary classification accuracy by 10% and enhancing generalization on imbalanced datasets through iterative refinement.
- **End-to-End Delivery Engineering:** Automated structural biology workflows with Dockerized toolchains for 3D simulation and catalytic site identification, streamlining protein engineering pipelines.

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.
- **De Novo Antibody Sequence Reconstruction Tool:** Developed an in-house tool for high-confidence antibody sequence reconstruction using public databases and genomics/proteomics/informatics integration, ensuring fast turnaround and IP protection.

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

- **Advanced Sampling:** Optimized high-dimensional biological sampling using robotics locomotion algorithms adapted from NASA/JPL, improving the accuracy of protein-ligand interaction studies.

Skills

- **Programming:** Python, Pandas, R, Bash, MySQL
- **ML/DL Framework:** Scikit-learn, PyTorch, Hugging Face Transformer
- **Computing Environment:** Git, Linux, Docker, AWS
- **Chemistry Tools:** RDKit, AmberTools, Amber, OpenMM, Rosetta, Schrödinger, ChimeraX

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

- S. Kyriacou, G. Meshulam-Simon, L. Clark, C. Chia and Inhee Park, "Enzyme engineering by means of reaction space indexing coupled with machine learning algorithms", Society for Industrial Microbiology and Biotechnology (SIMB), Aug 2022, San Francisco
  #enzyme engineering, #graph neural network (gnn), #machine learning

- 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

- S. Bhowmik, D. H. Jones, H.-P. Chiu, Inhee Park et al. "Structural and Functional Characterization of BaiA, An Enzyme Involved in Secondary Bile Acid Synthesis in Human Gut Microbe" Proteins.

82(2):216- 29. 2014

#QUANTUM CHEMISTRY, #CATALYTIC ACTIVITY

- 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

- A. Jain, Inhee Park and N. Vaidehi, "Equipartition Principle for Internal Coordinate Molecular Dynamics" J Chem Theory Comput. 8(8):2581-2587. 2012 #FREE ENERGY

- V. Gangupomu, J. Wagner, Inhee Park et al. "Mapping Conformational Dynamics of Proteins using Torsional Dynamics Simulations" Biophys J. 104(9):1999-2008. 2013

  #MOLECULAR SIMULATION, #FREE ENERGY, #COMPUTATIONAL STRUCTURAL BIOLOGY

- G.S. Balaraman, Inhee Park et al. "Folding of Small Proteins Using Constrained Molecular Dynamics" J Phys Chem B. 115(23):7588-96. 2011

  #PROTEIN FOLDING, #MOLECULAR SIMULATION, #FREE ENERGY, #COMPUTATIONAL STRUCTURAL BIOLOGY

- W. Harvey, Inhee Park, O. Rubel, V. Pascucci, P.T. Bremer, C. Li and Y. Wang, "A Collaborative Visual Analytics Suite for Protein Folding Research" J Mol Graph Model. 53:59-71.2014

  #GRAPH, #FOLDING, #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

- S. Chattier, J. V. Cooley, Inhee Park et al. "Design, Synthesis and Biological Studies of Survivin Dimerization Modulators that Prolong Mitotic Cycle" Bioorg Med Chem Lett. 23(19):5429-33. 2013

  #IN SILICO DESIGN, #DOCKING, #MOLECULAR SIMULATION, #CHEMINFORMATICS, #FREE ENERGY

- Co-authored with 16 authors, "Impairment of Glioma Stem Cell Survival and Growth by a Novel Inhibitor for Survivin–Ran Protein Complex" PClin Cancer Res. 19(3):631-42. 2013

  #IN SILICO DESIGN, #DRUG REPURPOSING, #TRANSLATION OF BREAST CANCER LEAD TO BRAIN CANCER