Luke Perrin

Enterprise R&D Data Strategy | AI/ML Enablement | Digital Transformation in Drug Discovery

San Diego, California
+1 (630) 442.3168 | luke@lukeperr.in
Github | LinkedIn | portfolio.lukeperr.in

Summary

Results-oriented professional with expertise in drug discovery operations, utilizing advanced data techniques to enhance R&D efforts. A polished consultant and data specialist skilled in creatively applying these principles to various initiatives across the pharmaceutical industry. Possessing exceptional interpersonal skills, I consistently deliver impactful data solutions and foster productive client engagements by translating complex technical insights into actionable executive strategies. Motivated to drive digital innovation in the industry, I am passionate about facilitating the delivery of new therapies to patients by utilizing effective data insights.

Skills Overview

- Effectively translates & communicates requirements and feedback between scientific and technical leaders to ensure proper alignment and delivery of digital functionality.
- Demonstrates a unique skill set in drug discovery, chemical data, artificial intelligence, data engineering, and data analytics.
- A critical and creative thinker, capable of designing and developing complex solutions that specifically address functional requirements.
- Well-organized in managing individual and team tasks, ensuring timely execution and high-quality outcomes.
- Proficient in programming and has a comprehensive understanding of cloud infrastructure (GCP & AWS), including high-performance computing (HPC), networking, storage, and databases, from both technical and strategic perspectives.
- A process-oriented problem solver who maintains an awareness of the "bigger picture" before pursuing specific solutions.
- Possesses extensive knowledge of business needs, operational processes, and challenges in pharmaceutical research and development (R&D) and clinical settings, particularly in data management and system efficacy.
- Intimate familiarity with software product development & delivery cycles, DevOps, MLOps, Scrum agile frameworks.

Professional Journey

Scientific Software Experience

Schrödinger | San Diego, CA
Senior Strategic Deployment Manager
LiveDesign®**, Enterprise Informatics Cloud SaaS Solution & Digital Lab for Molecular Design**

Digital Transformation via Strategic Implementation

- Collected clients' executive requirements related to organization-wide priority shifts toward digital transformation, data strategy improvement planning, and ML / AI solution enablement– proposing various approaches and outlining identified risks of different digital solutions that align with their requirements and corporate initiatives.
- Identified known and unrealized pain points experienced by partner organizations; leveraged intimate industry knowledge to implement digital solutions to optimize R&D operations, information systems, and data accessibility to maximize the value of the organization's data.
- Determined scientifically-cognizant software solutions and designed digital infrastructure improvements for supporting R&D information systems for small molecule and biologics.
- Engaged with clients' technical stakeholders to define system specifications, ensure security policy adherence, assess compatibility with existing digital infrastructure, and scope requirements for seamless adoption and integration of proposed scientific software solution.
- Designed solutions for Schrodinger's partner organizations that would attempt to centralize global R&D data, optimize user interfaces through which research scientists interact with data, and accelerate scientific R&D by leveraging cutting-edge predictive ML and agentic AI tools and methods.

Enterprise Platform Support & Technical Project Delivery

- Served as the primary contact for a $5.5 MM book of business and facilitated a 22% YoY average expansion of over 18 different partnerships by establishing strong, trust-based relationships with clients, delivering value to their organizations. Five of these partners are amongst the top 20 pharma (by revenue).
- Fulfilled and negotiated RFIs & RFPs as required by current or prospective partner organizations; drafted and revised technical SOWs for platform deployments, advanced technical projects, and software co-development opportunities.

- Ensured both the fulfillment of granular support inquiries and the timely delivery of enterprise-level projects, mirroring a partner's specific priorities & goals by serving as a liaison between scientific & technical stakeholders both internally and externally.
- Scoped, outlined, and managed a wide variety of technical projects, including enterprise cloud SaaS deployments, assay & registration data system integrations, custom feature & third-party tool development, HPC integrations for supporting *in silico* research, and more.
- Encouraged widespread platform adoption by highlighting the immediate and direct value of the new state to crucial scientific stakeholders; provided additional recommendations for strategies that would maximize the value received from the platform by configuring customized workflows, ETL methods, and supervised learning ML/AI strategies on their behalf.

Market Knowledge & Product Development

- Learned intimately about the similarities and differences between specific pharma organizations' digital obstacles and broader challenges across the industry by frequently engaging with each via conferences and partner site visits, respectively.
- Identified and documented high-impact product features that would further enable or optimize R&D workflows for multiple organizations.
- Continuously monitored the market for additional relevant tools and software that could complement our platform's functionality or pose a competitive risk.
- Negotiated feature requests between our partner organizations and internal product & engineering teams, oftentimes providing alternative solutions or interim workarounds if the feature development does not align with the broader product vision or immediate priorities.
- Authored & reviewed user stories, identified risks & mitigation strategies, assessed impact against technical lift of new feature proposals; presented new feature concepts through presentations, process flow diagrams, technical descriptions, and example UX mockups to paint a clear vision of the intended functionality and user experience to internal product and engineering teams.

