BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

# DOE Bioenergy Technologies Office (BETO) 2023 Project Peer Review

# Optimizing Bio-jet Fuel Blends with the Feedstock to Function tool

April 5, 2023

Data, Modeling, and Analysis Program (DMA)

Vi Rapp

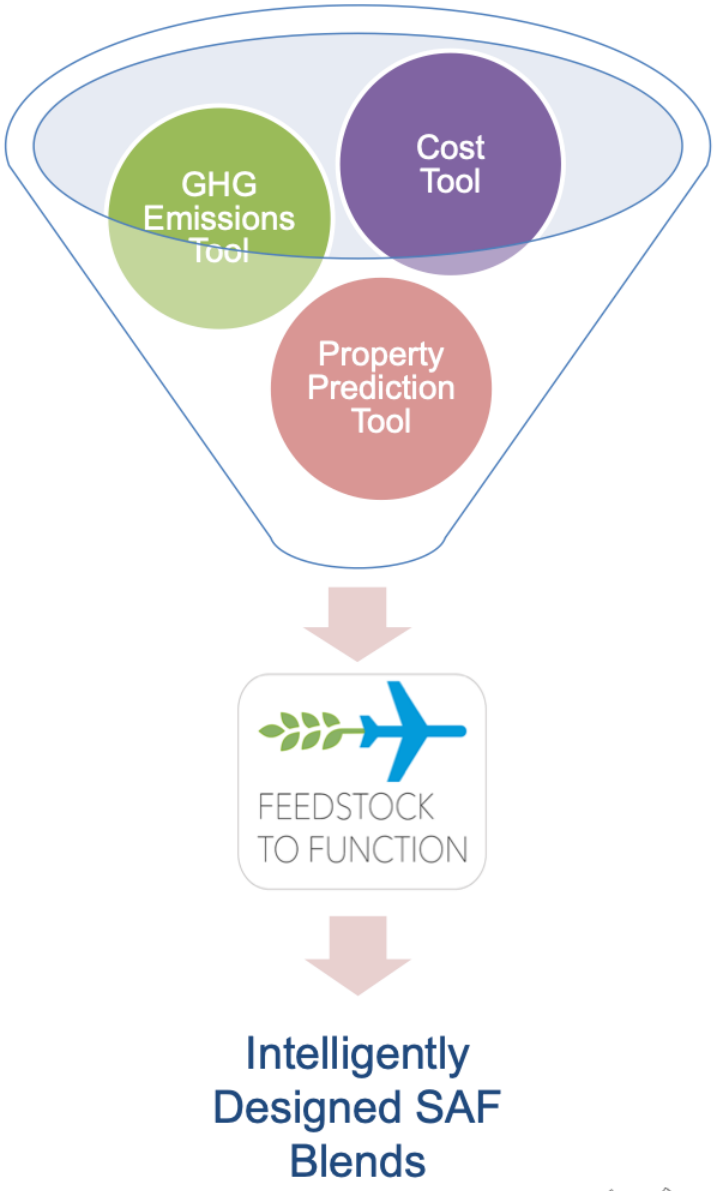
Lawrence Berkeley National Laboratory

# Project Overview

- Development of sustainable aviation fuels is limited by significant technical, social, and regulatory barriers
- Feedstock to Function is the first comprehensive webtool that predicts promising molecule properties and evaluates the costs, benefits, and risks for faster, less expensive bioprocess optimization, certification, and scale-up
- This project, if successful, will accelerate sustainable aviation fuel innovation and early R&D by enabling users to rapidly screen and identify fuel blends

