High-throughput analytical workflows that reduce time and resource needs to enable synthetic biology research **WBS# 2.5.3.702**

Alex Apffel

Agilent Research Laboratories

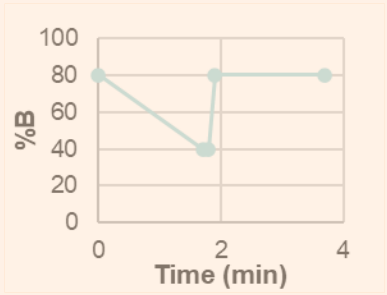
Project Overview

Goal: to use expertise at PNNL, LBNL, and SNL to solve a challenge for Agilent - how to advance and expand their metabolomic workflows

- **Mission of project**: To couple powerful Mass Spectrometry platforms (QQQ, IM-QTOF-MS) with Agilent's novel liquid chromatography (UHPLC) fast metabolomic workflows and perform machine learning to datasets generated.
- **Current Limitations**: Current analytical approaches for Synthetic Biology are too slow, limited to < 30 samples/day with time consuming method development and optimization.
- **Technical Approach**: Combination of ultrafast UHPLC-MS methodologies and machine learning based method selection to increase throughput and eliminate development/optimization time.
- **Justification**: Rational optimization of synthetic biology production yield requires measurement of pathway intermediates for a large number of strain variants to identify bottlenecks and carbon sinks.
- **Technical Risks:**
  - Introduction of modified workflows and data processes must merge seamlessly with existing functional infrastructure.
  - Integration and actionable conclusions based on metabolic pathway intermediate abundances is a challenging process that also must be addressed.
  - COVID-19 related multisite remote collaboration has NOT proven to be limiting or problematic.

PROJECT GOALS AND STRUCTURE

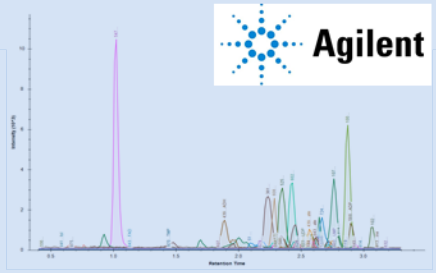
1. Implement Agilent Ultra-fast HPLC methods



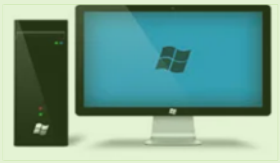
2. Develop methods for small molecular detection and quantification on Agilent MS instruments

QQQ

IM-QTOF



3. Establish data analysis and sharing pipeline

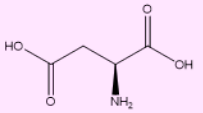


EDD

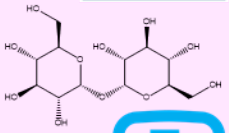
jupyter

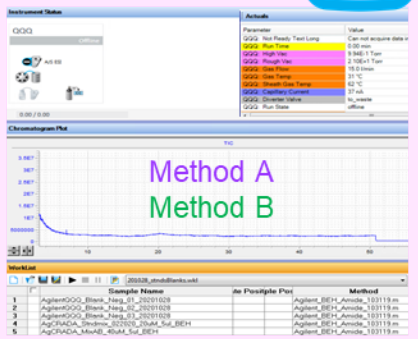
4. Implement and expand Agilent's Automated Method Selection software

Compound X



Compound Y





Machine Learning applied to datasets generated

1. Management: Organizational structure

Task Responsibilities

- ***Agilent Research Laboratories:*** Alex Apffel, Anya Tsalenko
  - *Optimization and Transfer of Analytical Methodologies*
  - *Implementation of new Physico-Chemical Parameter Prediction Algorithms for Method Selection*
  - *Validation of new Physico-Chemical Parameter to existing machine learning models.*
- ***PNNL:*** Kristin Burnum-Johnson, Jon Magnuson
  - *Method implementation*
  - *Data Generation and Processing for triple quadrupole (QQQ) mass spectrometers*
  - *Data Generation and Processing for ion mobility (IM-QTOF) mass spectrometers*
- ***LBNL:*** Chris Petzold
  - *Method implementation*
  - *Data Generation and Processing for QQQ*
  - *Data Management and Curation*
- ***SNL:*** Kunal Poorey
  - *Transfer of Physico-Chemical Parameter Prediction Algorithms to Agilent Labs*
  - *Development and testing of Machine Learning Approaches to Method Selection.*

1. Management: Risk Abatement

Team management:

- Leads are experts in their fields: Novel high-throughput analytical workflows (Agilent), mass spectrometry instrumentation and data analysis (PNNL & LBNL), and Machine learning (Agilent & SNL)
- Contributions from all labs to all teams
- Shared project team members from Agile BioFoundry ensure well-designed target-host applications

Team meetings:

- Biweekly meetings with all team members

Project technical risk mitigation:

- The project team is incorporating and advancing Skyline data processing workflow for all mass spectrometry measurements
  - All raw data and results are shareable
  - These modified workflows and data processes steps merge seamlessly with existing functional infrastructure
- The project team is utilizing purchased standards to demonstrate new analytic workflow and add confidence to target-host measurements

