BETO 2023 Project Peer Review Biochemical Process Modeling and Simulation (BPMS)

April 6th, 2023
Biochemical Conversion & Lignin Utilization Session
Yannick Bomble
NREL

Project Overview

Overarching goal: accelerate experimental and engineering efforts in BETO to reduce the cost of SAF, biochemicals, and bio-derived materials by **providing insights and solutions to process bottlenecks**. More specifically some of our goals are to enable:

- Improvement of metabolic enzymes and metabolic pathways to reduce the cost of SAF production.
- Engineering of plastic degrading enzymes and assessment of depolymerization schemes to guide plastics upcycling strategies.
- Determination of best fermentation conditions and provide reliable models for TEA
- Improvement of enzyme stability and specificity for cell free applications to help de-risk this technology.

These goals will directly impact BETO's ability to achieve its 2030 goal of 3B gallons SAF and 2050 goals of 60B gallons renewable hydrocarbon fuels and 40B pounds of renewable chemicals.

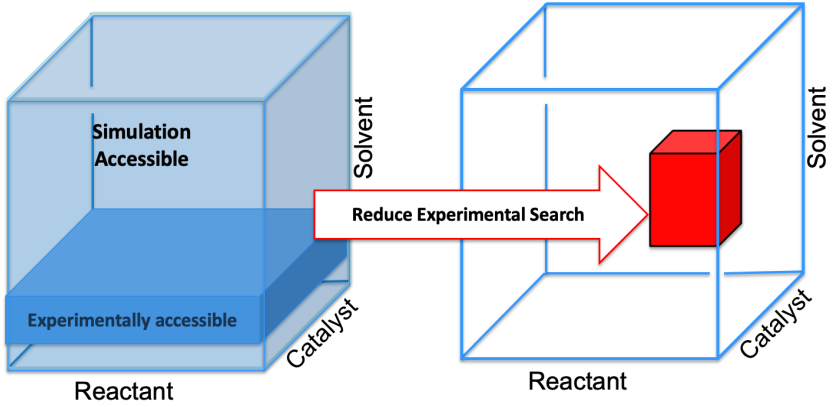
Context:

- This project has been ongoing for more than 2 funding cycles. Current funding cycle started in FY22. Current budget: $1M/year.
- This project has enabled many breakthroughs in the BETO biochemical plaform notably in lignin utilization, plastics depolymerization, production of 2,3 BDO, and scale up of these fermentations which will lead to lower the costs of conversion.

Why should we care? Relevance of modeling

1. Lower time to solution by reducing experimental work and time

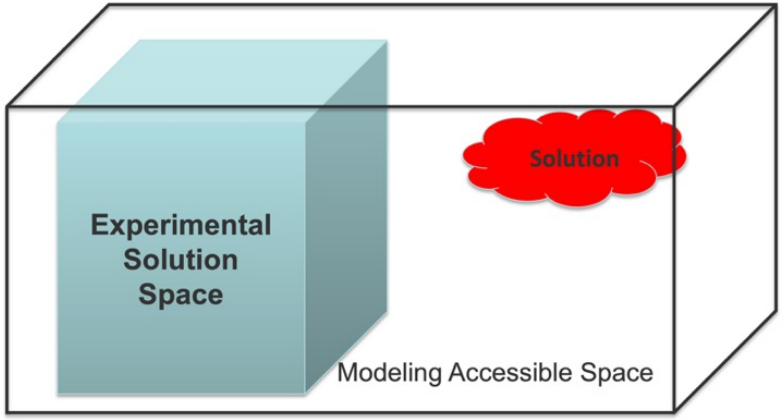
Solution space is too big for experiment but accessible by modeling.



Example: Determine the right aliphatic compound to produce based on ease of extraction before experimental efforts are implemented.

2. Solutions inaccessible experimentally

Modeling can find solutions unavailable to standard experimental search.



Examples: Risk too high

- Mutations/knockouts believed to be fatal to microbes.
- Testing reactor designs at industrial scale.
- Exploring mutants with very high number of mutations.

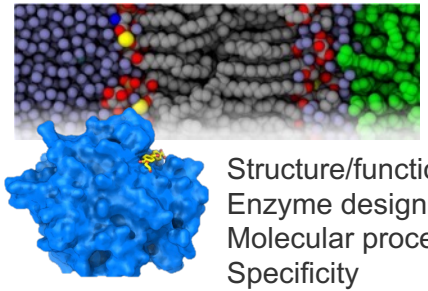
Relevance: Accelerates research and provides complementary insights leading to increased efficiency, reduced costs, and reduced risks for SAF and biochemical production in BETO.