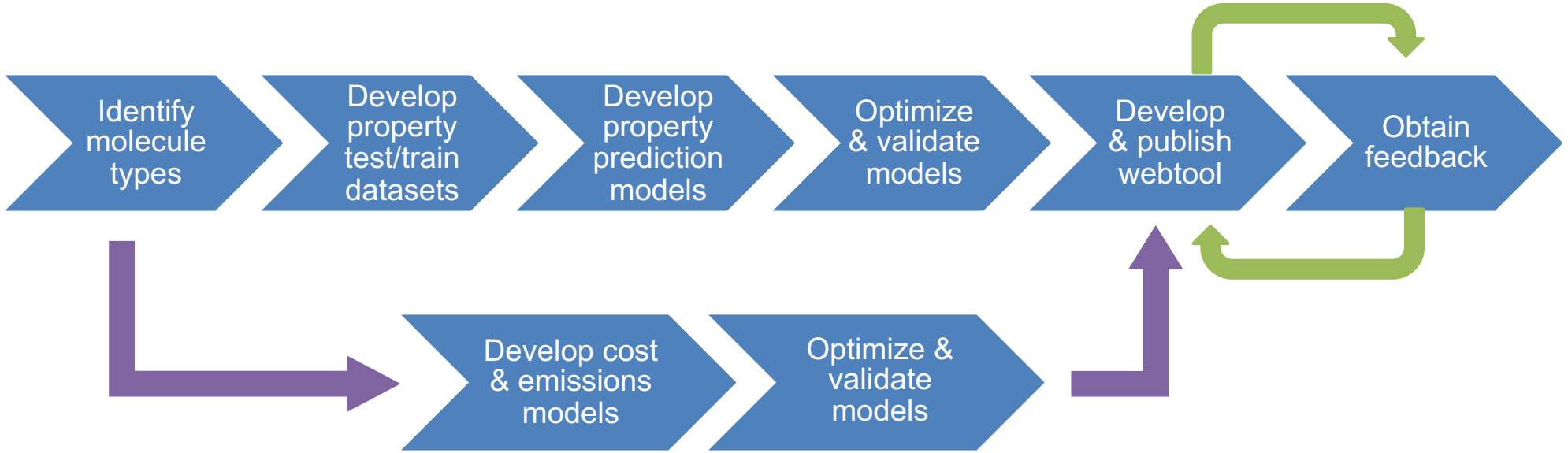
**Project Goals:**

- Expand webtool to include property prediction of fuel blends
- Enable users to design and explore new SAF blends by matching fuel blends to desired properties
- Link blends to lightweight minimum selling price and greenhouse gas emissions estimation tool



# 1. APPROACH

# 1 – Approach: Creating the Feedstock to Function Tool




## Technical Approach (2019-2021)

- Collect and clean experimental property data from published sources
- Predict fundamental properties of pure (neat) molecules for aviation fuels using machine learning
- Optimize and validate accuracy of models by comparing to published experimental data
- Link to light-weight technoeconomic analysis and life-cycle assessment tool
- Publish free webtool that allows users to explore viable molecules and associated costs and emissions with production and scale-up

# 1 – Approach: Current Feedstock to Function Tool

**An opensource webtool that rapidly screens bio-derived molecules for replacing or substituting petrochemical intermediates, fuels, and chemicals**

## Input:

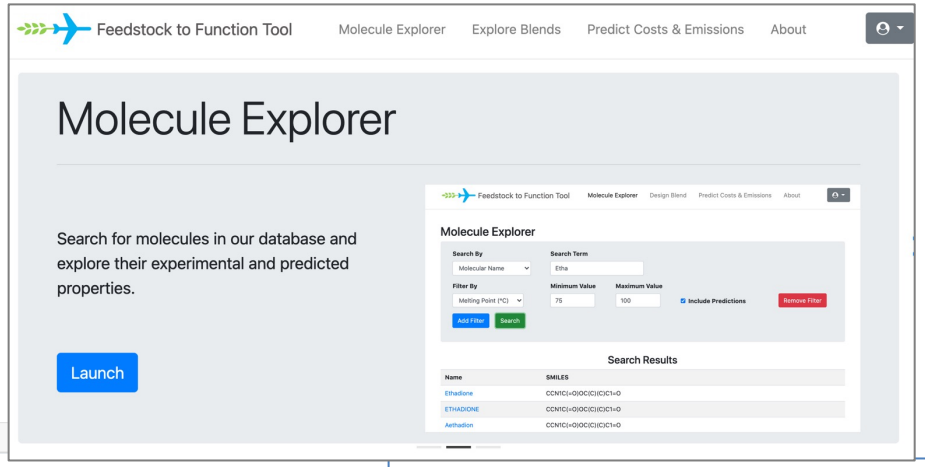
- User-selected molecule or target property ranges

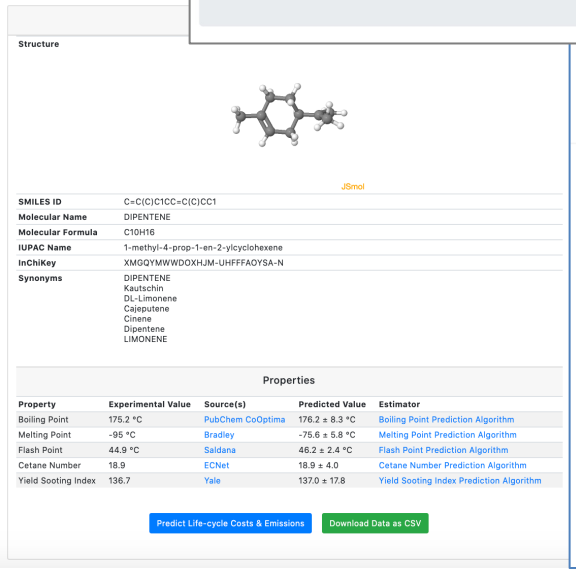
## Output:

- Chemical properties (flash point, melting point, boiling point, heat of combustion, yield sooting index)
- Costs and emissions of viable aviation fuel molecules

## Model Success & Connections:

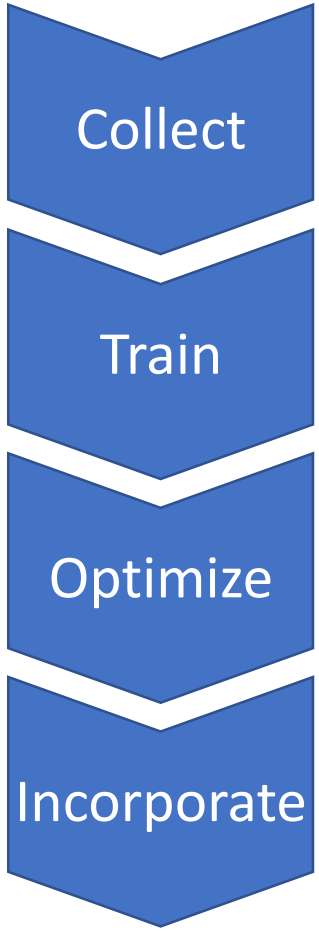
- Predictive models are validated, peer reviewed, and publicly available
- Bio-Cradle-to-Grave (BIOC2G) for cost, water, and emissions analysis