U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

2. Approach

Approach overview

- ***The general approach is to integrate analytical methods and automated method selection algorithms developed by Agilent Research Laboratories into a functional and established analytical pipeline operating in the Agile BioFoundry (ABF) test laboratories (PNNL and LBNL) to improve efficiency and throughput without disrupting or sacrificing current capabilities or performance.***
- ***Several technical challenges have been encountered (and overcome):***
  - *Original definitions and studies with high speed UHPLC-MS methods and method selection were conducted with QTOF type mass analyzers. Adaptation to IM-QTOF required modification of automated method generation procedures as well as production of new types of data. While the method generation procedures have been successfully complete, the value and potential of the new data types are still under evaluation.*
  - *During testing of new physico-chemical parameter estimation algorithms from SNL, compound nomenclature ambiguities due to protonation states were discovered. A new module was generated by Agilent Research Labs to address this ambiguity and exploit the added specificity with respect to compound charge state.*
  - *To fully test machine learning based method prediction models, a comprehensive matrix of all methods vs all compounds would be ideal. In the absence of such comprehensive data, partial validation is possible and incremental learning methodologies need to be implemented.*

Automated Method Selection

List of All Metabolite Targets in Each Sample

Unambiguous Target Structures

Spectral Database Searching for previous methods

Physicochemical Properties for Each Target

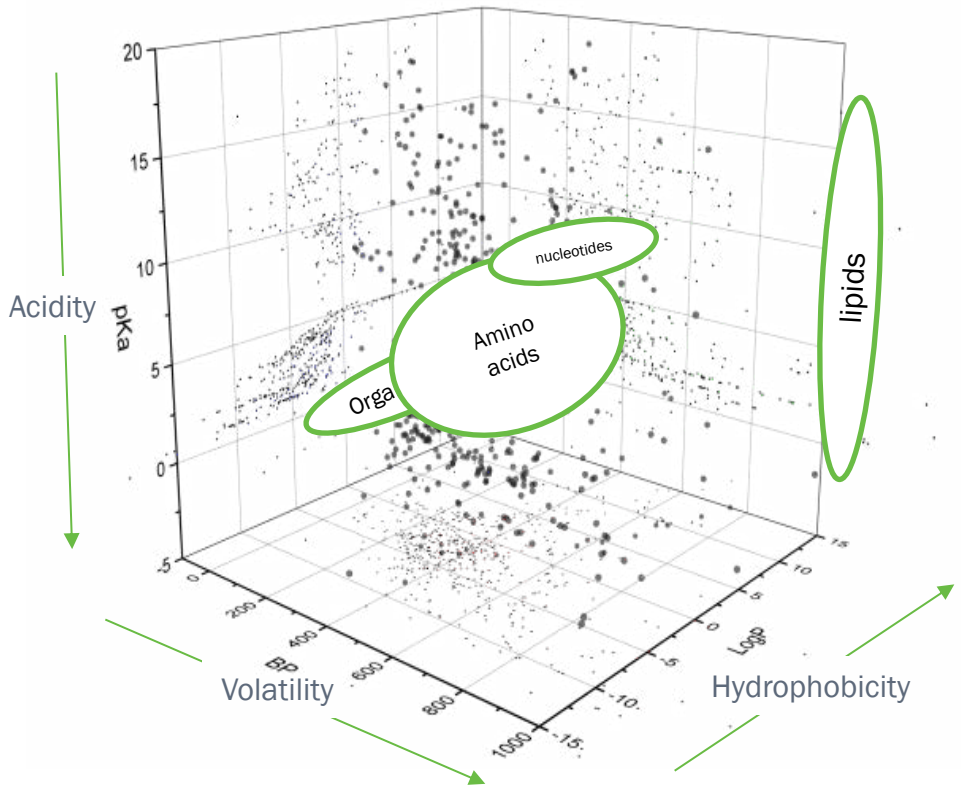
Method Scores for Each Target for Each Method

Minimal Set of Methods for Each Sample

Instrumental Worklist specifies Method(s) for Each Sample

Agilent

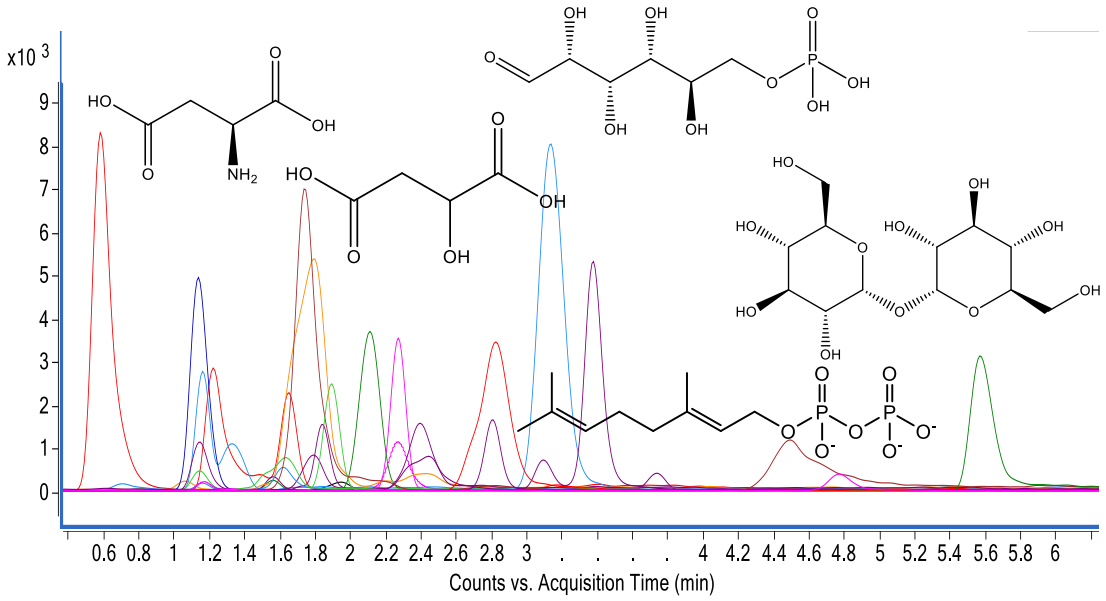
- Many chemical classes
- Hypothesis driven targeted analytical approaches
- Novel compounds require analytical flexibility