Strateos | San Diego, CA

Cloud Lab Digital Solutions Engineer

***Strateos Software*, Cloud SaaS for Remote Orchestration of Autonomous Robotic Lab Systems**

Digital R&D Solution Design & Business Analysis

- Contributed to professional service software projects for three of the top 10 pharma companies (by revenue).
- Led consultation meetings for clients' stakeholders, organizing progress updates and solution proposals into concise presentations and process diagrams.
- Analyzed clients' IT infrastructure and operational workflows to identify improvement opportunities within their data practices, analytic pipelines, and R&D digital assets.
- Identified clients' scientific, technical, and business requirements, using these to determine strategic software & cloud solutions, system integrations, and R&D optimizations that addressed the organization's goals as a part of their digital transformation initiative.
- Determined a client-specific software product solution stack and created a comprehensive 100+ page proposal for an enterprise data strategy and implementation plan.

Product Management

- Crafted comprehensive user stories that conveyed the needs of scientific users and addressed gaps in workflows in layman's terms; collaborated closely with product management and engineering teams to develop a meaningful set of features.
- Assisted engineering teams in determining which sprint tasks to prioritize or delay based on client needs and challenges encountered in using the software effectively.

Strateos, San Diego, CA

Data & Automation Engineer, Chemistry

***Strateos Software*, Cloud SaaS for Remote Orchestration of Autonomous Robotic Lab Systems**

Data Integration & Automation

- Developed a data architecture for seamless integration across endpoint systems and configured synchronous ETL to streamline data entry and processing for chemists.
- Utilized chemical vendor APIs to consolidate building block inventories and automated the decomposition of targets into available constituents. Created an algorithm to identify optimal fragment choices based on synthetic pathways, lead times, availability, and price.
- Established a process for real-time data transformation, providing management with a live BI dashboard and regular snapshots of key lab KPIs to guide senior leadership in decision-making.

Cheminformatics & Machine Learning

- Used chemistry software and libraries, mainly RDKit, Open Babel, and CACTVS, to extract and store chemical data.
- Proficiently constructed cheminformatic pipelines, such as calculating physicochemical properties, R-group enumeration, reaction enumeration with SMARTS, substructure matching, fingerprinting, atom mapping, and chemical structure visualization.
- Familiar with formats for structural information, including SMILES, SMARTS, MOL/SDF, CDXML, RXN, XYZ, and PDB.

- Applied regression analysis, classification, ensemble learning, neural networks, and K-means clustering for SAR analysis to predict properties by training ML models with chemical data and experimental results.

Scientific R&D (Wet Lab) Experience

Strateos | Eli Lilly Biotechnological Center, San Diego, CA
Associate Scientist II, Synthetic Chemistry (Small Molecule)
***Strateos Software*, Cloud SaaS for Remote Orchestration of Autonomous Robotic Lab Systems**

- Supported and conducted high-throughput chemical synthesis production in an automated lab for campaigns consisting of library sizes over 100. Then, developed workflows to accommodate even larger libraries
- Developed operational workflows and systems to support greater sample sets of chemical synthesis, allowing these to be processed by the automation platform and proceed to purification, QC analysis, and biological assay experimentation.
- Oversaw and optimized processes to purchase and intake large orders of building blocks and reagents for chemical synthesis. Worked to expedite and economize building block purchasing and receipt by surveying across various vendors using batch query tools with various chemical identifiers.

Integrity Bio | Thousand Oaks, CA
Research Associate II, Biochemistry (Large Molecule)

- Worked to perform analytical biochemical methods and techniques for profiling changes in clients' proteins and peptides under various stress conditions in the determination of the drug's active component stability in various formulation buffers.
- Performed various assays on sample sets, working to minimize errors and alteration of results due to fault by the analyst or the instrument. These assays include SE- and IEX-HPLC and UPLC; IsoQuant; SDS-PAGE; CE-SDS; icIEF; freezing point osmometer; concentration determination by UV-VIS spectrometry (including a variable pathlength spectrophotometer); subvisible particle analysis by DLS, FlowCAM, MFI, and HIAC.

Education

Data Analytics & Visualizations Bootcamp | UC Berkeley Extension – *Certificate*

- Microsoft Excel: VBA macros, statistics modeling
- Python Data Analytics and Visualization: Pandas, NumPy, JSON, APIs, Matplotlib, BeautifulSoup
- Databases: SQL, PostgreSQL, MongoDB, ETL Process
- Web Visualization: HTML, CSS, JavaScript, AJAX, Leaflet
- Big Data: Tableau, Hadoop, Machine Learning
- PurifAI (final project): Collaborated with a small team to develop a supervised learning model designed to forecast suitable chemical purification and LCMS method parameters based on the chemometric characteristics of compounds. The training dataset and processed information were stored in AWS RDS and were preprocessed using PySpark. The machine learning model was fine-tuned by utilizing Balanced Random Forest, XGBoost, Easy Ensemble AdaBoost, and Logistic Regression.