1. Approach: Management

Project: Biochemical Process Modeling and Simulation – Yannick Bomble

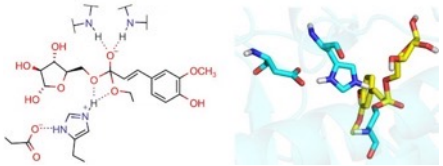
1- Molecular Modeling - Brandon Knott

Molecular dynamics
Quantum mechanics



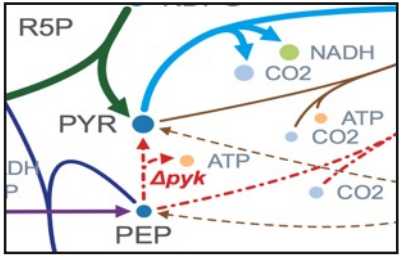
Structure/function
Enzyme design
Molecular processes
Specificity

QM and QM/MM approaches to upgrading chemistry and catalysis



2 - Metabolic Modeling and Machine Learning - Yannick Bomble

- Metabolic models
- Machine learning
- DBTL Learn efforts and omics analyses
- Redox enzyme / cofactor engineering

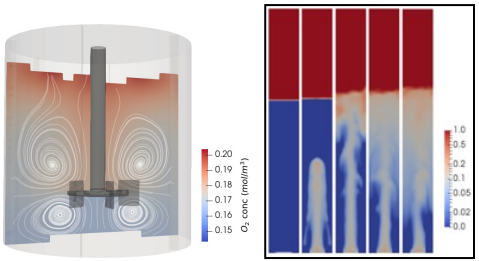


New subcontract with Sophie Barbe at INRA (France) “ML guided multistate modeling for protein engineering” ($60K)

3 - Mechanistic Process Modeling - Hari Sitaraman

Coupled CFD/Rxn-diffusion
Multi-scale modeling

New subcontract with Joseph Samaniuk at CSM “Modeling Transport and Reaction Kinetics of Reactor Systems Utilizing Cell-Free Biocatalysis.” ($70K)

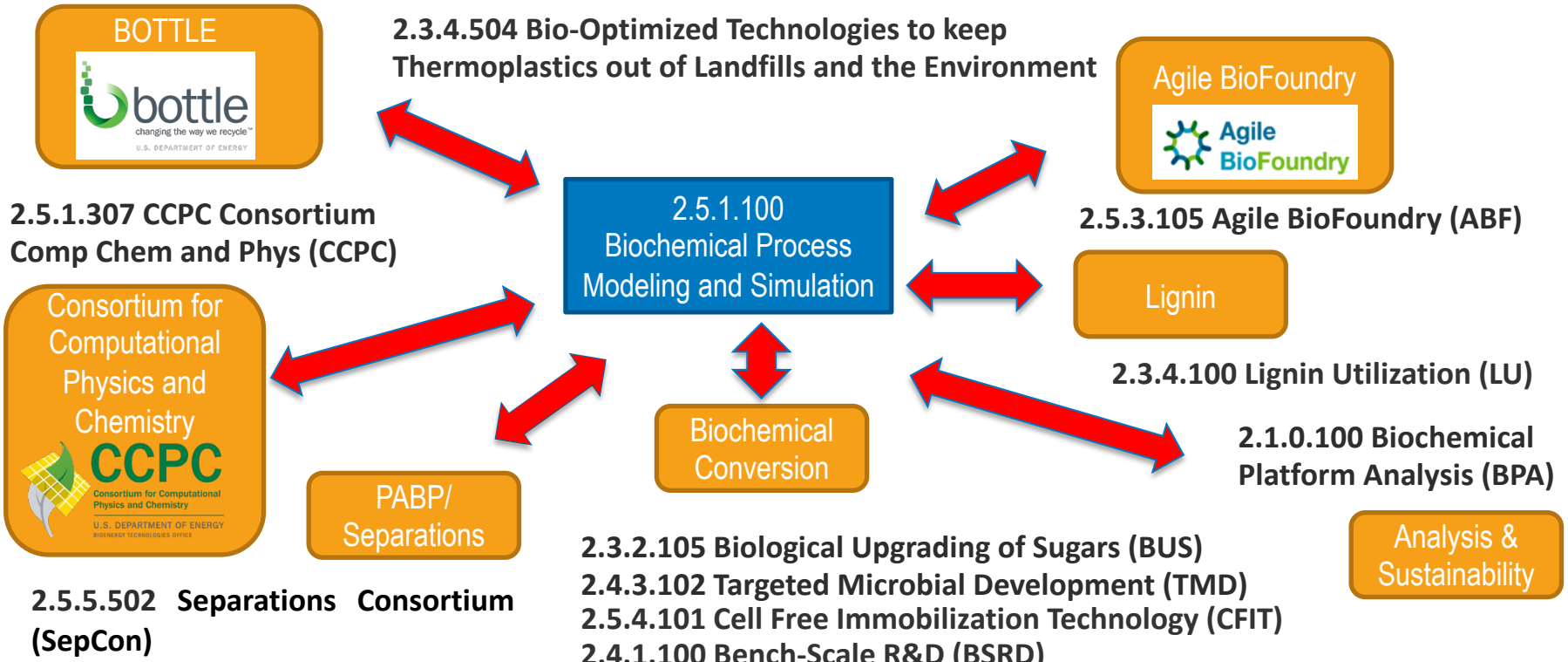


Project split into tasks by modeling type, managed by person with appropriate expertise
Task Managers responsible for:

- **Relevance**
- AOP, Milestones, quarterly reporting according to the guidance of BETO
- **Communication with other projects**

1. Approach: Management

The **most important** aspect in managing this project to **mitigate risks is the identification and prioritization of modeling activities** that are most meaningful and impactful to experimental projects.



