

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 and advanced data strategies that enhance R&D performance. A polished consultant and data specialist with a proven ability to creatively apply informatics, AI, and engineering principles across pharmaceutical initiatives. Skilled at translating complex technical concepts into actionable executive strategies, fostering strong stakeholder engagement, and delivering impactful digital solutions. Passionate about accelerating therapeutic development through effective data infrastructure, machine learning enablement, and enterprise digital transformation.

Skills Overview

- Effectively translate and communicate requirements between scientific and technical leaders to ensure alignment and successful digital delivery.
- Demonstrate expertise across drug discovery, chemical data systems, artificial intelligence, data engineering, and analytics.
- Design and develop complex, purpose-built solutions that address functional and organizational requirements.
- Manage individual and team priorities with strong organizational discipline, ensuring timely, high-quality execution.
- Proficient in programming and cloud infrastructure (GCP and AWS), including high-performance computing (HPC), networking, storage, and database systems from both technical and strategic perspectives.
- Apply a process-oriented, systems-level mindset while maintaining awareness of broader business and scientific objectives.
- Possess deep knowledge of pharmaceutical R&D and clinical data management workflows, operational processes, and system optimization challenges.
- Intimate familiarity with software product development & delivery cycles, DevOps, MLOps, Scrum agile frameworks.

Professional Experience

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

- Gather executive-level requirements related to enterprise digital transformation, data strategy modernization, and ML/AI enablement; propose aligned solution pathways and outline associated technical and organizational risks.
- Identify both explicit and latent operational pain points across partner organizations; implement digital solutions that optimize R&D workflows, data accessibility, and system interoperability.
- Architect scientifically informed software solutions and digital infrastructure improvements supporting small molecule and biologics research.
- Collaborate with technical stakeholders to define system specifications, ensure adherence to security policies, assess infrastructure compatibility, and scope seamless SaaS integrations.
- Design strategies to centralize global R&D data, enhance scientific user interfaces, and accelerate research through predictive ML and agentic AI methodologies.

Enterprise Platform Support & Technical Project Delivery

- Serve as the primary strategic contact for a \$5M+ book of business, driving an average 22% year-over-year expansion across 18+ partnerships, including five top-20 global pharmaceutical companies.
- Lead RFIs and RFP responses; draft and negotiate technical statements of work (SOWs) for enterprise deployments, advanced integrations, and co-development initiatives.
- Manage enterprise cloud SaaS deployments, assay and registration system integrations, custom feature development, third-party tool integrations, and HPC enablement for in silico research.
- Act as liaison between scientific, product, and engineering teams to align technical execution with business priorities.
- Promote platform adoption by configuring customized workflows, ETL processes, and supervised ML strategies tailored to each organization's objectives.

Market Knowledge & Product Development

- Develop deep insight into industry-wide digital challenges through direct engagement with global pharmaceutical partners.
- Identify high-impact product enhancements that optimize R&D workflows across multiple organizations.
- Monitor competitive and complementary technologies to inform product positioning and roadmap strategy.
- Negotiate feature requests between partners and internal teams, balancing client needs with long-term product vision.
- Author user stories, assess risk-mitigation strategies, evaluate technical lift versus impact, and present feature concepts via technical documentation, process diagrams, and UX mockups.

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 services engagements for three of the top 10 global pharmaceutical companies.
- Led stakeholder consultations and translated complex progress updates and solution proposals into concise presentations and process diagrams.
- Analyzed enterprise IT infrastructure and scientific workflows to identify data modernization and analytics optimization opportunities.
- Defined scientific, technical, and business requirements to design strategic cloud software stacks and integration roadmaps aligned with digital transformation initiatives.
- Developed a comprehensive 100+ page enterprise data strategy and implementation proposal.

Product Management

- Authored detailed user stories that articulated scientific workflow gaps in clear, accessible language.
- Partnered with engineering and product teams to prioritize sprint objectives based on customer impact and technical feasibility.

Strateos, San Diego, CA

Data & Automation Engineer, Chemistry

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

Data Integration & Automation

- Designed and implemented data architectures enabling seamless integration across endpoint laboratory systems.
- Configured synchronous ETL pipelines to streamline data capture and processing for chemists.
- Integrated chemical vendor APIs to consolidate building block inventories and developed algorithms to optimize fragment selection based on synthetic feasibility, availability, cost, and lead times.
- Established real-time data transformation pipelines powering executive BI dashboards and KPI reporting.

Cheminformatics & Machine Learning

- Leveraged RDKit, Open Babel, and CACTVS for chemical data extraction, transformation, and storage.
- Built cheminformatics pipelines including physicochemical property calculations, R-group enumeration, reaction enumeration (SMARTS), substructure matching, fingerprinting, atom mapping, and visualization.
- Applied regression, classification, ensemble learning, neural networks, and clustering techniques for SAR analysis and predictive modeling.

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 high-throughput automated synthesis campaigns involving libraries of 100+ compounds.
- Developed workflows to scale synthesis throughput and integrate automation with purification, QC, and assay systems.
- Optimized reagent sourcing by surveying vendors via batch chemical queries, improving cost and procurement efficiency.

