Patrick Finnerty

patrickfinnerty.com
pat@patrickfinnerty.com linkedin.com/in/patrick-finnerty
github.com/patimus-prime gitlab.com/patimus-prime

Summary

Multidisciplinary Engineer/Scientist — Bioinformatics — Software Engineering — Chemical Engineering

- 3+ years of experience in pharmaceutical R&D, bioinformatics, and data analysis/visualization
- Proficient in Python, JavaScript/TypeScript, R, machine learning and web development
- Strong expertise in developing and architecting software systems, including high-throughput data extraction and analysis, web applications, NGS and computational chemistry pipelines, and statistical analyses of high-dimensional data

Education

Universitat Autònoma de Barcelona **September 2020 – September 2021**
M.Sc. Bioinformatics (Thesis: Analysis of Hemoprotein Binding Sites) *Barcelona, Spain*

University of Arizona **2014 – 2018**
B.Sc. Chemical Engineering *Tucson, AZ*
B.Sc. Molecular and Cellular Biology
B.A. Biochemistry

Professional Experience

Eli Lilly, Bioproduct Research & Development **June 2018 – September 2020**
Scientist – Formulation Chemist *Indianapolis, Indiana*

- Designed and executed experiments to characterize antibody formulations, assess susceptibility to degradation, and determine optimal formulations. Extensive experimental design and testing: dozens to 100s of samples, numerous assays evaluating stability/activity at multiple timepoints and temperatures. Chemicals selected from GRAS list: buffer systems, excipients, osmotic agents, and stabilizers/surfactants
- Authored and evaluated technical documents such as process development and validation reports, SOPs, and lab notebooks to ensure FDA regulatory compliance and support GMP activities; approx. 150 documents written or reviewed during 2 year tenure
- Analyzed data from approx. 20 different assays and variable experimental conditions, and built predictive models, data dashboards with Excel and PowerBI. Assessed results from UV-Vis spectroscopy, mass spectrometry, viscosity/rheometry, HPLC, bioassays, particle count, pH, and more. Maintained data integrity and experimental reproducibility via LIMS system.
- Developed SOP, served as subject matter expert (SME), and managed vendor relationship for flow imaging microscopy instrument. Method development for the instrument involved ensuring reproducibility of subvisible particle count/images at 2-100µm range. Vendor coordination: resolved problems related to instrument performance, quality of results, scientific validity

Technical Skills

Languages	Python, JavaScript/TypeScript, R, Bash/Shell, GraphQL, SQL, LATEX
Technologies	Github, GitLab CI/CD, Docker, Nextflow, Linux, VSCode, Jupyter Notebooks, RStudio
Cloud	AWS, GCP, Azure, Vercel: EC2, Compute Engine, Virtual Machines, Google Kubernetes Engine, Route 53, Lambda, AppService, IAM
Databases	SQL, PostgreSQL, GraphQL, MongoDB, S3, Cloud Storage
Machine Learning	PyTorch, XGBoost, scikit-learn, DeepChem, Statsmodels, Hugging Face
Statistics	PCA/MDS, ANOVA, SVM, GNB, linear/logistic regression, clustering and multi-class classification, time-series analysis, signal processing, detrending
Data Visualization	Streamlit, Dash, Quarto, matplotlib, seaborn; Rmarkdown, xaringan, ggplot2, plotly
Web Development	Nextjs, nodejs, React, Mantine, Apollo Client/GraphQL, Prisma/SQL, Streamlit, Vercel
Data Wrangling	pandas; edgeR, tidyR, dplyr, data.table, knitr, clipr, stringr, kableExtra
Package Managers	poetry, yarn, npm, conda, pip, renv
Lab Techniques	HPLC, pH, flow imaging microscopy/particle analysis, HIAC particle counter, UV-Vis spectrophotometry, IR spectroscopy, gas chromatography, osmometry, rheometry, bacterial and mammalian cell culture, centrifugation, aseptic technique/fume hood/biosafety cabinet

Research & Projects below!

Research & Projects

Rare Genetic Disease Investigation **Oct 2022 – Present**

Python, Nextflow, Kubernetes, Google Cloud, Azure, gnomAD

- Investigating population prevalence of a rare genetic disorder, variant annotation of 3500 mutant proteins
- Implement Nextflow pipeline that ingests protein sequence data and distributes jobs to Kubernetes cluster to generate cheminformatic data, protein structures via OpenFold/AlphaFold; model protein interactions with ligands and chaperone proteins at relevant pH conditions, i.e. 6 to 7.4., simulate interaction energies via DiffDock
- Conduct statistical analysis of interaction energy results to determine if there are significant differences between approx. 160 predicted deleterious variants and wildtype proteins

Full Stack App Development **May 2022 – Present**

JavaScript/Typescript, Next.js, React, Python, Streamlit

- Developed website that presents portfolio work, prototype notebooks, and apps, linked from this resume:
- Front-end development: Designed and developed the user interface and user experience (UI/UX) of the website, using React and Mantine components, totaling 20+ pages/customized components, ensuring a visually appealing and responsive design that is optimized for various screen sizes and devices and highly maintainable
- Back-end development: core website is developed with Nextjs framework, React, Javascript, ESL 6; data-intensive apps are implemented with Python, Streamlit, embedded as iframes; website and apps are connected to databases by clients for SQL and GraphQL
- CI/CD and version control: Static site updated and deployed to Vercel upon Git update; Streamlit embeddings are managed via Git, Poetry, Hugging Face

Machine Learning and Statistical Analysis Projects **March 2022 – Present**

Python, PyTorch, Sci-kit, XGBoost, Statsmodels, Imbalanced-learn, Jupyter

- Identified 100 prime candidates from 6,000 samples of high-throughput screening data; detrended spectrophotometer absorbance data, ANOVA, compared candidates to baseline expression level
- Predicted classification of protein target of 500,000 small molecules given SMILES codes: generated QSAR data with RDKit/Mordred, handled imbalanced dataset, conducted feature engineering,hyper-parameter tuning
- Analyzed high-dimensional RNASeq TCGA data (20,000 genes x 1000 tumor samples): gene expression analysis; merged with clinical data, successfully predicted negative cancer status

Master's Thesis: Analysis of Hemoprotein Binding Sites **April 2021 – September 2021**

Python, R, R Markdown, UCSF Chimera *Barcelona, Spain*

- Designed and programmed robust pipeline to simulate/extract structural data from 100 proteins sourced from RCSB PDB; Python used to generate data from simulation software, R to organize and visualize data
- Implemented high-throughput functionality for 8+ features for Chimera molecular visualization software
- Wrangled and integrated data from 9 hetereogenous datasets; visualized and analyzed data, and automated reporting

Professional Affiliations and Interests

Society of Hispanic Professional Engineers **2016 – 2023**

American Institute of Chemical Engineers **2016 – 2023**

Interests: Cooking, reading, achieving fluency in Spanish, salsa dancing, learning piano, learning to surf, skateboarding