1. Approach: Setting priorities and goals

Goals:

- Improve metabolic enzymes and metabolic pathways to reduce the production cost of SAF.
- Engineer plastic degrading enzymes and assess depolymerization schemes to guide plastics upcycling strategies.
- Determine best fermentation conditions and provide reliable models for TEA.
- Improve enzyme stability and specificity for cell free applications to help de-risk this technology.
- Build long-term partnerships with minority-serving institutions (MSI) to deliver collaborative research opportunities

Approach:

- Use MultiScale Approach: Molecular (Task 1), Metabolic/Cellular (Task 2), and Macroscopic (Task 3) simulation. Leverage EERE computer resource: **Kestrel (NREL).**
- **Assess state of the art computational techniques and develop new and improved ones if needed.**
- Target most **relevant bottlenecks and barriers** in **most BETO-relevant processes.**
- Apply to mentor interns for the STAR program (Student Training in Applied Research). Actively recruit interns from the GEM program. Actively distribute presentations on the impact of modeling in SAF research to Minority Serving Institutions to motivate students to pursue scientific studies with an emphasis in modeling.

1. Approach: Identifying risks and mitigation strategies

Risks:

- Software and methods need to be developed to meet the questions and necessary speed for timely answers (MD, CFD, QM/MM, FE, analysis).
- Local computer hardware needs to stay at state-of-the-art.
- Project and time management given the number of projects.
- Turnover over the last 2 years as researchers in this field are in high demand.

Mitigation strategies :

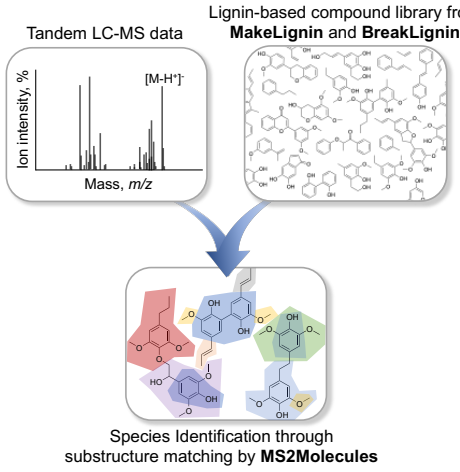
- Leverage **CCPC** (Consortium for Computational Physics and Chemistry) collaborations using all theory and modeling expertise **across laboratories**
- **Strong and regular communication and joint metrified milestones with other experimental projects.**
- **Placed 2 new subcontracts to ensure continuity in our research (CSM and INRAE)**. We are also recruiting new graduate students.

2. Progress and Outcomes: Task 1: Enabling better identification of lignin derived species

RELEVANCE: Characterizing lignin-derived compounds is critical for subsequent upcycling to lower the cost of conversion.

OUTCOME: LigninWrangler, a publicly available user-friendly computational tool that allows the identification of closely related lignin compounds. This tool can drastically improve lignin analytics and therefore lignin valorization efforts.

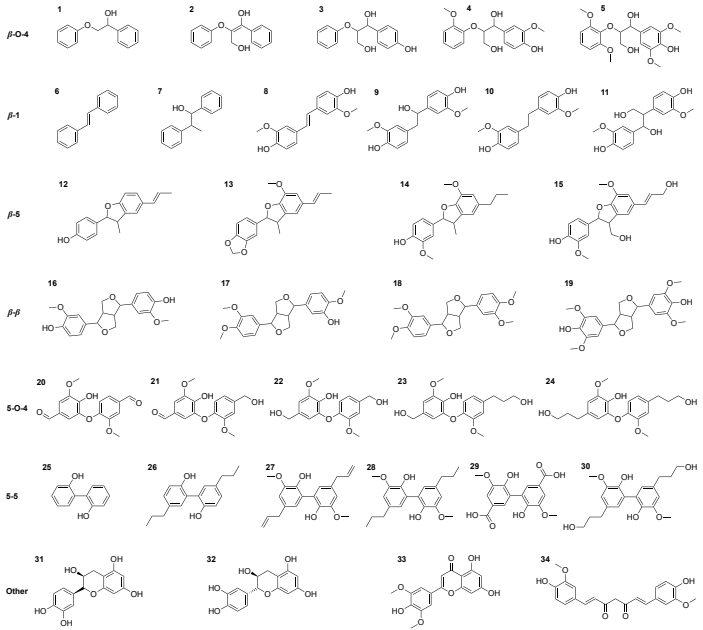
- Lignin heterogeneity is a significant challenge for separation and conversion efforts.
- Enhanced analytical techniques are very promising to identify lignin-derived compounds but the interpretation of the data is still difficult.
- To interpret high accuracy mass spec data, our computational tool (LigninWrangler) can dramatically help and assign molecular structures.



Create libraries of lignin-based species by 1) generating an ensemble of lignin molecules and 2) applying relevant chemical reaction rules to generate potential lignin decomposition product species.

Coupled with relevant mass spec data, MS2Molecules can **ID likely parents** by matching observed to expected m/z values for fragments/substructures.