Westmont College | Santa Barbara, CA – *Bachelor of Science, Chemistry*

- Selected the compounds and designed syntheses of carboxy-amino-substituted 1,10-phenanthroline bidentate ligands to complex with platinum (II) centers, purposed to form bridged and ionically linked macrostructure networks.
- Total synthesis of several epoxides from various substrates via asymmetric epoxidation catalyzed by highly selective, synthesized Chiral (salen) Mn(III) complexes (relating to the Jacobsen Epoxidation). The resulting products were studied to investigate the stereochemical, thermodynamic, and kinetic effects of the catalyst structure and substrate on the synthetic outcome.
- Selected a recently isolated natural product to attempt to propose a reasonable synthesis pathway. By compiling literary evidence that provided explicit or inductive support for each of the twenty steps in the highly enantioselective synthesis of a *nakijiquinone* derivative, a reasonable total synthesis was proposed.

Additional Assets

Scientific Domain Knowledge

Chemistry & Biochemistry – synthesis & purification, analytical, computational, organic, inorganic, catalysis, organometallics, polymer, physical/theoretical, peptides, and proteins.

Therapeutic Modalities – small molecules, peptides (linear, branched, and cyclic), conjugates, and monoclonal & polyclonal antibodies.

Pharmacology – *in silico* computational and predictive modeling, high-throughput screening (HTS), structure-activity relationship (SAR) analysis, potency & selectivity optimization, applied pharmacokinetics & pharmacodynamics (PK/PD) theory, toxicology & stability risk mitigation, *in vitro* and *in vivo* assay experimentation, etc.

Scientific Software Experience

- **Computational Chemistry –** Schrödinger Suite, ORCA, SAMSON, GaussView, Gaussian 16, Tinker, AutoDock, Dalton, Avogadro, PyMol
- **Registration & Informatics (ELN/LIMS) –** LiveDesign, Dotmatics, CDD Vault, Signals ELN, RegMol, Scilligence ELN
- **Chemistry Software –** ChemDraw, ChemOffice, ChemSketch, ChemAxon Suite, PyMol, JMol, OpenBabel, DataWarrior, Optibrium modules, DataGrok, Spotfire, KNIME
- **Analytical Chemistry Software –** MestreNova (MestreLab), ChemStation (Agilent), Chromeleon (Thermo Fisher), Empower (Waters), TopSpin (Bruker)

Technical Skills

- **Programming Languages –** Python, SQL, JavaScript, HTML/CSS, Rust, R, Julia, Unix
- **Data Analysis, Visualization & BI Tools –** MS Excel, KNIME, MS Power BI, Tableau, TIBCO Spotfire, Sisense, Periscope Data, Airtable, Jupyter Notebook/Hub, Google Looker (Data Studio), Sigma Compute (BI software), Google Colab, RStudio, MATLAB, Kaggle, Pandas, NumPy, JSON, APIs, Matplotlib, BeautifulSoup, React, Leaflet, several machine learning modules (e.g., scikit-learn, deepchem, etc.)
- **Developer Tools –** IDEs (VS Code, Jetbrains), cloud service CLIs, Github/GitLab, Claude Code, Tabnine, Terraform, Docker, Kubernetes
- **Product & Project Management –** Atlassian Cloud (Jira & Confluence), Asana, Monday, Airtable, Figma, Lucid Chart, Miro
- **Big Data & Databases –** SQL database tools (pgAdmin, Navicat, Postico, DataGrip), MongoDB (Atlas & Compass), Neo4j (GraphQL), Google Cloud BigQuery, AWS RDS, Postman API, Hadoop, PySpark
- **Cloud Service Providers –** Google Cloud Platform (GCP) & Amazon Web Services (AWS) cloud console, architecture solutions, & many of their services