# 1 – Approach: Expanding the Feedstock to Function Tool

**Incorporate fuel-blend property predictions and user-designed blend features to support optimization and deployment of sustainable aviation fuels**




Collect and clean experimental fuel blend data for model training/testing

Train machine learning models to predict fundamental properties of fuel blends and identify blends that match target properties

Optimize and validate accuracy of models by comparing to experimental data

Incorporate data and models into webtool and obtain user feedback

# 1 – Approach: Project Management & Risks

## Project Management:

- Regular meetings
  - Team: 1/month
  - BETO: 1/month
  - Stakeholders: 2/year
- Reporting
  - Quarterly progress reports
  - Annual diversity, equity, and inclusion report
- Risk management
  - Short term: regular meetings/updates
  - Long term: annual operating plan
- Stakeholder engagement
  - Model development and feedback

## Berkeley Lab

**PI**
Vi Rapp

**Experimental Research**
Sharon Chen

**Applied Math**
Ana Comesana

**Software Development**
Tyler Huntington

## Oregon State University

**OSU PI**
Kyle Niemeyer

**Mechanical Engineer**
Ali Martz

# 1 – Approach: Technical Risks and Mitigation Strategies

| Technical Risk/Challenge | Mitigation Strategy |
|---|---|
| Data inaccuracy, inconsistency, or lack of data availability | Compare multiple published data sources when available; directly measure FTIR spectra and properties |
| Machine learning algorithm unable to predict properties of blends to within 30% of experimental values | Investigate data decomposition methods, feature selection methods, and consider a wide breadth of algorithms to achieve at least 30% accuracy |
| Optimization method unable to identify fuel-blends that match target property values | Explore diverse optimization algorithms, including stochastic global methods |

# 1 – Approach: Metrics, Communications, & Collaborations

## Metrics of Success

- Demonstrate that the fuel-blend property prediction model predicts properties to within 15% of published experimental values
- **Go/No-Go in FY23:** Demonstrate that the fuel-blend design feature is capable of accurately identifying at least one blend that is within 30% of at least two target property values
- Integrate blend features into webtool; obtain user feedback and incorporate suggestions

## Diversity, equity, and inclusion

- Attend trainings, participate in institutional programs, recruit from underrepresented, and provide opportunities for all staff to lead activities that fosters career growth

## Communication & Collaboration

- The Center of Excellence for Alternative Jet Fuels and Environment (ASCENT)
- Existing BETO Biojet efforts: Sandia, NREL, PNNL, Dayton University, Georgia Tech
- Chemical Industry
- Naval Air Warfare Center Weapons Division
- Air Force Research Laboratory

MOLECULAR FOUNDRY

RESEARCH MADE POSSIBLE BY THE MOLECULAR FOUNDRY
foundry.lbl.gov

This project uses high performance computing resources located at Oak Ridge National Laboratory and provided by the Bioenergy Technologies Office.