34 model compounds studied
Encompassing 7 lignin interunit bond types



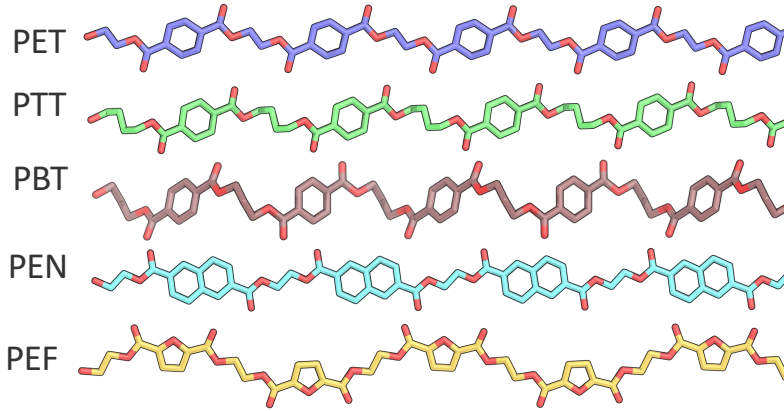
Dong, Mayes *et al. ChemSusChem* **2023**
Mayes *et al. in preparation*

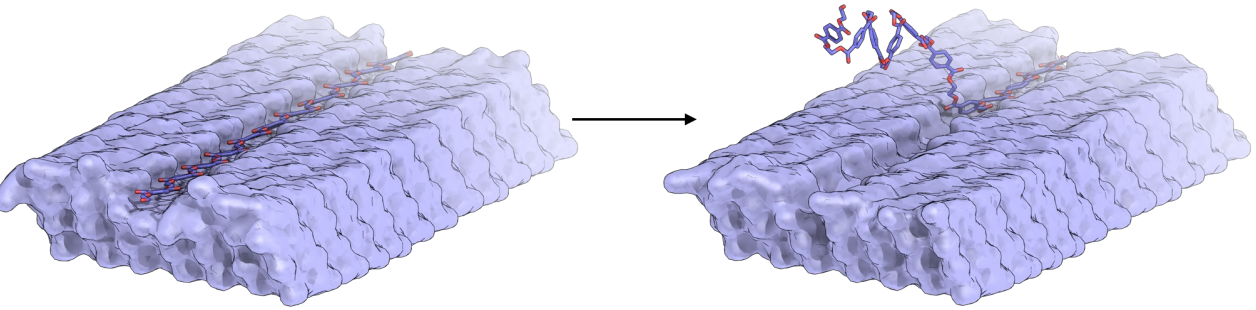
2. Progress and Outcomes: Task 1 - Improving plastics decrystallization

RELEVANCE: Many commodity plastics are semi-crystalline and decrystallization is the first step in many depolymerization schemes, *e.g.* via depolymerase enzymes.

OUTCOME: Strength of the basic structural interactions in different solvents. Provides guidance for plastics upcycling strategies in BETO.

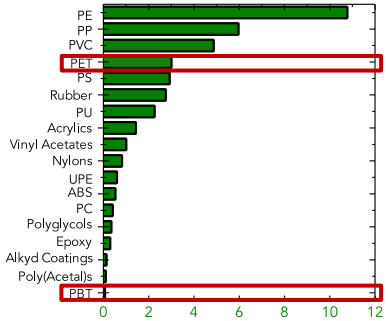
- Molecular dynamics based procedure to compute the free energy to decrystallize a single chain.
- Commercial soft drink bottles are about 16% crystalline, but PET, polyolefins, and nylons can be > 50% crystalline.
- Next steps will target non-aqueous solvents, and other polymer classes (polyolefins, nylons).





Polymer chain in crystal surface

Polymer chain extracted from surface



US Consumption, MMT/year

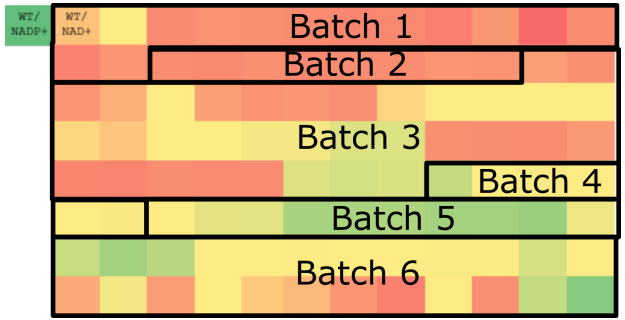
Nicholson *et al. Joule* **2021**

2. Progress and Outcomes: Task 2 - Using Machine Learning to improve key enzymes in BETO

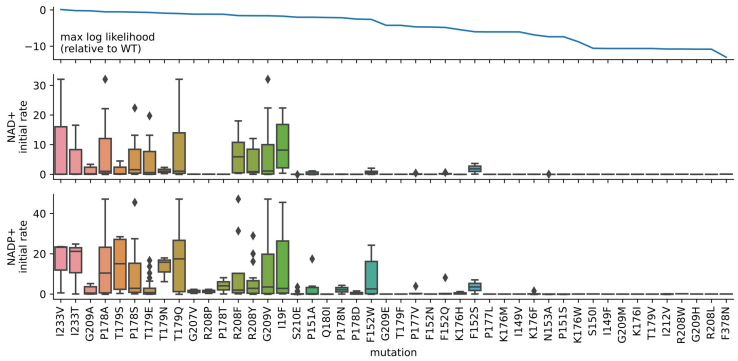
RELEVANCE: There is a need for new methodologies to improve key enzymes for applications *in vivo* or in cell-free systems.