Demonstration

Demonstrate Agilent Methods on LBNL's and PNNL's QQQ mass spectrometers

> 50 standards were selected and analyzed by LC-MRM in a QQQ, to optimize a fast LC separation method using protocols shared by Agilent

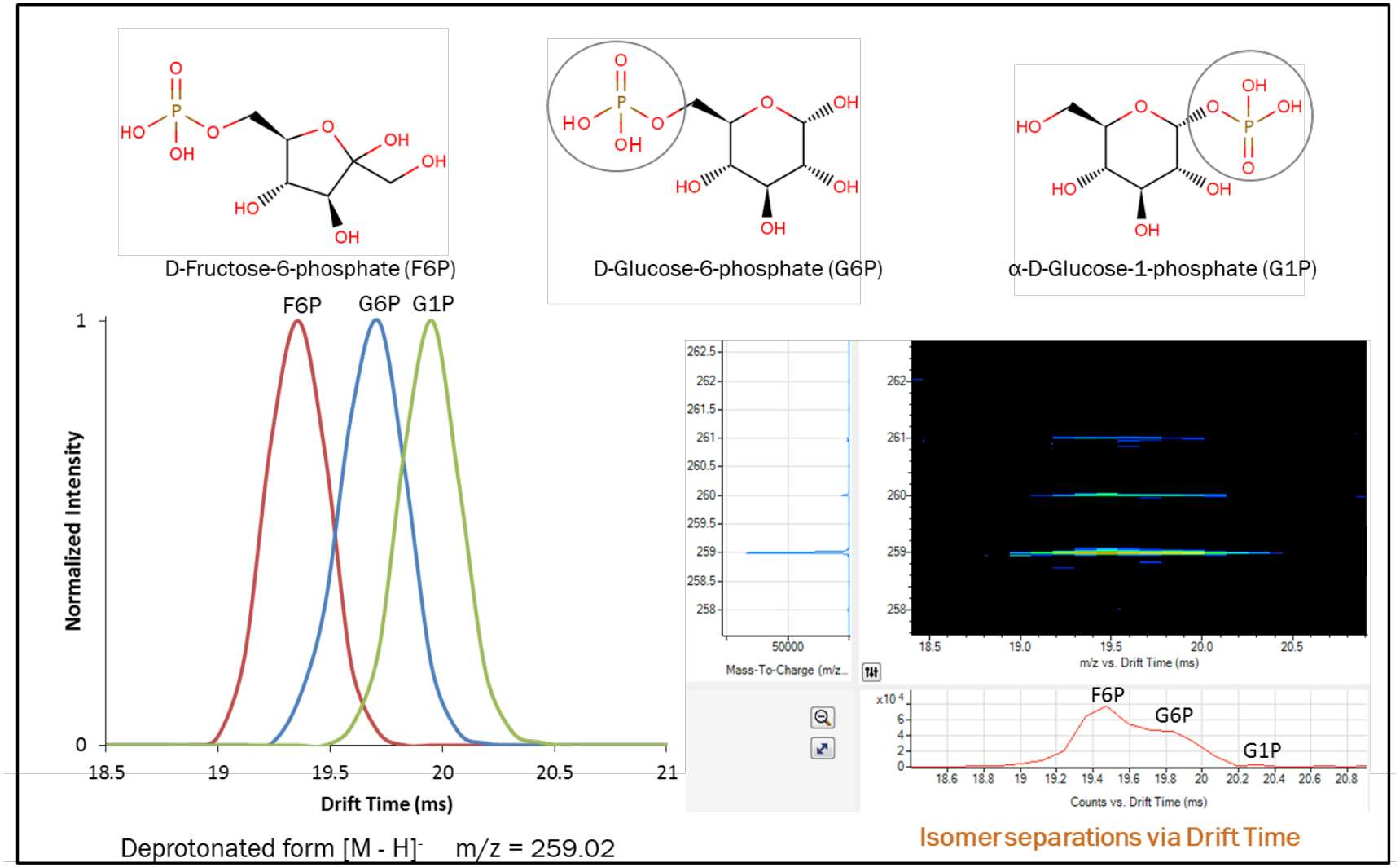


Metabolite standards are derived from diverse pathways

- Embden-Meyerhof-Parnas (EMP) pathway
- peripheral oxidative pathways
- pentose phosphate pathway
- Entner-Doudoroff pathway
- Tricarboxylic acid cycle
- Anaplerotic pathways
- EDEMP cycle