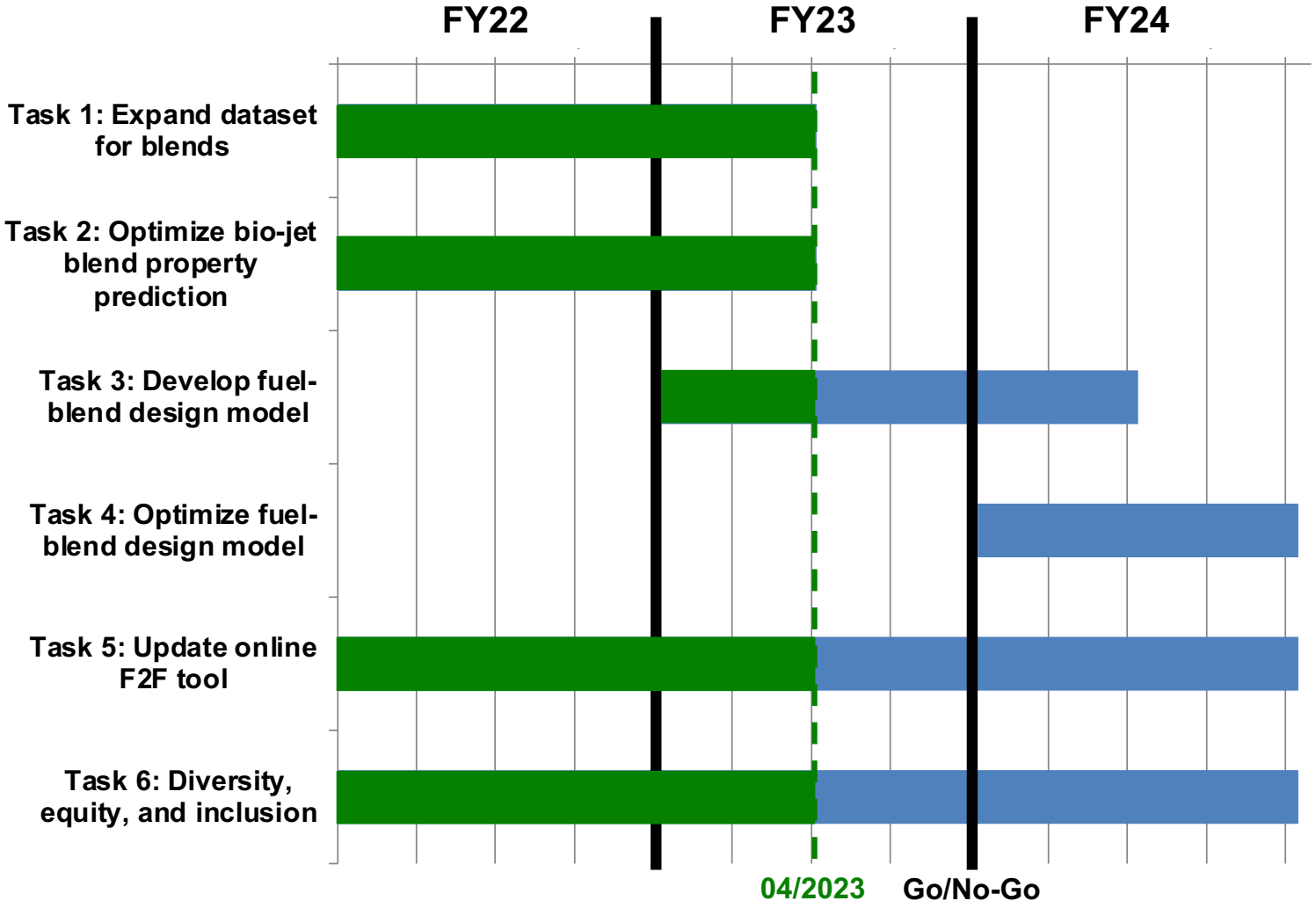


BERKELEY LAB

# 2 – PROGRESS AND OUTCOMES

# 2 – Progress and Outcomes: Schedule

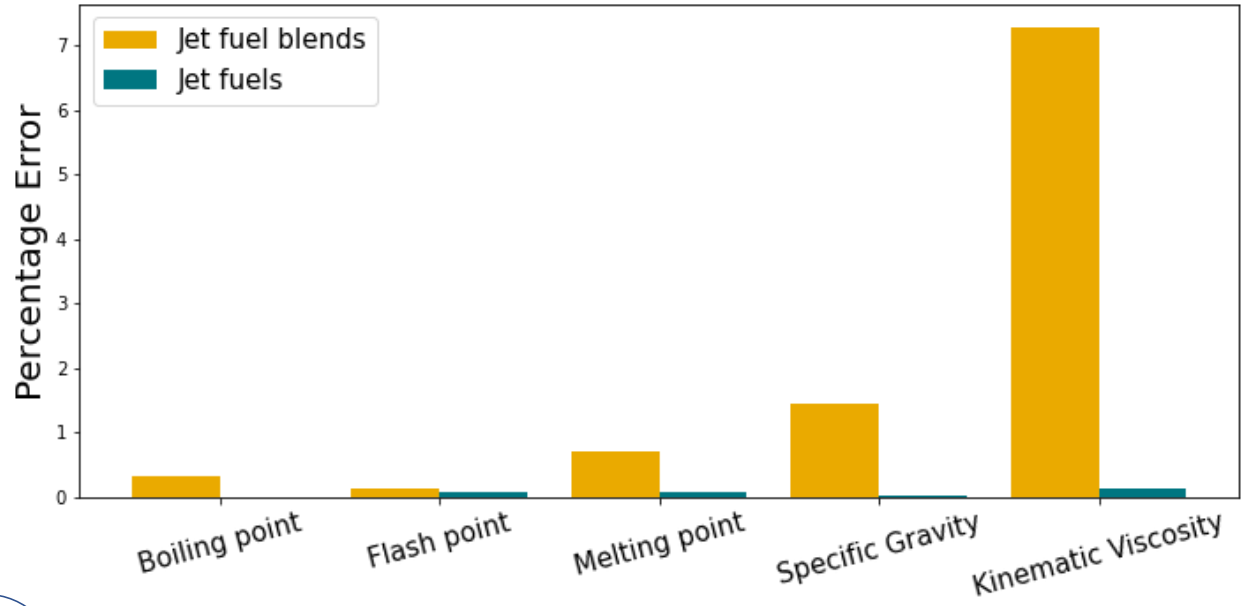
**Project is on schedule, achieved FY2022 milestones, and on target to complete future milestones**