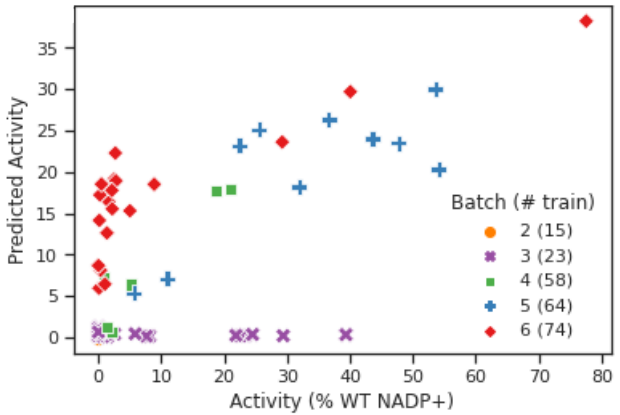
OUTCOME: Demonstrated a new machine learning (ML) approach for engineering enzymes using sparse experimental datasets.

- Proof of concept: GapN is essential for cell-free biocatalysis, but $NADP^+$ cofactor is labile and restricts the choice of pathways.
- Achieved >80% WT activity by hand-crafting mutants over 6 rounds of modeling/experimental cycles.
- Evolutionary-scale language models can guide enzyme engineering toward functional mutations even with sparse experimental data.



Our model can successfully predict relative NAD^+ activity after training with only 10s of experimental examples.

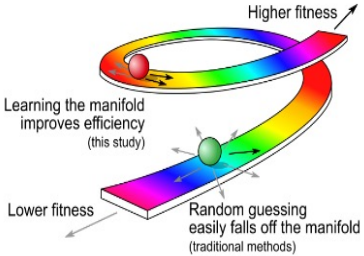




2. Progress and Outcomes: Task 2- Enabling Machine Learning-based Directed Evolution

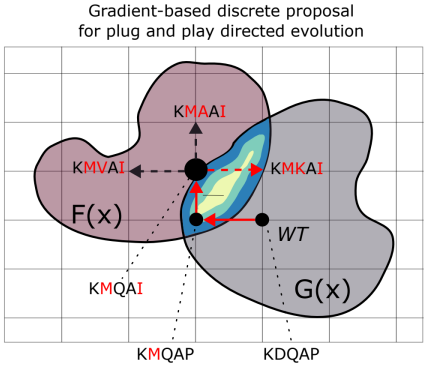
RELEVANCE: Develop tools to facilitate ML-assisted enzyme engineering

OUTCOME: ML-based directed evolution platform that can be applied across the BETO project portfolio for enzyme engineering



Hie et al. *bioRxiv 2022*

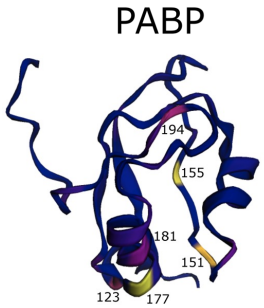
- Large majority of mutations detrimental to protein function. Direct sampling from this large, discrete space requires new computational tools.
- Combining unsupervised ***evolutionary density*** **models** with ***supervised fitness*** **models** (650M parameters) can help generate new sequence predictions.
- Adding density model helps prevent finding adversarial attacks.
- Demonstrated success on 3 proteins with mutation data available –better results than other models with much smaller training set.

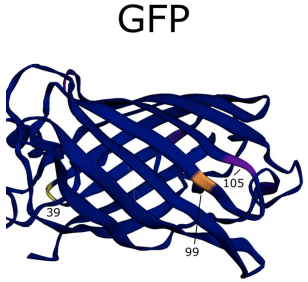


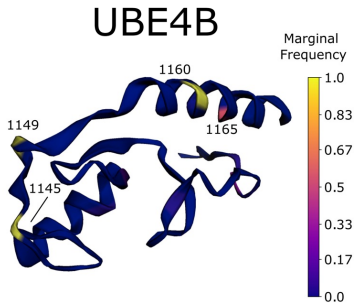
Product of experts
$P(x) = F(x)G(x)$

F(x): Unsupervised evolutionary distribution
Masked Language Model (e.g., ESM2), Potts model, etc.

G(x): Supervised protein activity distribution
CNN, MLP, Transformer, etc.







Work with Patrick Emami and Dave Biagioni, NREL (LDRD)

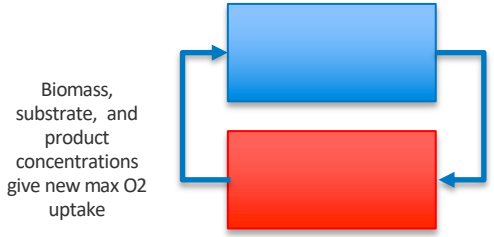
In review at *Machine Learning: Science and Technology* (2023)

2. Progress and Outcomes: Task 3 - Enabling modeling of fermentation scale up

RELEVANCE: Difficult to model fermentation scale up and produce models that can handle the fermentation and diffusion time scale mismatch.

