tetrascience

Improving CMC with analytics-ready chromatography data

CASE STUDY

The development of biologics begins in the laboratory, where scientists optimize production and analytical methods before scaling them up for pilots and, ultimately, commercial manufacturing. At a leading biopharmaceutical company, a team of scientists and data scientists in the Chemistry, Manufacturing, and Controls (CMC) division focuses on process development for purifying active pharmaceutical ingredients. They use preparative and analytical chromatography in their workflows—fast protein liquid chromatography (FPLC) and ultra-performance liquid chromatography (UPLC), respectively.

The goal is to identify purification methods that maximize drug purity and yield, among other factors. A typical study involves conducting four preparative runs with different FPLC conditions (Figure 1). Around 55 fractions are collected per run, and each fraction is subsequently analyzed by UPLC for purity and yield. Based on the results, scientists select the best preparative run and range of fractions or design a new set of experiments.

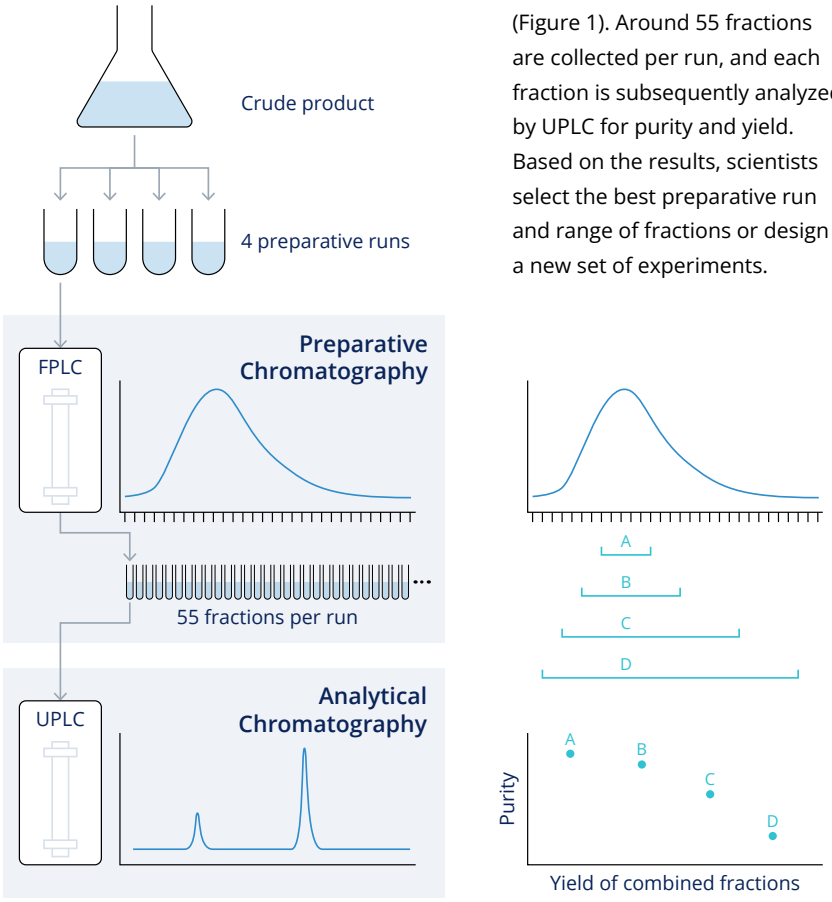

Figure 1. Lab workflow for process development of chromatographic purification. The objective is to maximize yield (quantity of product) while achieving high purity.

Challenge:

CMC scientists at a top 10 biopharma develop preparative and analytical methods for purifying biologics. They use inefficient, manual-intensive practices to assemble and analyze heterogeneous sets of chromatography data.

Solution:

The Tetra Scientific Data and AI Cloud™ replatforms and engineers chromatography data, powering an in-house application for streamlined analysis.