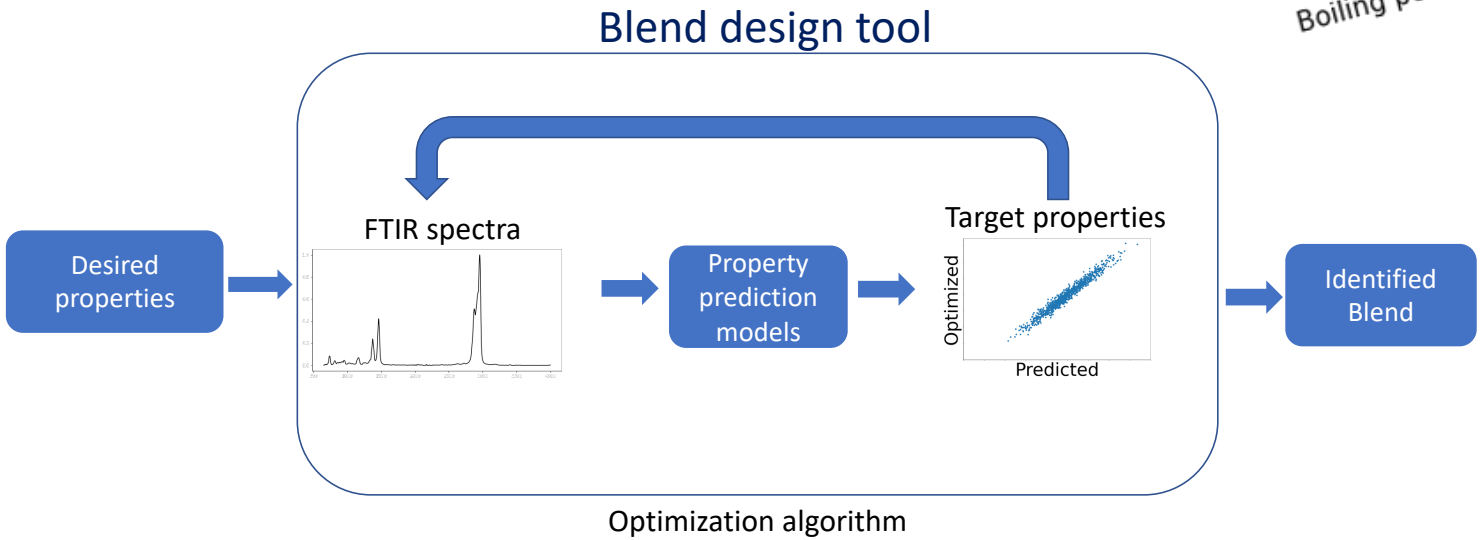



- ✓ FY2022: Demonstrate fuel-blend property prediction model predicts properties of fuel-blends at different blend ratios to within 15% of published experimental values
- FY2023 (Go/No-Go): Demonstrate preliminary fuel-blend design feature accurately identifies a two-component blend two within 30% of target property values
- FY2024: Complete optimization of fuel-blend design feature; demonstrate it identifies blends (up to three components) to within 15% of target property values; complete integration of features into webtool

# 2 – Progress and Outcomes: Technical Accomplishments

Fuel-blend property prediction model achieved < 7%, surpassing 15% error goal!




Established optimization method for blend design feature

# 2 – Progress and Outcomes: Technical Accomplishments

## Released preliminary fuel blend explorer tool

Feedstock to Function Tool  Molecule Explorer  Explore Blends  Predict Costs & Emissions  About

### Blend Explorer

Search for desired blends using the options below. Use "Add Filter" to search for multiple blends within property ranges. If available, experimental values are listed. Click on a desired blend to see predicted property values if experimental values are not available.