Ion Mobility Mass Spectrometry

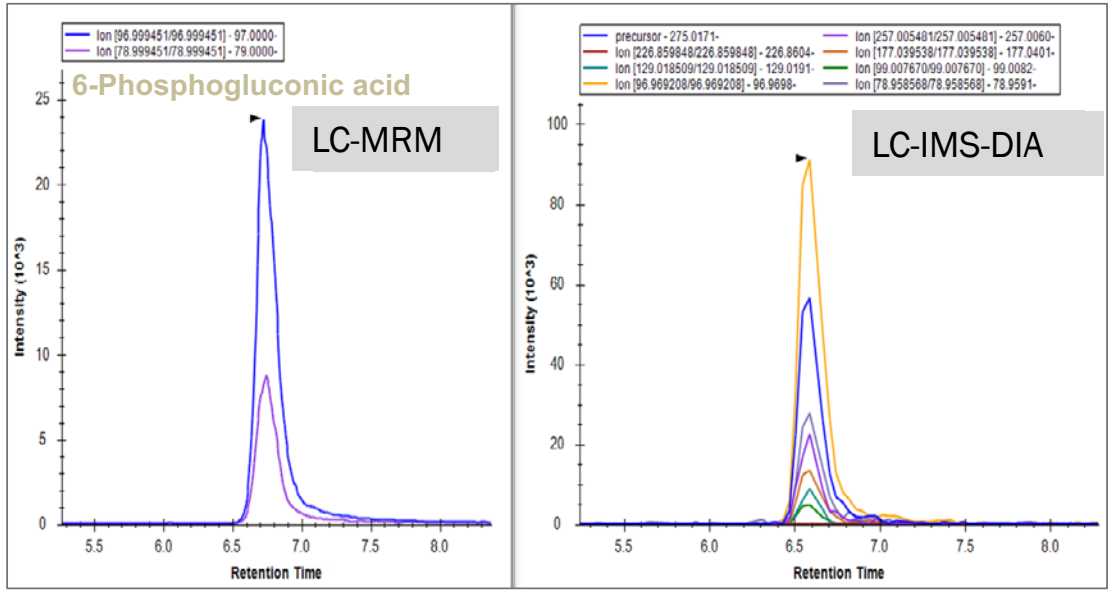
Utilize ion mobility for improved efficiency and throughput of metabolomic measurements

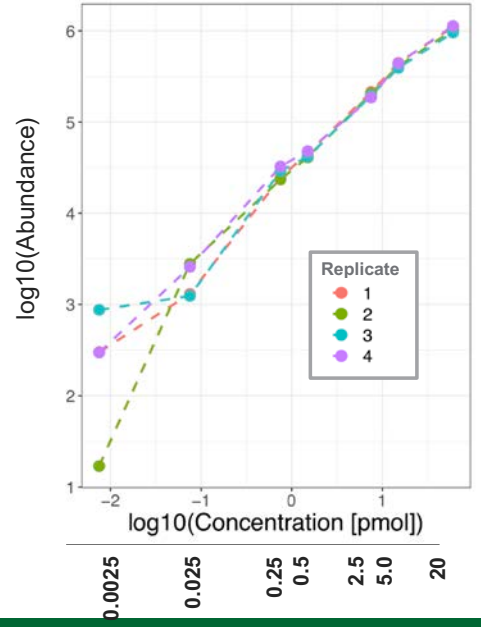


The Agilent 6560 Ion Mobility QTOF was developed through a previous PNNL/Agilent collaboration

Demonstrate Agilent Methods on PNNL's ion mobility platform

- Standards were acquired in a UHPLC-ion mobility MS system.
- Library established with 60 compounds (retention time, exact mass, collision cross section, MS/MS).
- For a selected group of metabolites, calibration curves were evaluated. Data linearity and reproducibility was observed across 4 orders of magnitude.
- For 60% of standards ion mobility-data independent acquisition (DIA) provided more transitions determined post-acquisition for increased detection confidence.



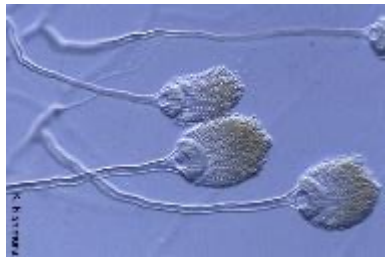


Ion mobility metabolomic analyses of ABF Target-Hosts

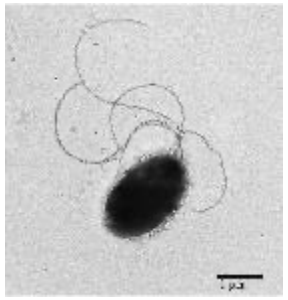
Go/No Go Milestone: (PNNL) Perform high-throughput UHPLC-ion mobility MS analyses. More than 100 analysis of samples from ABF host microorganisms were acquired in PNNL's Agilent UHPLC-ion mobility mass spectrometer. Samples were analyzed using the LC methods developed and optimized by Agilent and ion mobility methods developed and optimized by PNNL. Data was processed with a software pipeline developed by PNNL and results were uploaded to Experiment Data Depot (EDD), a crucial element on ABF infrastructure.

- ***A library of more than 50 metabolites (that are relevant to ABF hosts due to being part of central carbon metabolism or pathways that have been engineered to generate a bioproduct) was established. Metabolites were characterized with retention time, exact mass, collision cross section and MS/MS.***
- ***A total of 116 samples of ABF hosts were acquired in the ion mobility instrument.***

Organism	Acquisition mode (Collision energies)	Number of runs	Acquisition date
Pseudomonas putida	2 (20 and 40 V)	22	8/1/20
Aspergillus pseudoterreus	2 (20 and 40 V)	30	8/25/20-8/26/20
Aspergillus niger	2 (20 and 40 V)	64	8/25/20-8/26/20
Total		116	



Aspergillus pseudoterreus



Pseudomonas putida KT2440

Agile BioFoundry

U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

3. Impact

Impact overview

Within the context of a Biofoundry, rapid DESIGN-BUILD-TEST-LEARN Cycles are required to meet the need to address increasing numbers of new production targets in decreasing amounts of time. One of the key challenges of optimizing this cycle is to improve the efficiency and impact of rapid LEARN Cycles, which, in turn can only be as good as the quantity and quality of TEST data allows.