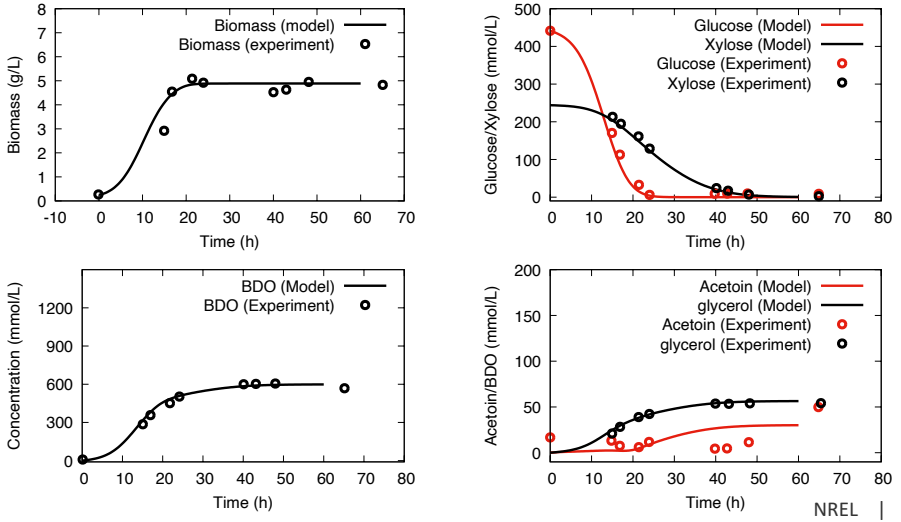
Outcome: Improved subcycling methodology and minimal surrogate metabolic model to make overall model faster.

Developed a novel reactor subcycling methodology to circumvent fermentation and diffusion time scale mismatch capture batch dynamics over a long batch reaction.



- Improved reduced surrogate developed from previous metabolic modeling efforts.
- Reaction model for sugars to BDO
  - Species: Microbes, glucose, xylose, acetoin, glycerol, BDO, oxygen.
- Captures important effects
  - High O_2 -> more acetoin, less BDO
  - Low O_2 -> more glycerol, less BDO

Reaction model compared to experimental data

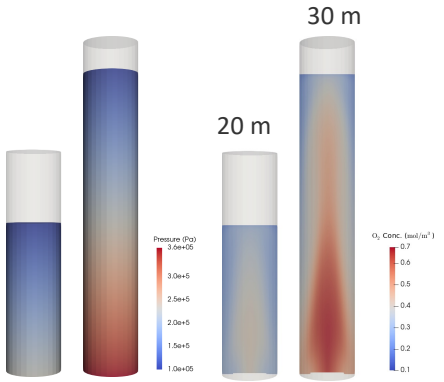


2. Progress and Outcomes: Task 3 - Enabling 2,3 Butanediol (BDO) production at scale

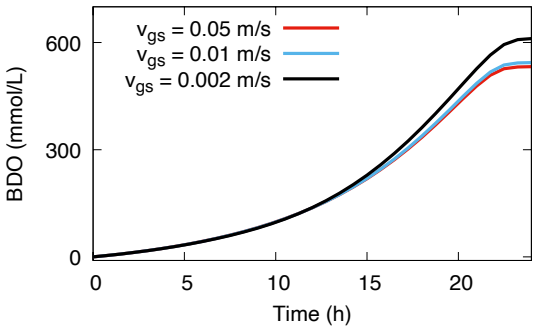
RELEVANCE: Predicting and driving the scale-up of fermenters for 2,3 BDO production. Inform techno-economic analysis.

Outcome: Identified fermenter configurations to maximize 2,3 BDO at scale.

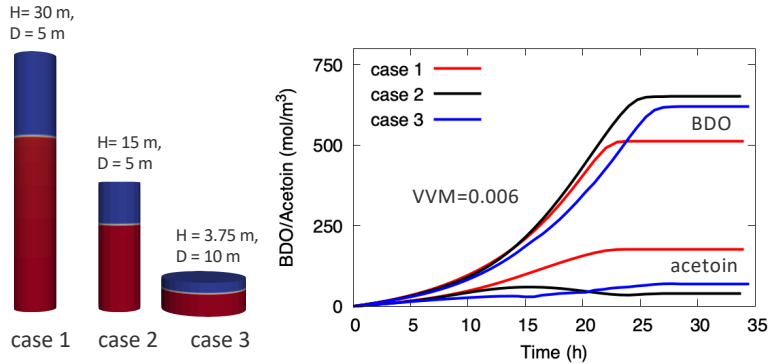
- Coupling multiphase flow with microbial bioreaction
  - Disparate time scales: reactions ~ hours, fluid-dynamics ~ 100 sec.
  - Subcycling scheme: steady-state fluid-solves interleaved with reaction update.



Higher O_2 conc. in taller reactors



Higher BDO at lower aeration rates



Optimizing reactor geometry for BDO production

- **Lower aeration rates and shorter bubble columns favor higher BDO production**
  - Taller columns (case 1) leads to higher O_2 concentration at the bottom triggering more acetoin and less BDO production.
  - Pond like reactor (case 3) leads to lower radial mass transfer.
  - Intermediate case (case 2) results in highest BDO yield (20% improvement).

3. Impact

Reduce Cost of Research and Time-to-Solution

- Computational tools to characterize lignin-derived compounds for subsequent upcycling to lower the cost of conversion.
- Plastics decrystallization free energies in **different solvents to help guide deconstruction strategies**.
- **Enhanced machine learning models to enable enzyme engineering efforts in multiple projects.**
- Reactor models **predict BDO fermentation yield at scale (~ 500 m^3), accelerating transition from lab-scale.**
- Reactor modeling predicts mass-transfer effects at scale and provides optimal reactor designs.

