Chiral Screening Workflows: A Proof of Concept Use of ADF in Non-Linear Laboratory Workflows

Allotrope Foundation

CASE STUDY

Goals

- Partner with instrument vendors to model non-linear workflows which mirror those used often during pharmaceutical development
- Demonstrate the ability to convert proprietary vendor data formats to ADF format, focusing on LC-UV and SFC-UV as key instrument examples
- Demonstrate the ability to extract key meta data according to Allotrope controlled vocabulary into a repository
- Demonstrate that use of ADF facilitates automatic transfer of meta data, reducing costs and error potential

Challenges

- Consensus LC-UV and SFC-UV taxonomies, ontologies, and data models did not exist
- The semantics concepts (RDF triples) represented extremely novel technology for scientists, instrument vendors, and software engineers
- Synchronization of ADF files updates as additional metadata throughout the workflow had to be solved

Results

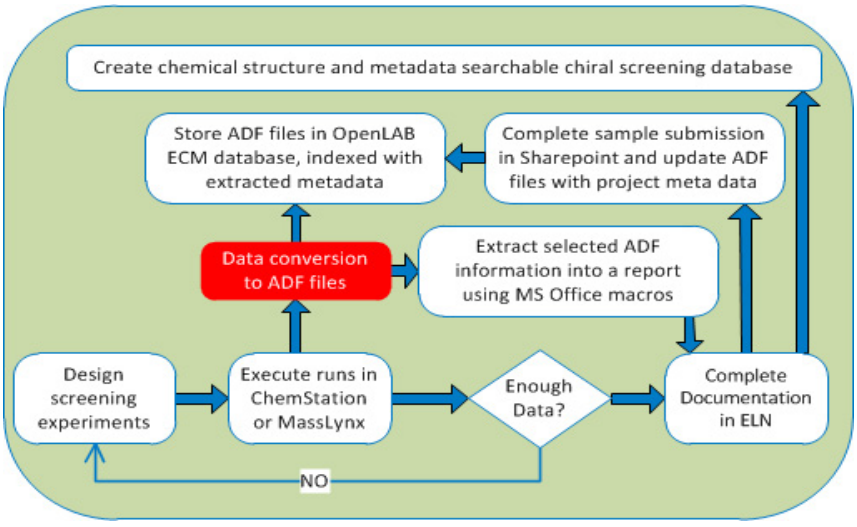
- Robust taxonomies, ontologies, and data models in CMAPs were created for LC-UV and SFC-SQD
- Agilent and Waters created proof of concept (PoC) ADF converters for LC-UV using their products ChemStation and MassLynx, respectively
- Agilent created an OpenLab ECM filterpack to extract key metadata and index ADF files in the repository
- ACD Labs created a software solution to dynamically update ADF files with project metadata from Sharepoint
- ACD Labs created a chiral screening database designed to be structure or metadata searchable from the Merck search engine
- Osthus created Pipeline Pilot scripts and MS Office macros to extract information from ADF files for automated creation of screening reports for storage in ELN

Background

High performance liquid chromatography (HPLC) is one of the most prevalent techniques employed in the pharmaceutical development laboratory, with uses spanning from identification to quantitation, to detailed characterization of both chiral and achiral materials. Supercritical fluid chromatography has also gained widespread use as a rapid and powerful means to separate chiral molecules, especially in a higher throughput or screening format where initial experimentation is prescriptive based upon chemical knowledge and prior separations experience. However, as with all aspects of research and development, screening workflows rapidly become non-linear when initial screening experimentations don't yield all of the results which were desired, and follow on experiments are planned. It is critical for the scientist to be able to associate all these available experimental results with experiments and materials which produced them, to both identify key separations trends and inform the successful design of future experiments.

Chiral screening protocols have been advanced over the years which have defined combinations of chiral stationary phases and mobile phases proven to be most successful to achieve chiral selectivity. While chiral screening protocols have standardized some of the experiments run, the resulting data exist in a myriad of proprietary vendor formats which are not inter-operable, often resulting in data "dead ends". To maximize the value of chiral screening, these results should use a single data format and common terminology regardless of which vendor's instrument was used and which vendor's chromatography data system (CDS) was used to acquire the chromatogram. With adoption of the Allotrope Framework, results would exist in a common data format and indexed with common terminology, permitting limitless potential to rapidly retrieve and aggregate the data for further uses such as analytics, trending, and automated report generation. The project described here uses the Allotrope Framework in a proof of concept study to address this workflow.