The impact of this project is to allow the generation of more high-quality analytical test data to feed the LEARN process by directly increasing sample throughput and eliminating manual method selection and optimization time.

With the implementation of ion mobility and data independent acquisition (IM-DIA) analysis, data from all detectable analytes and transitions are automatically acquired. The subset of the best transitions can be determined post-acquisition and are utilized by specialized DIA software to detect metabolites with increased confidence.

Dissemination of the methods and results of these developments will allow ready adoption of high throughput methodologies by the synthetic biology community.

The software infrastructure is based on open-source tools. Resources (Skyline projects and IMS library) were shared through Agilent's sharepoint and results were uploaded to EDD. A publication (in preparation) will describe the complete software workflow. The created library can be applied for metabolite identification in IM-DIA data from other relevant samples.

Impact to the research community

Benefits from the success of this project

- The Agile BioFoundry will implement Agilent's novel method selection software in everyday analyses. This will increase throughput by reproducible and accurately selecting the **minimal set of methods for each target-host sample to characterize metabolites in pathways of interest.**
- Agilent will advance the utility of their analytical workflows to new mass spectrometry platforms (QQQ & IM-MS) and advance their prediction models with machine learning.
- This fills a great need in the research community for high-throughput analytical workflows that reduce time (i.e., efficient DESIGN-BUILD-TEST-LEARN Cycles) and resource needs to enable synthetic biology research.

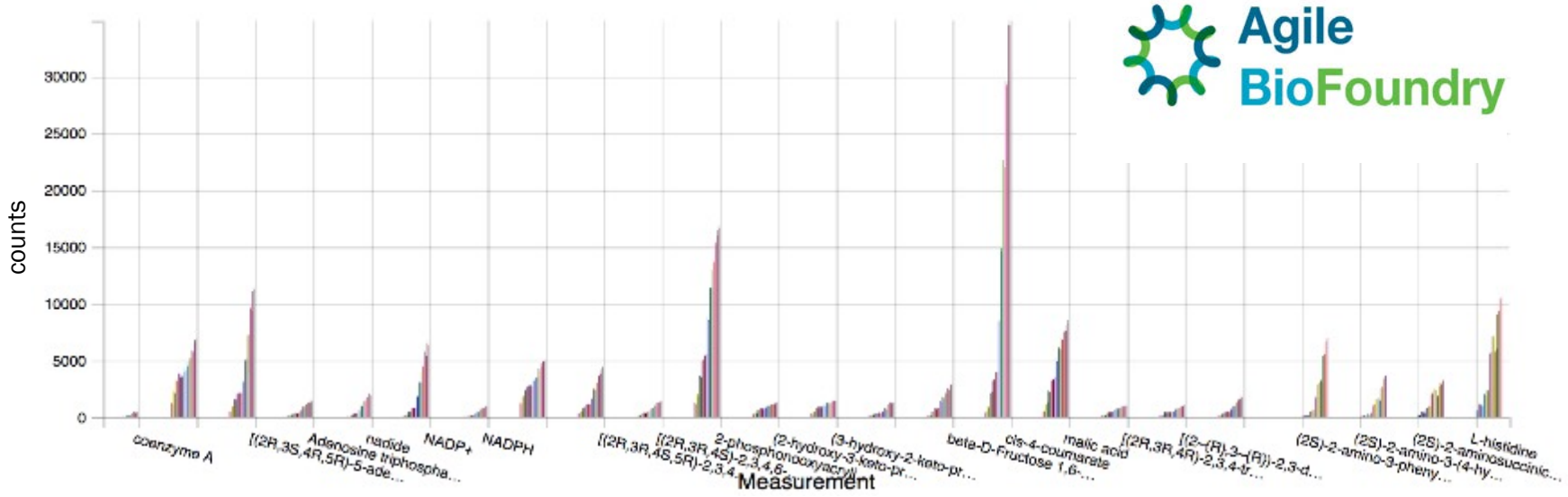
U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

4. Progress and Outcomes

Data deposit and sharing on the ABF's Experiment Data Depot (EDD)

- EDD is a publicly available online tool designed as a repository of experimental data and metadata.
- EDD can uptake experimental data, provide visualization of these data, and produce downloadable data in several standard output formats.



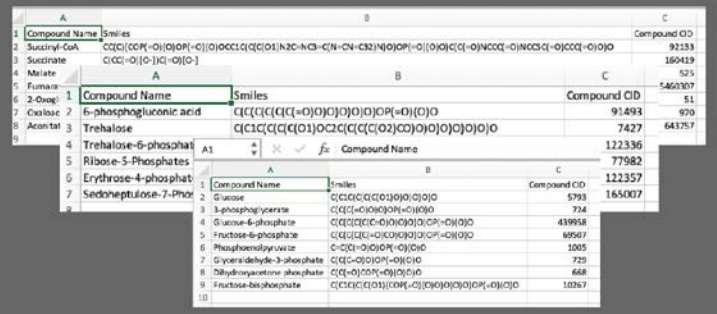
We have successfully implemented a workflow to incorporate Agilent's novel method selection software and fast analytical methods in everyday Agile BioFoundry analyses.