Outcomes:

- Accelerate chromatography process development
- Redirect thousands of hours of manual data handling to model building
- Search and retrieve chromatography data faster and easier
- Streamline troubleshooting during scale-up to facilitate tech transfer
- Generate AI-native data for future AI applications

The Challenge

Each study generates a large and complex set of data, including over 200 chromatograms, that must be aggregated and processed prior to analysis. The primary data sources are the UNICORN chromatography data system (CDS) for FPLC and Empower CDS for UPLC.

The raw data from these sources are not compatible due to proprietary formats and heterogeneous structures. In response, the CMC team previously relied on labor-intensive practices to prepare this data for analysis (Figure 2). They manually gathered, organized, and harmonized results from each experiment. This process consumed the vast majority of their data analysis workload, amounting to thousands of hours per year.

Given the magnitude of this effort and the tight deadlines inherent in drug development, data scientists have had little time to build mechanistic models for process optimization. Such models can greatly accelerate process development and lead to better purification methods.

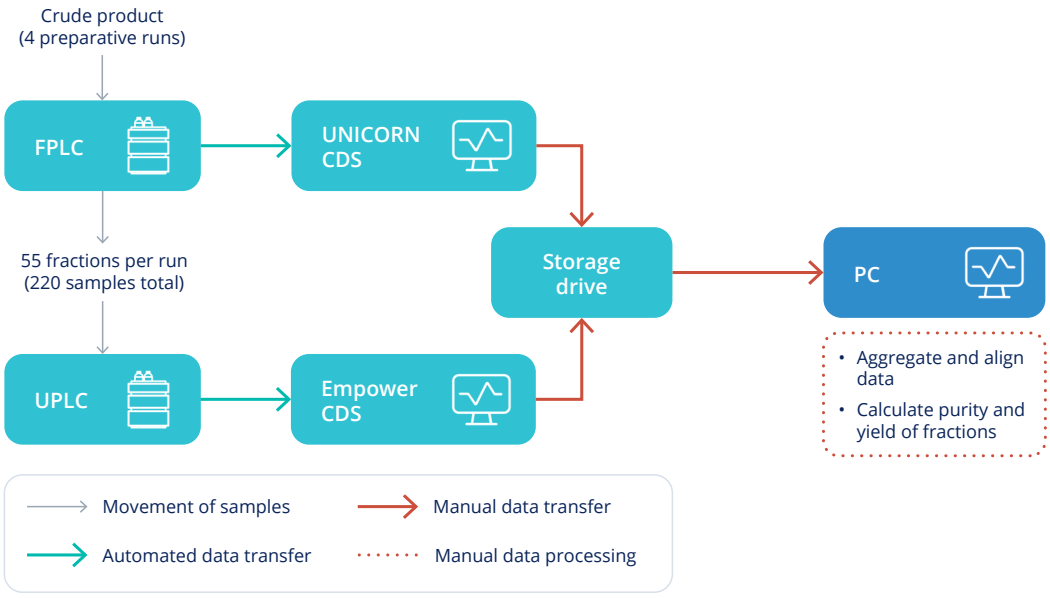

Figure 2. Initial data workflow

The Solution

The CMC group sought to fully automate its chromatography data workflow and turned to TetraScience for help. The resulting solution centers around the Tetra Scientific Data and AI Cloud (Figure 3). The platform automatically centralizes raw data from all sources in the workflow and converts it into an open, vendor-agnostic format. The engineered data, known as Tetra Data, is compliant, large-scale, liquid, and purpose engineered for science. As a result, it is fully optimized for analytics and artificial intelligence (AI).

Leveraging Tetra Data, the CMC team developed an in-house application to analyze the chromatography data. The app retrieves complete datasets from each experiment using contextual metadata added by the Tetra Scientific Data and AI Cloud. It then generates plots, performs analyses, and calculates purity and yield from the harmonized data. With Tetra Data, CMC data scientists can readily build mechanistic models to predict the results of future experiments and recommend optimal processes.