Cheminformatics & Bioinformatics

- **Molecular Encoding & File Formats –** Daylight CTFiles (MDL, MOL, SDF), SMILES, SMARTS, SELFIES, InChI, fingerprints (ECFP & Morgan), XYZ, molecular graphs, CDXML
- **Biologic Sequence Encoding & Notation –** HELM, FASTA, monomer graph, ESM-2
- **Molecular Registration & Standardization –** registration hashes, deduplication logic, tautomerization & stereoisomism, delocalized aromaticity, salt forms, charge neutralization
- **Molecular Vectorization & Featurization –** Fingerprint bit vectors (MACCS keys, ECFP, Morgan), 1D/2D/3D descriptors, graph neural & convolutional networks (GNNs & GCNs), atom & pharmacophore maps
- **Molecular Assessment –** MCS analysis, Tanimoto scoring, ROCS analysis, PCA, UMAP, clustering methods (hierarchical, K-means, Butina, DBSCAN), RDKit postgres database cartidge
- **Molecular Transformation –** Decoding and encoding across formats, R-group decomposition, R-group & reaction-based enumerations (RXN SMARTS), transformations via atom map specifications
- **Biologics Sequence Analysis & Transformation –** peptide sequence & MAFFT alignment, monomeristic to atomistic transformations, single- & multi-point residue mutations
- **Cheminformatics Tools –** RDKit, OpenBabel, CACTVS, ChemAxon, Schrödinger Suite (Python API), Pandas RDKit wrapper, NextMove Software

Languages

English (*native*), French (*intermediate*), German (*intermediate*)

Past Project Achievements

PurifAI

Developed a supervised learning model that intended to forecast which SPE (purification) & LCMS (analysis) method conditions would best suit specific compound samples. The model was trained using historical observations collected from lab notebooks, and molecules were featurized using both fingerprint vectorization and chemometric characterization (RDKit). The training dataset and transformed data were stored in AWS RDS and were preprocessed using PySpark. The machine learning model was fine-tuned using Balanced Random Forest, XGBoost, Easy Ensemble AdaBoost, and Logistic Regression via scikit-learn.

The validated model was deployed in-web using React and JavaScript, enabling chemists to evaluate which SPE & LCMS methods were forecasted to be successful for newly synthesized compounds.

Retrosynthetic AI

Implemented a hybrid approach using both MCS and Tanimoto similarity scoring to query *NextMove*'s Pistachio reaction database and identify similar compounds with associated synthetic pathways. Once adequately similar targets were identified and scored for confidence, the reaction schemes were iteratively parsed against the original target (via SMARTS pattern matching) to specify the detachment points for decomposition. The decomposed components are then enumerated into the corresponding starting materials at each reaction step, and the list of initial starting materials (SMILES) and the conditions & catalysts at each reaction step were programmatically recorded into ELN via API.

Algorithmic Survey of the Commercial Availability for Chemical Building Blocks

Using several commercial chemical building block distributors' inventory databases (namely, *eMolecules*, *Chemspace*, and *Enamine*), a list of required starting materials for chemical synthesis (e.g., the reagents specified as the outputs from the *Retrosynthetic AI* project above) could be programmatically queried by structure similarity using each distributor's API. The search tool was configured so that commercial inventory filters could be applied (e.g., amount available, estimated lead times, price limits, etc.), and one could specify suitable atom alternatives for specific reagents (e.g., in the case of amine alkylations, either alkyl bromide or chloride would be suitable hits).

Once the raw vendor hit data was returned via API query, the data was reformatted to a single, standardized table. The inventory filters specified were applied to the cleaned data. Later, the tool was additionally equipped with the ability to track specific vendors' price and lead time performance, which was considered in the preferential selection of the algorithm's survey across all inventory databases. The tool was wrapped with a lightweight Flask GUI and containerized via Docker for broader use by the chemistry team.

Automated Sample Registration, ELN Entry, & Campaign Status Report Generation

Developed a programmatic mechanism to take raw compound structure, synthesis reaction scheme, purification & analytical method information to automatically 1) register compounds in the registrar (*CompReg*), 2) create new synthesis & assay requests in the tracker (*Airtable*), 3) create a new experiment entry in the ELN (*Signals* ELN), and 4) generate project status reports & dashboards weekly by each synthesis campaign.

This was accomplished by configuring a cronjob-based ETL process that pulled and pushed updates from an intermediate database using each endpoint system's respective API.

Cheminformatics Toolbox

Developed a series of Jupyter notebooks outlining a series of various different cheminformatic processes useful for research chemists needing to manipulate chemical structures in bulk. The notebooks covered basic chemical structure file or text reading & writing, molecular vectorizations & featurizations, transformations with R-group decomposition & enumeration, and molecular dataset analysis using MCS, clustering, and SAR analysis. These incorporated RDKit, OpenBabel, Schrödinger's BBChem API, and other Python-based chemistry modules.

LiveDesign Python Client (LDClient) Tutorials

Due to the complexity of LiveDesign's Python Client API (LDClient) and the sourcecode-like format of the documentation, LiveDesign users found it challenging to navigate the client effectively. I developed a series of Jupyter notebooks that provide an overview of the client's major functions and how the layers of objects are used within each. Each notebook covers a specific topic and focuses on user features, specific real-world use cases, and scientific workflows rather than individual client requests lacking context. This has been widely distributed across the organization and externally for LiveDesign users.