Integrity Bio | Thousand Oaks, CA

Research Associate II, Biochemistry (Large Molecule)

- Conducted analytical and stability assessments on proteins and peptides under varied stress conditions.
- Performed and optimized assays including SE-HPLC, IEX-HPLC, UPLC, SDS-PAGE, CE-SDS, idEF, DLS, FlowCAM, MFI, HIAC, and UV-VIS spectroscopy.
- Ensured analytical rigor, reproducibility, and data integrity across experimental workflows.

Education

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

- Excel (VBA, statistical modeling)
- Python (Pandas, NumPy, APIs, Matplotlib)
- Databases (SQL, PostgreSQL, MongoDB, ETL)
- 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, Sciligence 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, RESTful API
- **Web Development** – Flask, Django, React, Angular, Astro, Tailwind
- **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, Hugging Face, 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 & stereoisomerism, 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, monomeric 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 machine learning model designed to forecast which solid-phase extraction (SPE) purification methods and LCMS analytical conditions would be most suitable for newly synthesized compounds. Aggregated and cleaned historical experimental observations from electronic lab notebooks, then featurized molecules using both fingerprint vectorization (e.g., Morgan/ECFP fingerprints) and chemometric descriptors generated with RDKit. Stored the structured dataset in AWS RDS and performed distributed preprocessing using PySpark to support scalable transformation and model training. Evaluated and fine-tuned multiple ensemble approaches—including Balanced Random Forest, XGBoost, Easy Ensemble AdaBoost, and Logistic Regression—optimizing for predictive performance across imbalanced experimental outcomes. Deployed the validated model through a lightweight web interface (React/JavaScript), enabling bench chemists to proactively assess purification strategy success probability before committing laboratory resources.

Retrosynthetic AI

Implemented a hybrid retrosynthetic planning pipeline combining Maximum Common Substructure (MCS) analysis and Tanimoto similarity scoring to query large-scale reaction databases (e.g., Pistachio) for precedent synthetic routes. Ranked candidate analogs by structural similarity and reaction confidence, then iteratively parsed reaction SMARTS patterns to identify bond disconnections and feasible decomposition points for the target molecule. Programmatically enumerated intermediate precursors and starting materials at each reaction step, capturing required reagents, catalysts, and conditions. Automatically recorded synthesized routes and associated metadata into ELN systems via API integration, ensuring reproducibility and knowledge retention. This system provided chemists with data-informed synthetic suggestions grounded in historical precedent while maintaining flexibility for expert override.

Algorithmic Survey of the Commercial Availability for Chemical Building Blocks

Designed and implemented a programmatic survey tool to assess commercial availability of required starting materials across major chemical vendors (eMolecules, Chemspace, Enamine). Queried vendor APIs using structure- and similarity-based search logic, incorporating configurable filters such as quantity thresholds, lead times, pricing constraints, and acceptable functional group substitutions (e.g., halide interchangeability). Standardized and merged raw vendor responses into a unified data schema, applied performance-weighted vendor scoring, and tracked historical fulfillment reliability to inform optimal supplier selection. Extended functionality to monitor vendor price trends and lead-time variability over time. Wrapped the tool in a lightweight Flask GUI and containerized it with Docker to support broader team accessibility and reproducible deployment.

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

Engineered a cron-based ETL orchestration framework that automated compound registration, experiment documentation, assay request creation, and campaign-level reporting across multiple endpoint systems. Programmatically registered compounds in CompReg, generated experiment records in Signals ELN, and created synthesis and assay tracking entries in Airtable through coordinated API integrations. Designed an intermediate database layer to manage synchronization, validation logic, and error handling between systems. Automated generation of weekly campaign dashboards and status reports to provide leadership with real-time insight into synthesis throughput, bottlenecks, and performance metrics. This automation significantly reduced manual data entry burden while improving data consistency and cross-system traceability.

Cheminformatics Toolbox

Developed a modular suite of Jupyter notebooks that provided reusable workflows for common cheminformatics tasks required by research chemists. Included functionality for structure file parsing and conversion (SMILES, SDF, MOL), molecular featurization (fingerprints, descriptors), R-group decomposition and enumeration, reaction-based transformations (RXN SMARTS), clustering analyses (hierarchical, K-means, Butina), and structure-activity relationship (SAR) visualization. Integrated RDKit, OpenBabel, Schrödinger APIs, and Pandas-based data manipulation to enable scalable dataset analysis. Designed notebooks to be instructional and reproducible, allowing chemists with limited programming experience to execute advanced structure-based analyses.

LiveDesign Python Client (LDClient) Tutorials

Created a comprehensive series of Jupyter-based tutorials to demystify the LiveDesign Python Client API, which was previously difficult for users to navigate due to its object-layered architecture and source-code-oriented documentation. Structured tutorials around real-world scientific workflows and use cases rather than isolated function calls, illustrating how objects interact across the API hierarchy. Provided step-by-step examples covering data retrieval, compound manipulation, assay data integration, and workflow automation. Distributed the materials internally and externally to accelerate adoption and reduce onboarding friction for scientific users leveraging programmatic access to the platform.