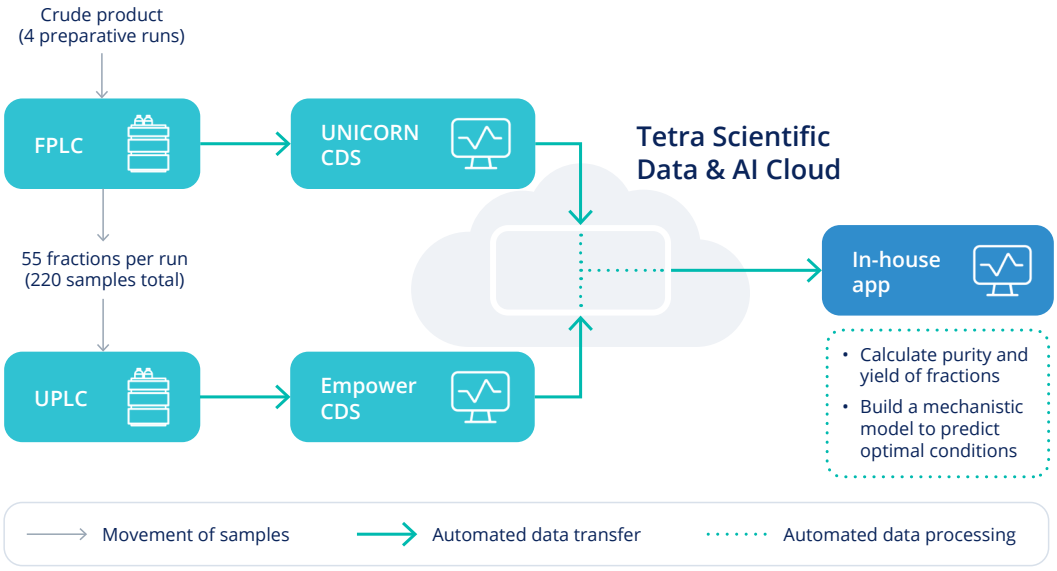

Figure 3. Tetra-enabled workflow

The Outcomes

The Tetra Scientific Data and AI Cloud accelerates the development cycle by clearing barriers faced by the CMC team. The solution eliminates the need for scientists to manually aggregate and prepare large datasets, saving thousands of hours annually. By replatforming and engineering chromatography data, the platform makes analysis far easier and scalable. It also enhances data integrity by minimizing error-prone manual tasks. As a result, the CMC team will base their decisions on more reliable, data-driven insights while ensuring traceability throughout the development process.

Bench and data scientists can now allocate time previously devoted to data handling toward more impactful work, such as developing mechanistic models. With liquid, large-scale Tetra Data feeding into these models, the CMC team will need fewer experiments to find optimal purification conditions. The upshot is faster time to insight, lower costs, and more efficient methods.

Another key outcome is the increased accessibility of chromatography data, as illustrated by two use cases:

1. **Hold-time studies:** Development scientists conduct experiments to assess how long materials can be safely stored between process steps. These investigations involve analyzing chromatograms produced over several days or weeks. With data centralized and contextualized in the Tetra Scientific Data and AI Cloud, scientists can query the database and rapidly find relevant chromatography results.

2. **Scale-up:** When quality issues emerge during scale-up, development scientists need to compare manufacturing data against lab results. However, the CMC group uses a separate, closed network for manufacturing. To retrieve data from this network, scientists had to physically go to the pilot plant and transfer the data onto a USB thumb drive. This process was not only inefficient but also prone to losing valuable contextual information while bearing compliance and data security risks. In response, the CMC team built a data pipeline to move data from the production network to the Tetra Scientific Data and AI Cloud at regular intervals. Now, scientists can quickly retrieve analytics-ready data from a pilot and compare it with relevant data from the lab. The new solution streamlines troubleshooting, saves time, and mitigates risk. Plus, it gives development scientists better access to higher-quality data.

AI readiness

Although the CMC team is not currently using AI for process development, its Tetra Data is inherently AI-native and future-proof. When ready, data scientists can leverage historical and new chromatography data to drive AI-enabled outcomes. The impact could be immense: faster development timelines, slashed costs, and higher-quality therapies.