Automatic Analytical Workflow: Agilent's method selection software ➡ Fast analytical liquid chromatography methods ➡ high-quality mass spectrometry data ➡ data analysis ➡ data sharing and visualization on EDD

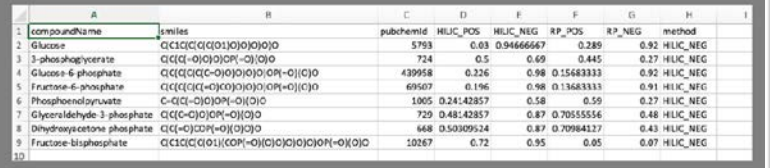
EDD is compatible with Automatic Analytical Workflows-QQQ metabolic measurements

- (Agilent/LBNL/SNL) Demonstrate automated method selection, data acquisition, and data processing.
- Detailed steps in the integration of Agilent's Automatic Analytical Workflows (AAW), with Experimental Data Depot (EDD)

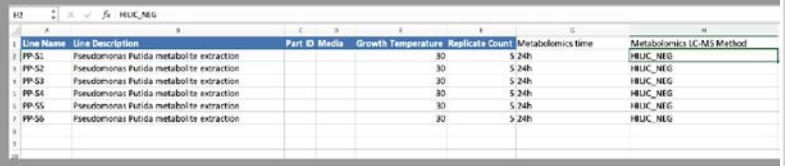
Step 1: Provide compound names and SMILES IDs for pathway intermediates as input to the AAW software



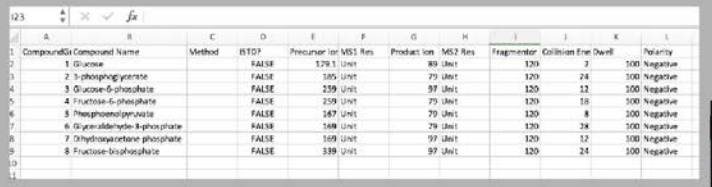
Step 2: AAW software output provides a prediction of which LC-MS method to use

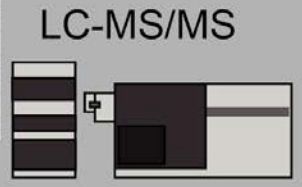


Step 3: Include LC-MS method prediction in EDD experiment description file



Step 4: LC-MS MRM method parameters for pathway compounds





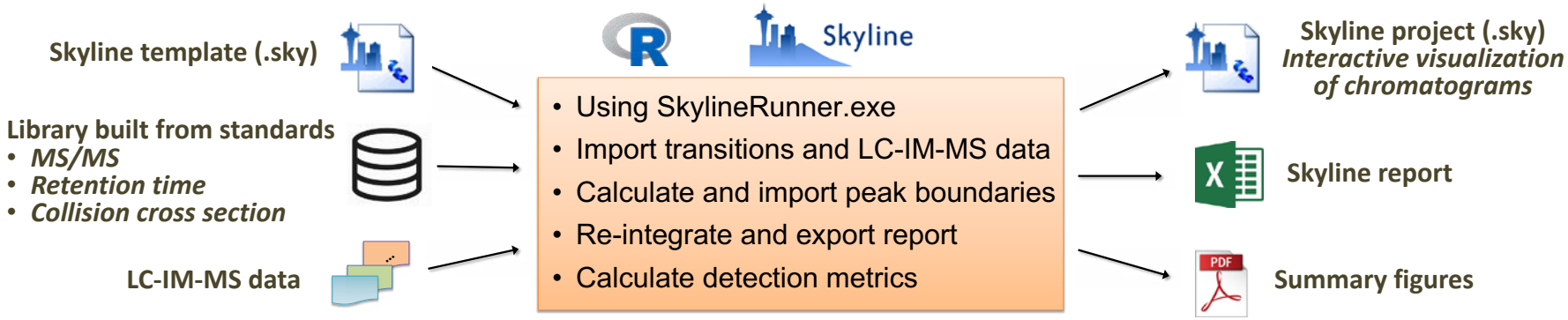
Step 5: Upload processed LC-MS data to the EDD

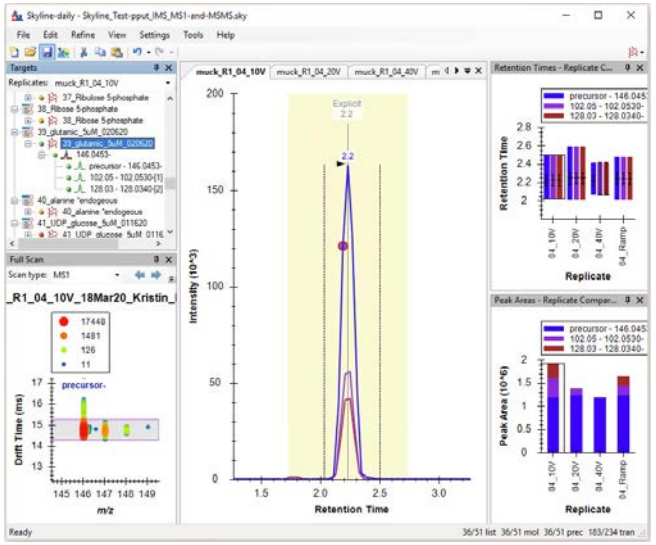
Our Analytical Workflow works with QQQ mass spectrometry data

Automatic Analytical Workflow: Agilent's method selection software ➡ Fast analytical liquid chromatography methods ➡ high-quality mass spectrometry (QQQ) data ➡ data analysis ➡ data sharing and visualization on EDD

EDD is compatible with Automatic Analytical Workflows-ion mobility metabolic measurements

- (Agilent/PNNL) Perform high-throughput UHPLC-ion mobility analyses
- Develop a data analysis workflow compatible with EDD.





- All shareable: raw data and results.
- Skyline projects can be opened from any computer.
- Projects shared in Agilent's Sharepoint.
- Studies corresponding to samples analyzed in this platform created on EDD.