**Search**: Jet-A Blends

**Search Term**: HEFA

**Filter By**: Kinematic Viscosity

**Minimum Value**: 4

**Maximum Value**: 5

Remove Filter

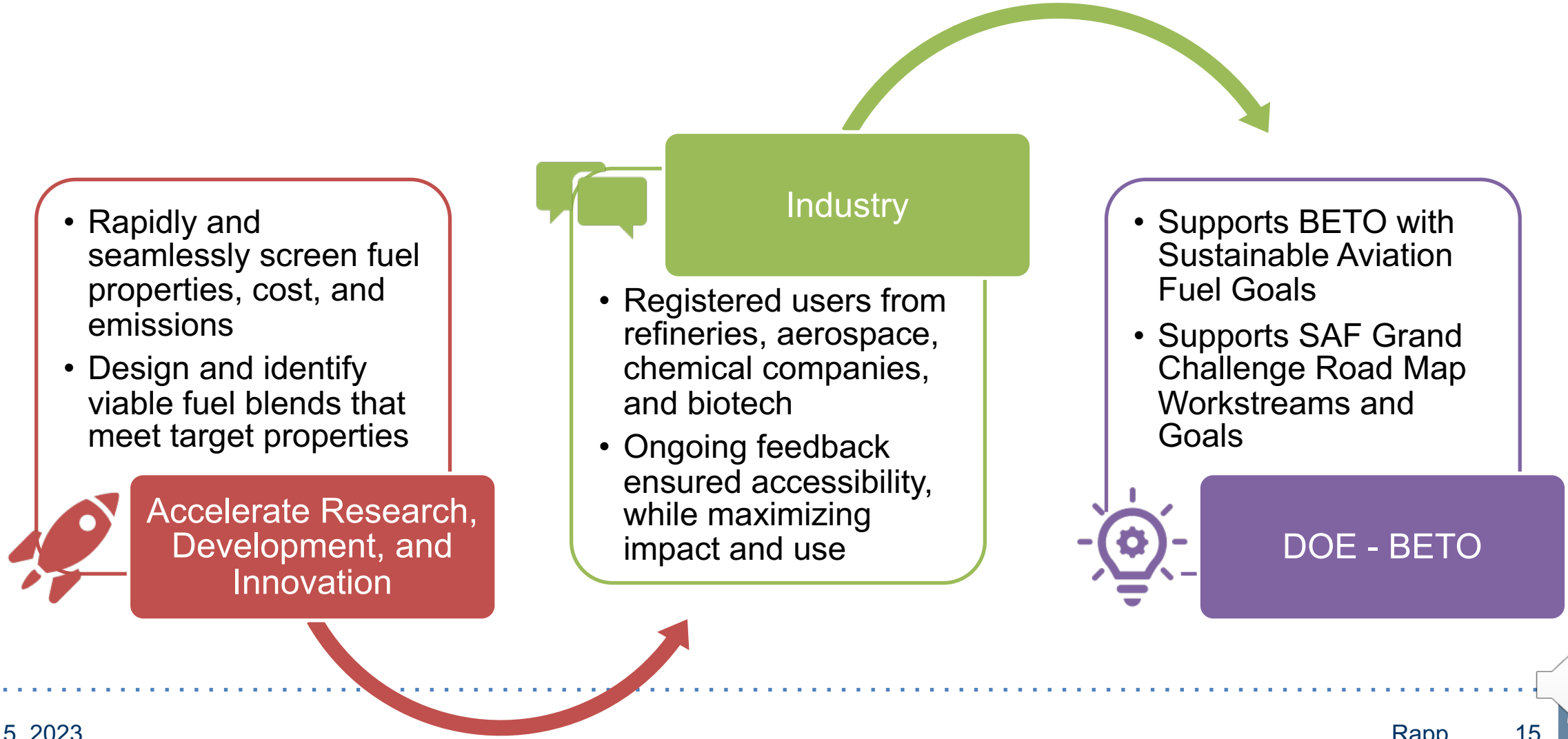
Add Filter  Search

| Blend | Boiling Point (K) | Flash Point (K) | Melting Point (K) | Density (RDW) | Density (kg/m^3) | Kinematic Viscosity (mm^2/s) |
|---|---|---|---|---|---|---|
| 50% HEFA and 50% Jet-A | 550.5 | 315.7 | 224.2 | 0.8 | 786.1 | 4.9 |
| 30% HEFA and 70% Jet-A | 553.6 | 316.7 | 225.2 | 0.8 | 796 | 4.7 |
| 20% HEFA and 80% Jet-A | 555.1 | 317.2 | 229.2 | 0.8 | 801 | 4.6 |
| 10% HEFA and 90% Jet-A | 556.5 | 317.7 | 227.2 | 0.8 | 805.9 | 4.5 |

# 3 – IMPACT

# 3 – Impact

**Feedstock to Function can help accelerate SAF R&D efforts, reduce expenditures, and enable more productive and successful experimentation**



# 3 – Impact: Supporting SAF Grand Challenge and BETO Goals

## Supports Conversion Technology Innovation and End Use Goals

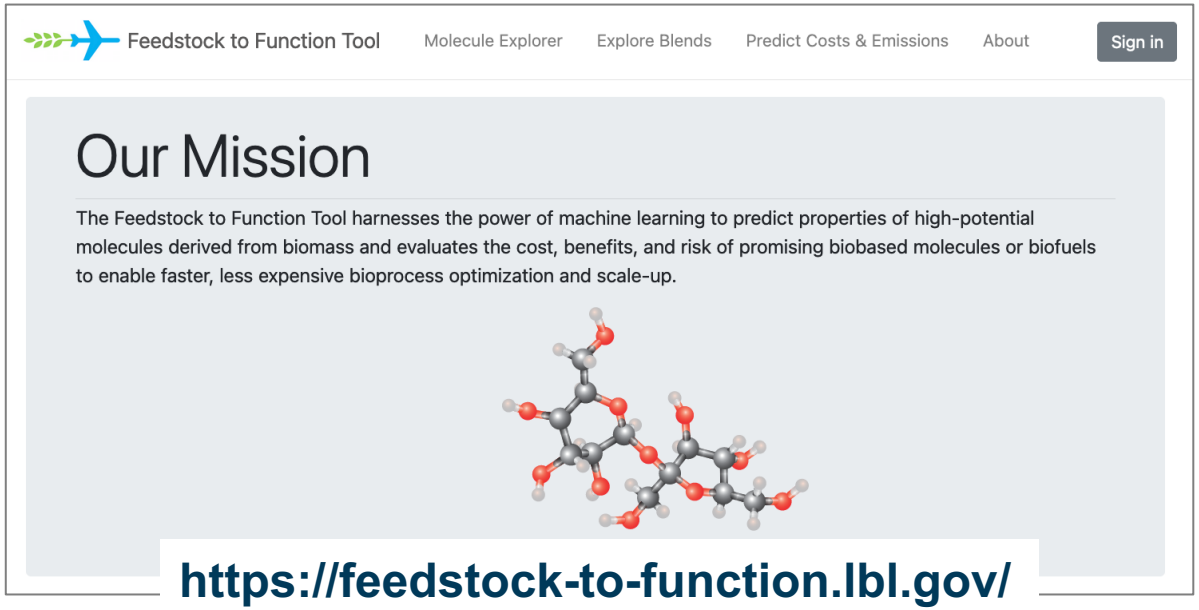
- EU.1. Support SAF Evaluation, Testing, Qualification, and Specification
- EU.2. Enable Use of Drop-In Unblended SAF and SAF Blends up to 100%
- EU.3. Investigate Synthetic Aviation Turbine Fuels Offering Performance or Producibility Advantages
- CT.3.Develop Biointermediates and Pathways for Compatibility With Existing Capital Assets
- CT.4. Reduce Risk During Scale-Up and Operations