Provide New Insights

- **TEA enhanced by accurate models; can now accurately include many reactor design variables at full industrial scale.**
- Sets of mutations to improve protein activity and stability.

The research conducted in this project will directly impact BETO's ability to achieve its 2030 goal of 3B gallons SAF and 2050 goals of 60B gallons renewable hydrocarbon fuels and 40B pounds of renewable chemicals by providing new solutions and enabling other projects in the BETO portfolio.

3. Impact

Technology Transfer/Scientific Dissemination

- LigninWrangler tool and libraries publicly available for lignin deconstruction technologies.
- Omics methods and metabolic modeling visualization tool released for public use.
- Reactor models and lignin models are publicly available for industrial use.
- Record of Inventions (1) and publications (8)

DEIA

- Provided training in machine learning and machine vision to a graduate student who joined our project from the GEM program.
- Actively distributed presentations on the impact of modeling in SAF research to Minority Serving Institutions to motivate students to pursue scientific studies with an emphasis in modeling.

Summary

Management:

- **Active management** with reevaluation of efforts every year done by communicating with other projects to **ensure that we are working on the most tractable and impactful projects**.

Approach:

- Target **most relevant bottlenecks and barriers** in most BETO-relevant processes.
- Use a **MultiScale Approach**: Molecular (Task 1), Metabolic/Cellular (Task 2), and Macroscopic (Task 3) simulation and leverage EERE computer resources

Progress and Outcomes (highlights):

- LigninWrangler, a computational tool for identification of closely related lignin-derived compounds.
- Guidance for plastics upcycling strategies of semicrystalline commodity plastics.
- Inclusion of evolutionary-scale language models to improve accuracy of sequence -> function prediction.
- Enabled ML guided directed evolution and protein engineering from sparse experimental datasets.
- Enabled simulation of batch fermentation at scale.

Impact:

- Reduced cost of research and time-to-solution for 2,3 BDO production, plastic depolymerization, cell free biocatalysis, and lignin utilization.
- Provided new insights to TEA team for scaling up 2,3 BDO production.
- Enabled technology transfer and scientific dissemination.

Quad Chart Overview

Timeline

- Project start date - 2021
- Project end date 2024

	FY20	Active Project
DOE Funding	(10/01/2021 – 9/30/2024) $1,000,000	$3,000,000 (FY21-FY24)

Project Partners*

Barriers addressed

Ct-N Multiscale computational framework accelerating technology

Ct-C Process Development for Conversion of Lignin

Ct-F Increasing the yield from catalytic processes

Ct-G Decreasing the cost to developing novel ind. relevant catalysts

Ct-K Developing methods for Co-product Production

Ct-L Decreasing devel. time for ind. relevant microorganisms

Ct-M Current reactors are not designed to handle many harsh conditions

Project Goal

Provide **actionable guidance** to experiments and TEA from mechanistic predictions and design principles:

- Mutations for enzymes
- Metabolic target products
- Selection of solvents for polymers breakdown
- Metabolic knockouts and insertions
- Reactor optimizations

Reduce research time and cost, **increasing efficiency,** using **theory, modeling, and simulation** to examine experimentally inaccessible solution space.

End of Project Milestone

Combine reactor design, new engineered enzyme designs, kinetic modeling to enable the production of pinene, limonene, and/or bisabolene at 25g/L from hydrolysate or ethanol, or 25g/L f from mevalonic acid with more than 90% yield with complete cofactor recycling (with **CFIT)**

Funding Mechanism

AOP as WBS# - 2.5.1.100

Acknowledgments

Funding

- U.S. DOE EERE Bioenergy Technologies Office
  - HQ: TM - Ian Rowe
  - NREL LPM and Platform Lead: Zia Abdullah, Mike Guarnieri

NREL Project Members

Vivek Bharadwaj
Yannick Bomble
Lintao Bu
Brandon Knott
Jeff Law
James Lischeske
Heather Mayes
Hari Sitaraman
Peter St. John
Mohammad Rahimi
Jonathan Stickle
Josh Vermaas
Chao Wu

Collaborations with other BETO Projects

- Biological Upgrading of Sugars (BUS)
- Lignin Utilization (LU)
- Targeted Microbial Development (TMD)
- Biochemical Platform Analysis (BPA)
- Cell Free and Immobilization Technologies (CFIT)
- Bench-Scale R&D (BSRD)
- Separations Consortium (SepCon)
- BOTTLE Consortium
- Agile BioFoundry (ABF)
- Consortium Comp Chem and Phys (CCPC)

www.nrel.gov

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Responses to Previous Reviewers' Comments

Q: Collaboration with 13 projects is a substantial undertaking, and this magnitude of workload will need to be resourced appropriately. It feels like perhaps the project group is spread thin. Self-selection with successful projects appears to be taking place, which is not a compliment or a criticism, but rather purely an observation.

A: We agree that we always need to remind other projects of the value of modeling and what improvements were or could be obtained from modeling to improve processes. We do aim to focus on research projects where modeling could have the most impact, and to do so we are in constant discussion with experimental project leads to demonstrate that modeling can help improve their bottom line. The risk associated with this project are twofold; the first is to consider too many projects and not focusing on the ones where modeling can have impact, which could lead to unproductive use of our resources. This is why we take the time to think about our approach in light of what our experimental collaborators in BETO need. The second is to not yet have the appropriate software, methods, and/or hardware to tackle a problem, which usually prompts us to seek fruitful external collaborations, which we have done successfully to BETO's benefit in the past.