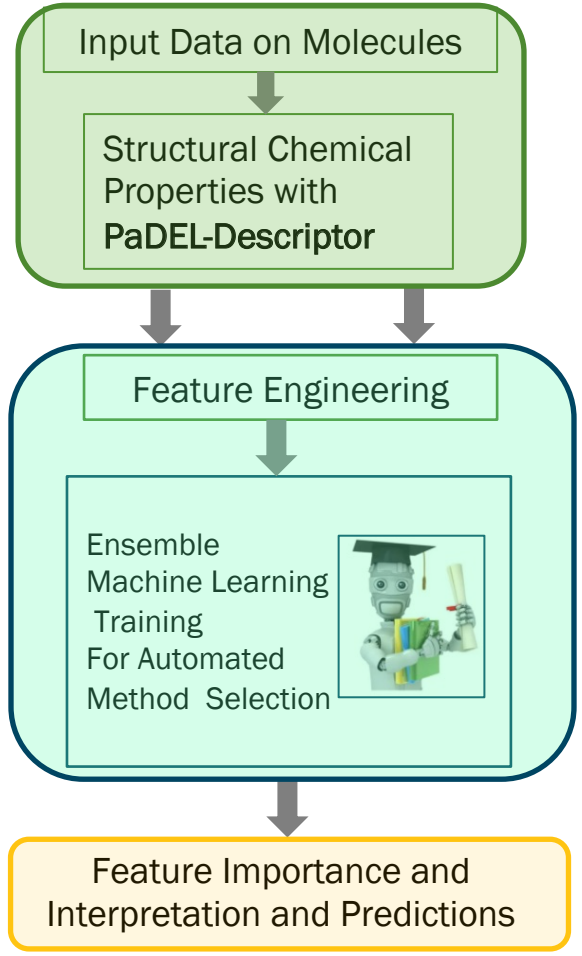
Our Analytical Workflow works with ion mobility mass spectrometry data

Automatic Analytical Workflow: Agilent's method selection software ➡ Fast analytical liquid chromatography methods ➡ high-quality mass spectrometry (ion mobility) data ➡ data analysis ➡ data sharing and visualization on EDD

Machine learning will advance Automatic Analytical Workflows

(Agilent/SNL) Use machine learning techniques to include additional physical/chemical properties and demonstrate model predictions.

ML Strategy Summary



SNL's custom machine learning software pipeline will highlight

- PaDEL descriptor for enhancing the physical-chemical properties of the dataset
- Apply ensemble machine learning strategy to pick the best performing model for method selection

We are now incorporating SNL's machine learning approach to advance our method selection software

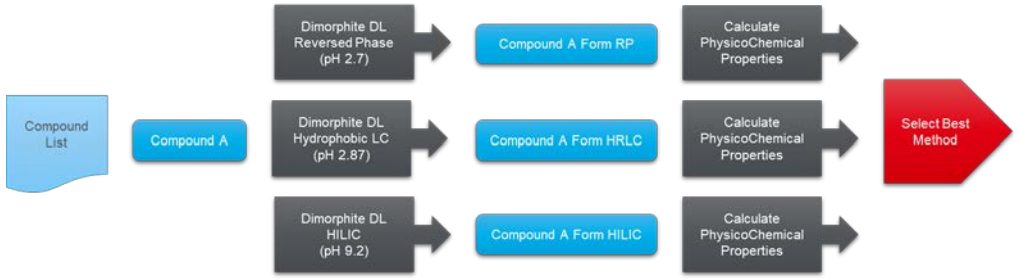
Automatic Analytical Workflow: Agilent's method selection software ➡ Fast analytical liquid chromatography methods ➡ high-quality mass spectrometry data ➡ data analysis ➡ data sharing and visualization on EDD

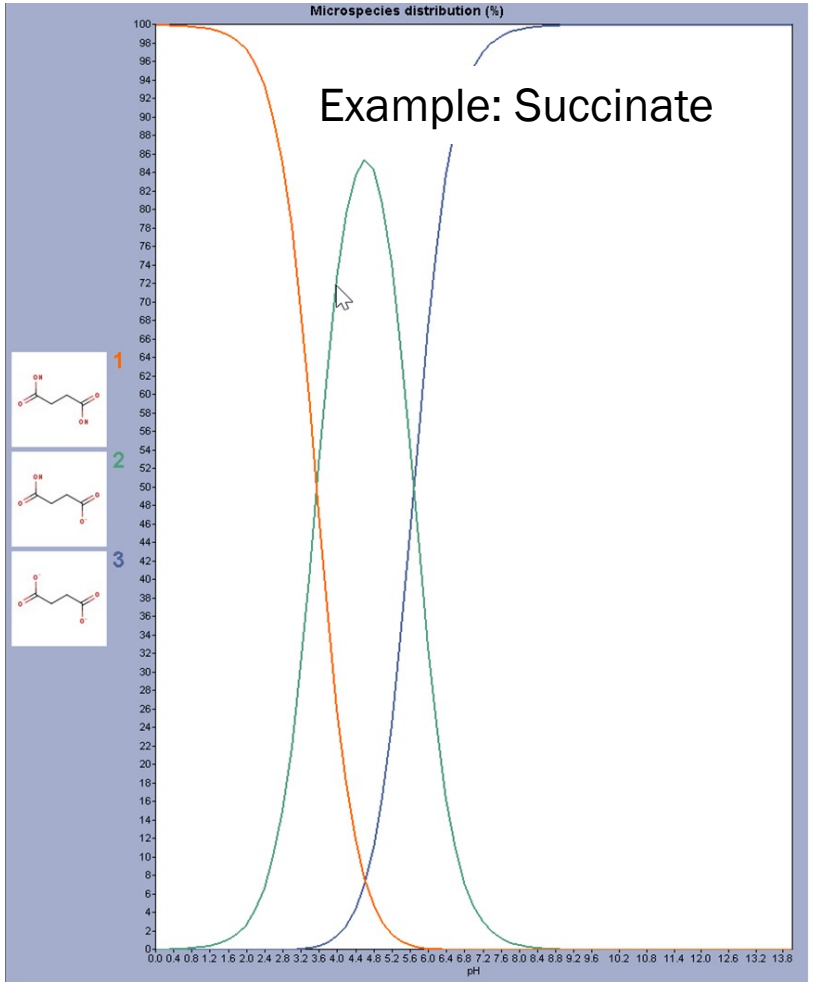
Machine learning will advance Automatic Analytical Workflows