## Links to BIOC2G that supports additional goals

- De-risk scale-up through R&D and integrated piloting of critical pathways by 2030 to accelerate fuel conversion technology scale-up (Workstreams CT.1–CT.4)
- Conduct RD&D on scaling and sustainability of biomass, waste, and residue feedstocks to enable innovations in technologies and strategies that increase the availability of biomass and waste resources at reduced carbon intensity and cost. (Workstreams FI.3 and FI.6)

# 3 – Impact: Disseminating Information

- Registered national and international users from industry, universities, national labs, research institutions, and government
- Published open-access webtool
- Presentations at high impact conferences
- Publications in journals
- Engage with industry and experts to promote discussions of outcomes and challenges





Fuel 321 (2022) 123836

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full length article

A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties

Ana E. Comesana [a], Tyler T. Huntington [b,c], Corinne D. Scown [a,b,c,d], Kyle E. Niemeyer [e], Vi H. Rapp [a,*]

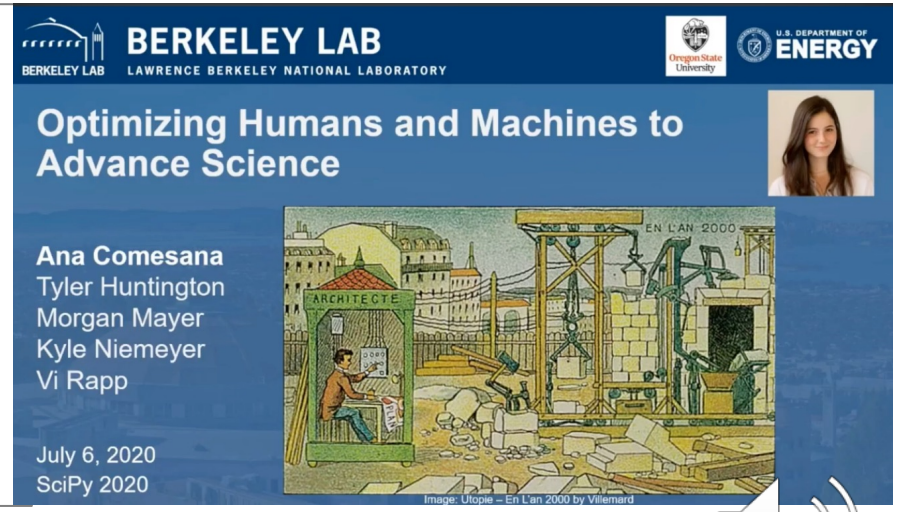
[a] Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, United States
[b] Biosciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, United States
[c] Joint BioEnergy Institute, Emeryville, CA, United States
[d] Energy & Biosciences Institute, University of California, Berkeley, CA, United States
[e] School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, OR, United States

ARTICLE INFO

*Keywords:*
Chemical descriptors
Machine learning
Biofuel
TPOT

ABSTRACT

Machine learning has proven to be a powerful tool for accelerating biofuel development. Although numerous models are available to predict a range of properties using chemical descriptors, there is a trade-off between interpretability and performance. Neural networks provide predictive models with high accuracy at the expense of some interpretability, while simpler models such as linear regression often lack in accuracy. In addition to model architecture, feature selection is also critical for developing interpretable and accurate predictive models. We present a method for systematically selecting molecular descriptor features and developing interpretable machine learning models without sacrificing accuracy. Our method simplifies the process of selecting features by reducing feature multicollinearity and enables discoveries of new relationships between global properties and molecular descriptors. To demonstrate our approach, we developed models for predicting melting point, boiling point, flash point, yield sooting index, and net heat of combustion with the help of the Tree-based Pipeline Optimization Tool (TPOT). For training, we used publicly available experimental data for up to 8351 molecules. Our models accurately predict various molecular properties for organic molecules (mean absolute percent error (MAPE) ranges from 3.3% to 10.5%) and provide a set of features that are well-correlated to the property. This method enables researchers to explore sets of features that significantly contribute to the prediction of the property, offering new scientific insights. To help accelerate early stage biofuel research and development, we also integrated the data and models into a open-source, interactive web tool.