Q: One area where the team should consider focusing more on is reactor modeling and computational fluid dynamics. Such simulations can greatly aid in fermentation scale-up. Also, such models are being ignored by most academic researchers (especially in the United States), so this is one area where the team can really drive the field and advance the state of technology. Lastly, it was unclear whether the team was developing open-source tools that will benefit the broader community. Disseminating their tools will improve the impact of this project.

A: Computational fluid dynamics is indeed an area of our project that we think has a lot of promise, notably to help with reactor design for atypical production routes including complex fermentations and cell-free biocatalysis. Scientific dissemination is an important aspect of our research, and we do make all the codes, force fields, and other computational tools available to the public.

Responses to Previous Reviewers' Comments

Q: The project's history of invention disclosures and publications are strong benchmarks. To some degree, the project is purely computational, even though it is informatics to a broad suite of users. The project staff might benefit from some cross-training or cross-functional activities. To get a better perspective and a deeper understanding, we find it to be valuable for our accountants to regularly tour the production facilities

A: We do agree that cross-training can be very important. For the last few years, we have encouraged researchers work on this project to attend and actively participate in relevant experimental group meetings to get their perspective and also to better identify needs to remain relevant.

New publications/IP since last review

- Roman Brunecky,[+] Brandon C. Knott,[+] Venkataramanan Subramanian, Jeffrey G. Linger, Gregg T. Beckham, Todd A. Vanderwall, Vladimir V. Lunin, Fei Zheng, Mercedes Garrido, Logan Schuster, Emily Fulk, Samuel Farmer, Michael E. Himmel, and Stephen R. Decker. "Natural diversity screening of glycoside hydrolase family 7 cellobiohydrolases." *In preparation.*
- Tucker Burgin, Benjamin C. Pollard, Brandon C. Knott, Heather B. Mayes, Michael F. Crowley, John E. McGeehan, Gregg T. Beckham, H. Lee Woodcock. "The reaction mechanism of the Ideonella sakaiensis PETase enzyme." *In preparation.*
- Heather B. Mayes, Brenna A. Black, William E. Michener, Rui Katahira, Xueming Dong, Grace Dorgan, John Ralph, and Gregg T. Beckham. LigninWrangler: An open-source tool for lignin modelling and characterization. *In preparation.*
- Heather B. Mayes, Brandon C. Knott, Nicholas A. Rorrer, Michael F. Crowley, and Gregg T. Beckham. Bond dissociation energies of commercial plastics as a roadmap for upcyling strategies. *In preparation.*
- Emami, Patrick, Aidan Perreault, Jeffrey Law, David Biagioni, and Peter C. St John. "Plug & Play Directed Evolution of Proteins with Gradient-Based Discrete MCMC." **2023** (Under Review at *Machine Learning: Science and Technology*)
- Xueming Dong,[+] Heather B. Mayes,[+] Kris Morreel, Rui Katahira, Yanding Li, John Ralph, Brenna A. Black, Gregg T. Beckham. Energy-Resolved Mass Spectrometry as a Tool for Identification of Lignin Depolymerization Products. *ChemSusChem*. **2023**, 16 (1) e202201441.
- Taku Uchiyama, Takayuki Uchihashi, Takuya Ishida, Akihiko Nakamura, Josh V. Vermaas, Michael F. Crowley, Masahiro Samejima, Gregg T. Beckham, Kiyohiko Igarashi. Lytic polysaccharide monooxygenase increases cellobiohydrolases activity by promoting decrystallization of cellulose surface. *Science Advances.* **2023**, 8 (51) eade5155.

New publications/IP since last review

- David Biagioni, Charles Edison Tripp, Struan Clark, Dmitry Duplyakin, Jeffrey Law, Peter C St John, “graphenv: a Python library for reinforcement learning on graph search spaces”, Journal of Open Source Software **2022**, 7, 4621
- Josh V. Vermaas, Michael F. Crowley, Gregg T. Beckham. Molecular simulation of lignin-related aromatic compound permeation through gram-negative bacterial outer membranes. *Journal of Biological Chemistry*. **2022**, 298(12) 102627.
- Christopher W Johnson, Peter C. St John, Gregg T. Beckham, Joshua R. Elmore, Adam M. Guss, Davinia Salvachua Rodriguez, Gayle J. Bentley, V George Lee Peabody, Taraka Dale, Ramesh K Jha, Niju Narayanan, “Engineered microorganisms for the production of intermediates and final products” (Patent U.S. **2022** 11518975).
- Andrew D McNaughton, Erin L Bredeweg, James Manzer, Jeremy Zucker, Nathalie Munoz Munoz, Meagan C Burnet, Ernesto S Nakayasu, Kyle R Pomraning, Eric D Merkley, Ziyu Dai, William B Chrisler, Scott E Baker, Peter C St. John, Neeraj Kumar, “ Bayesian Inference for Integrating Yarrowia lipolytica Multiomics Datasets with Metabolic Modeling” ACS Synthetic Biology **2021**, 10 2968
- Chao Wu, Ryan Spiller, Nancy Dowe, Yannick J Bomble, Peter C St. John, Frontiers in Bioengineering and Biotechnology **2021**, 9 707749