(Agilent/SNL) Use machine learning techniques to include additional physical/chemical properties and demonstrate model predictions.

- **Recent Agilent Contributions**

1. Incorporation of algorithms to account for method specific pH dependent target protonation states.





We are now incorporating Agilent's machine learning approach to advance our method selection software

Automatic Analytical Workflow: Agilent's method selection software ➡ Fast analytical liquid chromatography methods ➡ high-quality mass spectrometry data ➡ data analysis ➡ data sharing and visualization on EDD

Progress and Outcomes

Milestones	Due	% fin	Issues if any
Milestone 1: (PNNL, Agilent) Transfer Agilent Methods	9-30-19	100%	
Milestone 2: (LBNL, Agilent) Transfer Agilent Methods	9-30-19	100%	
Milestone 3: (LBNL, Agilent) Implement Agilent Methods	2-28-20	100%	
Milestone 4: (PNNL, Agilent) Demonstrate Agilent Methods	5-31-20	100%	
Milestone 5: (LBNL, SNL, Agilent) Demonstrate automated method selection, data acquisition, and data processing.	8-31-20	100%	
Milestone 6: (PNNL, Agilent) Perform high-throughput UHPLC-ion mobility MS analyses (GN)	8-31-20	100%	
Milestone 7: (SNL, Agilent) Use machine learning techniques to include additional physical/chemical properties and demonstrate model predictions.	2-28-21	80%	Testing physico-chemical properties including protonation state.
Milestone 8: (PNNL, LBNL, SNL, Agilent) Contribute and utilize data, specifications, for a finalized Automated Method Selection tool	6-30-21	20%	

Summary

Key Takeaways
Agilent has shared their ultra-fast HPLC methods with LBNL and PNNL
LBNL and PNNL implemented Agilent's HILIC HPLC methods for metabolite standard
LBNL and PNNL have developed methods for small molecule detection and quantification on Agilent MS instruments (QQQ/ IM-QTOF)
PNNL developed a data analysis workflow for samples acquired in the IMS platform compatible with existing data management infrastructure (EDD)
PNNL acquired and analyzed 100 ABF samples in the UHPLC-ion mobility MS system
LBNL, SNL and Agilent demonstrated automated method selection, data acquisition, and data processing.
Publications, presentations, and IP
Bilbao et al. Poster presented at the 68th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics (virtual June 1 - 12, 2020)
Bilbao et al. Lightning talk presented at the Skyline User Group Meeting (virtual May 27 - 28, 2020)
Manuscript for UHPLC-ion mobility MS workflow (in preparation; PNNL, Agilent, LBNL, SNL)

Quad Chart Overview

Timeline

- Project start date: June 2019
- Project end date: June 2021

	FY20 Costed	Total Award
DOE Funding	496K	800K
Project Cost Share	111K	240K

Project Goal

To couple powerful MS platforms (QQQ, IM-QTOF-MS) with Agilent's novel UHPLC fast metabolomic workflows and perform ABF ML to datasets generated.

End of Project Milestone

Contribute and utilize data, specifications, for a finalized Automated Method Selection tool

Project Partners

- PNNL
- SNL
- LBNL

Funding Mechanism

2017 ABF Directed Funding Opportunity

Acknowledgments

Agilent

- Alex Apffel
- Anya Tsalenko
- Rick Fasani
- Aaron Gee

LBNL

- Chris Petzold
- Yan Chen
- Edward Baidoo
- Jennifer Gin

PNNL

- Kristin Burnum-Johnson
- Jon Magnuson
- Young-Mo Kim
- Nathalie Munoz
- Aivett Bilbao
- Yuqian Gao
- Karl Weitz
- Daniel Orton

SNL

- Kunal Poorey
- Corey Hudson

Additional Slides

(Not a template slide – for information purposes only)

- *The following slides are to be included in your submission for evaluation purposes, but will **not** be part of your oral presentation –*
- *You may refer to them during the Q&A period if they are helpful to you in explaining certain points.*

Publications, Patents, Presentations, Awards, and Commercialization

- List any publications, patents, awards, and presentations that have resulted from work on this project
- Use at least 12 point font
- Describe the status of any technology transfer or commercialization efforts

Publications, presentations, and IP

- Bilbao et al. Poster presented at the 68th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics (virtual June 1 - 12, 2020)
- Bilbao et al. Lightning talk presented at the Skyline User Group Meeting (virtual May 27 - 28, 2020)
- Manuscript for UHPLC-ion mobility MS workflow (in preparation; PNNL, Agilent, LBNL, SNL)

Note: This slide is for the use of the Peer Reviewers only – it is not to be presented as part of your oral presentation. These Additional Slides will be included in the copy of your presentation that will be made available to the Reviewers.