…chiral screening protocols have standardized some of the experiments run, however the resulting data exist in a myriad of proprietary, non-interoperable vendor formats which represent data "dead ends".



This material was prepared by Allotrope Foundation, which is a not-for-profit association of companies developing a standardized framework for analytical laboratory data and information management. For details, please contact the Secretariat (more.info@allotrope.org).

Chiral Screening Workflows: A Proof of Concept Use of ADF in Non-Linear Laboratory Workflows

Allotrope Foundation

CASE STUDY

Project Overview

An integration projection was developed through close partnership between Agilent, Waters, Osthus, ACD Labs, and Merck and Co, Inc designed to address the basic chiral screening non-linear workflow depicted. In current state, chiral samples were either analyzed with an Agilent 1200 HPLC-UV using ChemStation software, or with a Waters UPC2 SFC-UV using MassLynx software until such time as an acceptable separation was achieved. Final chromatograms generated are archived in OpenLab in native format, which renders them unreadable unless opened in the native software. The results of the chiral screening experiments are then manually summarized manually in ELN, with no direct link to the raw data in OpenLab.

Working with semantics experts from Osthus, subject matter experts from Agilent, Waters and Merck identified key taxonomy current used in each vendors commercial product and transformed this information into the single vocabulary needed for LC-UV and SFC-UV examples. With significant contributions of expertise from the LC-UV working group, these vocabularies were transformed into ontologies in RDF format and ultimately, semantically-correct data models using Cmaps for these techniques.

Results

Software engineers from Agilent and Waters leveraged the ontologies and data models to create proof of concept (PoC) ADF converters for LC-UV using their products ChemStation and MassLynx, respectively. The process was highly collaborative and iterative as new terms or relationships were identified, and correspondingly data models updated.

Impact of Case Study

The LC-UV chiral screening case study demonstrates the feasibility of creating a workflow which is enabled end-to-end by Allotrope data standards and ontologies. This will help scientists to more efficiently access and re-use their data, which in turn will drive better knowledge sharing and decision making.

Initial example ADF files have been created focusing on population of critical instrument parameters, integrated results, and audit trail information in the data description layer; raw chromatographic results were transferred to the data cube; and finally native format raw data to the data package. These ADF files have been shared with the Allotrope Foundation and the Allotrope Partner Network to accelerate application of the framework.

For the purposes of the above PoC study, Agilent's OpenLab ECM was used as the data repository, mimicking the function of a future data cloud environment. Agilent created an OpenLab ECM filter pack to extract defined key metadata from these example ADF files, and use to index the files to assist future rapid data mining applications (searching an entire ADF file would be time-consuming).

ACD Labs created a software solution based upon their existing ACD/Spectrus platform, to dynamically update ADF files in the OpenLab ECM platform with new project metadata acquired throughout the course of the chiral screening workflow. This mirrors real-world situations, where all metadata is not available at the outset of an experiment, or when new metadata is added to a legacy experiment as it is "repurposed" for a new use.

For the purposes of the PoC, ACD Labs demonstrated updating ADF files with metadata from Sharepoint, including chemical structure. Additionally, ACD Labs created a structure and metadata searchable chiral screening database, designed to be queried from the Merck search engine, which pools data collected during a screening experiment using common metadata labels.

Osthus used Pipeline Pilot scripts and MS Office macros to extract metadata from ADF files in OpenLab ECM and create automated reports to document key outcomes of chiral screening experiments. These reports are added to an electronic notebook and are similarly searchable from the Merck search engine, leveraging the common metadata used throughout the case study.

About Allotrope Foundation

Allotrope Foundation is an international consortium of pharmaceutical & biopharmaceutical companies launched in 2012 with a common vision to develop innovative approaches for handling scientific data. Allotrope Foundation has developed a framework to capture and represent data generated by any analytical device in the laboratory in a standardized format, including more complete metadata related to each test and measurement event, expressed in a standardized vocabulary, which facilitates the exchange, utilization and integration of data beyond the boundaries of the originating instruments and laboratories.

This effort is fully funded by the members of Allotrope Foundation and is rapidly progressing on our common goals to improve data integrity, reduce wasted effort and allow us to realize the full value of our scientific data.

Allotrope Foundation
1500 K Street, NW
Washington, DC 20005
(202) 230-5855
http://www.allotrope.org