# Summary

| | |
|---|---|
| Goal | Building on the success of the Feedstock to Function tool, this project will expand the fuel-blending property prediction feature to support optimization and deployment of sustainable aviation fuels. |
| Approach | • Expand existing Feedstock to Function Tool to predict properties of fuel-blends<br>• Develop and optimize a fuel-blend design feature that enables rapid identification of bio-based blends that meet target property ranges |
| Progress & Outcomes | • Fuel-blend property prediction model surpassed 15% target error (achieved < 7%)<br>• Incorporating fuel-blend prediction model into webtool<br>• Established optimization method for blend design feature |
| Impact | • Engage industry and academic experts for input and feedback to ensure accessibility<br>• Support DOE – BETO goals and related projects<br>• Disseminate technical results through webtool, presentations and publications |
| Future Work | • Complete optimization of fuel-blend design feature and demonstrate blends identified match within 15% of target properties<br>• Complete integration of features into webtool and obtain feedback<br>• Complete Diversity, equity, and inclusion plan |

# Quad Chart Overview

## Timeline

- Project start date: 10/01/2021
- Project end date: 09/30/2024

| | FY22 Costed | Total Award |
|---|---|---|
| **DOE Funding** | (10/01/2021 – 9/30/2022) | $325,000 |
| **Project Cost Share** | N/A | N/A |

## TRL at Project Start: 2
## TRL at Project End: 3

## Project Goal

Accelerate innovation and early R&D of sustainable aviation fuels by expanding the Feedstock to Function tool to include a fuel-blend property prediction feature and a fuel-blend design feature that enables users to identify bio-based blends that meet target property ranges.

## End of Project Milestone

Complete development and optimization of the fuel-blend design feature and demonstrate it can identify available experimental blend data to within 15% of target property values. Identified blends will contain up to three components and be validated against blends that are well characterized. Complete development and integration of the fuel-blend design feature interface onto the webtool. Obtain and incorporate feedback from potential users (i.e., scientists and companies), and publish results in a peer-reviewed journal. Provide a summary of IDEA success metrics to DOE in the quarterly report for each IDEA activity listed in the plan.

## Funding Mechanism

Lab Call, 2021

## Project Partners

- Oregon State University

# ADDITIONAL SLIDES

# Responses to 2021 Reviewers' Comments

Comment 1: "Nice project with a great summary. I am not sure about the combination of physical properties, costs and LCA, especially process costs. It seems like the project should focus on properties only. LCA can be included as they are primarily based on heat and material balances. The estimation of feedstock handling and hydrogenation costs are a bridge too far for me, especially since they are based on process simulations. We have no biorefinery industry and as a long-time process modeler and experienced engineer, I don't think that we can predict the costs and should not try. We do not know the costs; the best that a model can do is to show relative costs and users of the model will not know this. I am excited that the project will add properties of blends and think that will be a huge benefit and impact. One of the issues is that they should do a better job of highlighting potential error and I think they should cut off property projections when the error is significant. Many times researchers are desperate for data and will use whatever they can find and ignore caveats. The project assess the level of error that is acceptable and only publish those values. If this was based on products only, I would probably give it all 5's. I think we dilute its effectiveness by including everything."

Response 1: We greatly appreciate the feedback and agree that continuation of this project should focus only on properties. Thank you for the excellent comment to highlight prediction error and appropriate cutoff thresholds. We will explore integrating functionality to let the user set prediction error threshold. We will also leverage the methods described above to determine thresholds for removing property predictions with significant error. Regarding our data, as stated in the slides, all property data used for training models is experimental.

# Presentations and Webtool Release

**Webtool Release**

Feedstock to function website https://feedstock-to-function.lbl.gov/

**Publications**

- Comesana AE, Huntington TT, Scown CD, Niemeyer KE, Rapp VH. (2022) A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties. Fuel, 321:123836. DOI: 10.1016/j.fuel.2022.123836.

**Presentations**

- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Predicting Bio-jet Properties Using a Tree-Based Pipeline Optimization Tool” paper accepted for presentation at Spring Western States Section of the Combustion Institute in Stanford, CA March 2022
- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Optimizing Humans and Machines to Advance Science” presented virtually at SciPy Conference in July 2020. (https://youtu.be/ENOf0IZDla8)
- Vi Rapp, “Machine learning for fuels, chemicals and food ingredients,” UC Congressional Briefing: The promise of Artificial Intelligence Research in Washington D.C. on Dec. 11, 2019.
- Corinne Scown, "Using TEA and LCA to Answer Your Burning (Cost and Carbon-Related) Questions," at the Joint BioEnergy Institute in Berkeley, CA on Oct 23, 2019.
- Corinne Scown, “"Sustainability at the Joint BioEnergy Institute," AIChE Bioenergy Sustainability Conference in Nashville, TN on Oct. 21, 2019.
- Ana Comesana, Tyler Huntington, Morgan Mayer, Kyle Niemeyer, Vi Rapp “Innovative Biofuel Development through Machine Learning” poster presented at Bay Area Scientific Computing Day in Berkeley, CA on Dec. 16, 2019.
- Morgan Mayer, Tyler Huntington, Ana Comesana, Vi Rapp, Kyle Niemeyer, “Can machine learning predict fuel properties accurately?” Fall 2019 Western States Section of the Combustion Institute Meeting in Albuquerque, NM on Oct. 14, 2019.