PHYO PHYO KYAW ZIN

Cheminformatics Scientist, Data Scientist & Software Developer

phyophyokyawzin1@gmail.com| 919.771.5803 | North Carolina | Blog | phyophyokyawzin | zinph | zinph

EXPERIENCE

Terray Therapeutics Inc. **(Remote)**

***Cheminformatics Scientist* (August 2022 – Present)**

- Design and optimize large virtual libraries containing up to hundreds of millions of molecules for chip-based synthesis, enhancing compound diversity, SAR understanding, hit optimization and research efficiency.
- Develop and automate rapid enumeration methods and workflows for virtual libraries, accelerating data generation and analysis.
- Create and manage SQL databases and metabase views for tracking library/protein targets, compound registration, status and results.
- Analyze large chemical data (millions of data in each chemical library) to extract actionable structure–activity relationship (SAR) insights, improving drug design and optimization.
- Developed user-friendly QSAR analysis pipelines and Streamlit apps for resynthesis selection, enhancing usability and decision-making.
- Collaborate with building block vendors to catalog and procure materials, expanding and diversifying Terray's chemical space.

Simulations Plus Inc. **(Remote)**

Scientist 1 – Cheminformatics (May 2021 – August 2022)

Postdoc – Cheminformatics (July 2020 – April 2021)

- Extracted and curated biological data on ADMET properties from literature, databases, and public sources, and developed automated pipelines for efficient chemical data processing and curation.
- Built machine learning models to predict biological endpoints, including metabolism sites and mutagenicity.

Collaborations Pharmaceuticals Inc. **(Hybrid)**

Cheminformatics Intern (June 2019 – August 2019)

- Developed MacrolactoneDB, a web application hosting macrolactones with bioactivity information curated from public repositories.
- Created QSAR regressors for multiple protein targets using various ML methods to predict biological endpoints of macrolides.

SKILLS

- **Areas of Expertise:** Cheminformatics, Data Science, Machine Learning, QSAR Analysis, Drug Design, Hit Identification, Lead Optimization, Molecular Docking, Virtual Screening
- **Programming Languages:** Python (pandas, numpy, scikit-learn, RDKit, regex, Streamlit, Jupyter Notebooks), SQL, R
- **Visualization Tools:** matplotlib/seaborn, ggplot2, Spotfire, Tableau, Plotly, Metabase, Spotfire, Chemaxon
- **Data Engineering & Automation:** Big Data, Workflow Automation, Cloud Integration, Web Scraping, Multi-GPU Computing
- **Machine Learning Techniques:** Clustering, Classification, Regression, Dimensionality Reduction, Ensemble Methods
- **Technical Communication:** Scientific Writing, Technical Documentation, Presentations, Dashboards & Reporting

FELLOWSHIPS

- AAUW International Doctoral Fellowship ($20k), 2019-2020
- Olive Ruth Russell Fellowship ($6k) (2016, 2017, 2019)
- 4-year tuition scholarship ($130k) at Berea College, 2012-2016
- Scholarship ($3.5k) at Pre-Collegiate Program of Yangon, 2011

EDUCATION

PhD, Chemistry (Specialty: Cheminformatics), 2016 - 2020
North Carolina State University, Raleigh, NC
Adviser: Denis Fourches

B.A., Computer Science & Chemistry (Dean's List), 2012 - 2016
Berea College, Berea, KY

Publications can be found on Google Scholar.

ACHIEVEMENTS & AWARDS

- Best Use of Google Cloud Platform at PackHacks2019 (Apr 2019)
- Second-Best Overall Project at DiamondHacks2019 (Mar 2019)
- Best Use of AI/ML at PearlHacks2019 (Feb 2019)
- CINF Scholarship for Scientific Excellence at ACS (Aug 2017)
- Storytelling prize at ETHDenver (Feb 2018)
- Class of 1958 Chemistry Award at Berea (Apr 2016)
- 2nd place in poster presentation at TRIWiC, Ohio (Feb 2016)
- Pugsley Award for Excellence in Mathematics at Berea (Dec 2012)

SELECTED PROJECTS

- **ROSE (Resynthesis Selection and Optimization Engine):** Developed a Streamlit app for suggesting compounds for resynthesis from large chemical libraries, leveraging multiple data-driven, SAR-derived algorithms and user-defined parameters for optimal results.
- **BB Mapper (Building Block Mapper):** Created a framework for mapping off-platform resynthesis compounds to on-platform compounds with various chemical tethers and locations, ensuring accurate mapping of building blocks and quality control.
- **Library Design Workflow:** Led the development and implementation of a comprehensive library design workflow, building core components and overseeing the process from design to registration and hand-off, to ensure efficiency, consistency, and seamless collaboration.
- **Library Enumeration Project:** Developed a reaction-based enumeration system and implemented an automated workflow to track, enumerate, and notify the team of library enumeration statuses via Slack, including detailed logs on successes and errors.
- **CNS-MPO Algorithm:** Implemented and integrated the CNS-MPO algorithm using vendor APIs into tDesign streamlit app for company-wide use.
- **BB-Swapper:** Developed a lead optimization algorithm for generating new compounds by swapping building block components based on user-defined parameters.
- **SoM (ADMET Predictor):** Mined and curated metabolism data to build ML models predicting atomic sites of metabolism and resulting metabolites.
- Ames (ADMET Predictor):Updated Ames models with new data, built classification models for predicting mutagenicity across major strains.
- **PKS Enumerator (PhD Thesis):** Developed a Structural Motif-based enumeration algorithm for generating in-silico libraries of macrolide scaffolds.
- **SIME (PhD Thesis):** Created a scaffold-based enumeration method integrating SMs and sugars to build virtual libraries of macrolides.
- **MacrolactoneDB (PhD Thesis):** Developed a web application for macrolactones, including bioactivity information curated and mined from public repositories and literature.
- **Imatinib DLNN (PhD Thesis):** Implemented rigorously validated and highly predictive QSAR regression models using 2D/3D/4D molecular dynamics derived & time-series based descriptors for Imatinib analogues
- **CryptoChem (PhD Thesis):** Designed a cheminformatics-based encryption system for secure chemical information transfer (Funded by DARPA).
- **ChemX:** Developed a program for generating molecule analogues by replacing fragments with similar building blocks.
- **Virtual Bio-profiling of PFAS Derivatives:** Conducted cheminformatics analysis and visualization of PFAS derivatives, enhancing understanding of environmental contaminants.