tetrascience

Increasing efficiency in purification process development

CASE STUDY

At a leading global pharmaceutical company, scientists develop and optimize purification methods for drug manufacturing. The team works with small molecules, peptides, and oligonucleotides. The crude product is first purified by fast protein liquid chromatography (FPLC) and then analyzed by ultra-performance liquid chromatography (UPLC) to assess purity and yield. The scientists also conduct forced degradation studies to evaluate how different purification methods impact chemical stability. In these studies, FPLC fractions are subjected to stress conditions prior to UPLC analysis.

The Challenge

This work is vital for chemistry, manufacturing, and controls (CMC) efforts. However, the scientists were relying on inefficient practices to transfer and compile data (Figure 1). They were manually recording results from Cytiva UNICORN 7 and Waters Empower 3 chromatography data systems (CDSs) into Revvity Signals Notebook, which was a tedious and error-prone process.

Moreover, it was challenging to find data from colleagues working on the same asset, as data was scattered across many folders and systems. FPLC data was primarily stored on instrument workstations, and each scientist maintained separate UNICORN and Empower projects. As a result, the scientists struggled to consolidate analytical data (from UPLC) and connect it with upstream processing data (from FPLC).

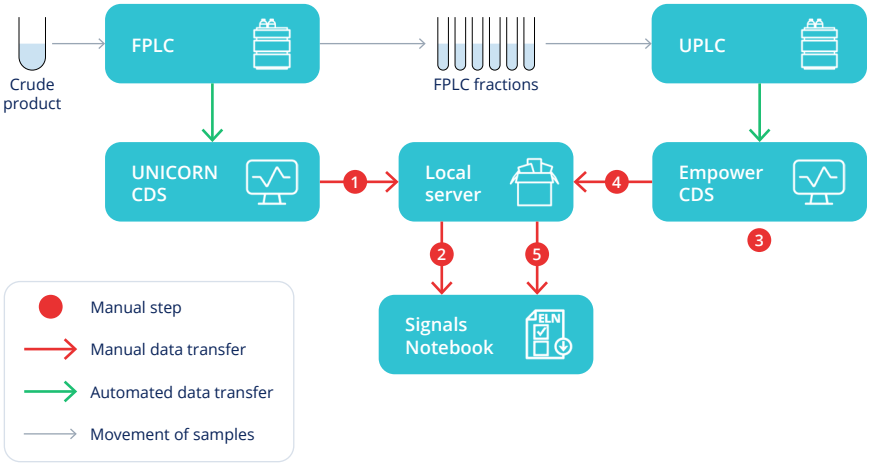

Figure 1. Initial workflow. (1) The scientist manually transfers FPLC data to the server. (2) They transcribe the results into the ELN. (3) They manually enter sample IDs into Empower CDS. (4) After the UPLC run, they manually transfer the data to the server. (5) Then they transcribe the results into the ELN.

Challenge:

Scientists at a top 10 pharma grappled with manual-intensive processes when documenting and assembling chromatography data for CMC activities.

Solution:

The Tetra Scientific Data and AI Cloud™ replatforms and engineers data, enabling a seamless flow of data between UNICORN, Empower, and Signals Notebook.

Outcomes:

- Eliminates 25 to 100 hours per week of manual data transcription
- Reduces data search times by up to 8 hours
- Streamlines report preparation, saving 10 to 20 days per year
- Speeds up process development by predicting optimal purification conditions to maximize yield and purity

This fragmentation, in turn, slowed down data analysis and reporting, particularly when trying to identify anomalies or issues. Compiling data from multiple runs became an arduous task, and gathering large historical datasets for analytical predictions was nearly impossible. Collectively, these inefficient data practices cost an estimated $375,000 annually in wasted scientists' time.

The Solution

The Tetra Scientific Data and AI Cloud streamlines the entire data workflow (Figure 2). FPLC data in UNICORN is automatically captured in the cloud and contextualized with relevant metadata. It is then converted into an open, vendor-agnostic format, resulting in AI-native Tetra Data. Sample IDs for the crude sample and fractions are auto-generated in Signals Notebook, and the sample list is automatically transferred to Empower, avoiding manual transcription by the user. Similar to FPLC data, UPLC data in Empower is replatformed and engineered in the Scientific Data and AI Cloud. Finally, an automated pipeline prepares and sends the results to Signals Notebook.

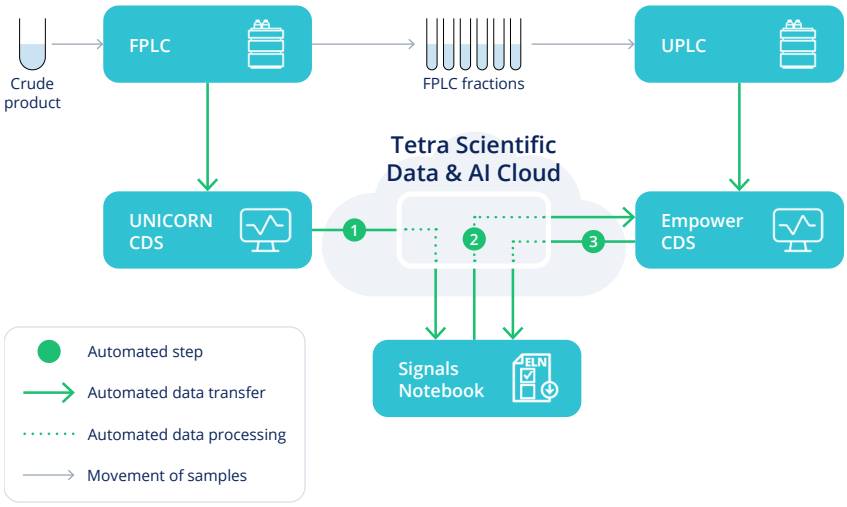

Figure 2. Tetra workflow. (1) The Tetra Scientific Data and AI Cloud automatically ingests the FPLC data, engineers it into Tetra Data, and moves it to the ELN. (2) The ELN auto-generates sample IDs, which are automatically sent to Empower CDS. (3) UPLC data from the CDS is automatically ingested into the cloud, engineered, and sent to the ELN.

The Outcomes

End-to-end automation significantly frees up scientists' time, allowing them to focus on more critical CMC activities. It also enhances data integrity by reducing the risk of human errors. Key outcomes of this new approach include:

- The new workflow eliminates **25 to 100 hours per week** of manual data transcription between the CDSs and the electronic lab notebook (ELN).
- With a centralized repository of contextualized data, scientists can more easily find chromatography data—including raw data, peak tables, and metadata—from colleagues, saving **4 to 8 hours per search**.
- Scientists can reclaim **10 to 20 days per year** that were previously spent transcribing data into forced degradation technical reports.

The Tetra solution streamlines method development and expedites the path to clinical trial readiness. Scientists can now compare multiple runs effortlessly, identify anomalies immediately, and access historical data instantly. They are much better equipped to make data-driven decisions and extract actionable insights from the entire team's extensive dataset.

AI Readiness

By replatforming and engineering chromatography results into Tetra Data, this top 10 global pharmaceutical company is producing the kind of data essential for advanced analytics and AI. These liquid, large-scale, compliant, and purpose-engineered datasets can fuel predictive models. For instance, scientists can leverage analytics and AI tools to predict the optimal FPLC conditions that maximize yield and purity, significantly reducing the need for wet lab experiments. This results in accelerated development timelines and lower costs.