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

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

# Project Overview

***Context:***

1. **Lower time to solution.**
2. **Explore solutions inaccessible experimentally.**

## This project focuses on the following:

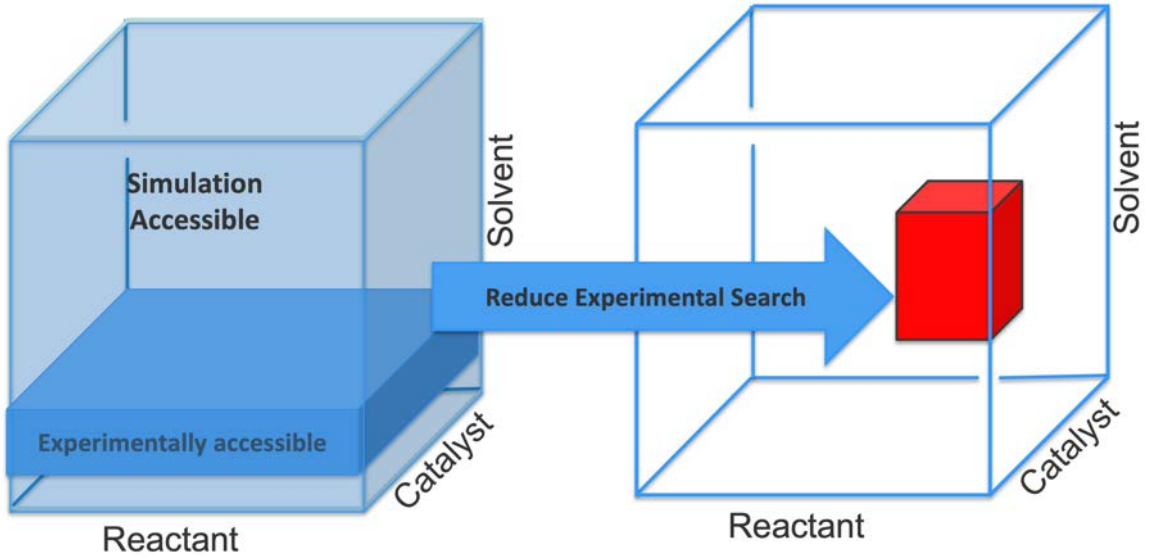
- Improving biomass, plastic degrading, and metabolic enzymes.
- Engineering and modifying metabolic pathways.
- Down selecting biochemical targets.
- Determining best fermentation conditions and providing reliable models for TEA

## *Heilmeier Catechism :*

- **What:** Reducing research time and cost, **increasing efficiency,** using modeling and simulation to provide **actionable guidance to experimental efforts.**
- **Today:** Modeling often conducted without discussion and input from experimentalists.
- **Importance: Modeling can drastically reduce time to solutions and allow new breakthroughs.**
- **Risk:** Considering too many projects and not focusing on the ones where modeling could have impact.

# Modeling Relevance

**Lower time to solution by reducing experimental work and time**.
Solution space is too big for experiment but accessible by modeling.

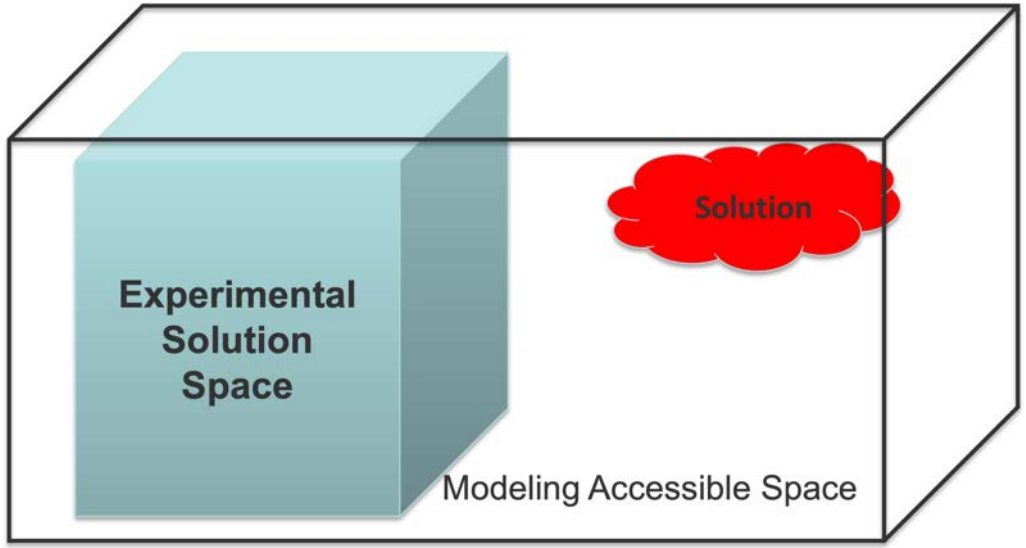



Example: Determine the right aliphatic compound to produce based on ease of extraction before experimental efforts are implemented.

# Modeling Relevance

**Solutions inaccessible experimentally**

Modeling can find solutions unavailable to standard experimental search.




Examples: Risk too high

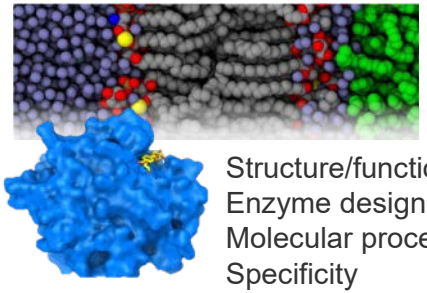
- Mutations/knockouts believed to be fatal to microbes.
- Testing reactor designs at industrial scale.
- Exploring mutants with very high number of mutations.

# 1. Approach: Management

**Project: Biochemical Process Modeling and Simulation – Yannick Bomble**

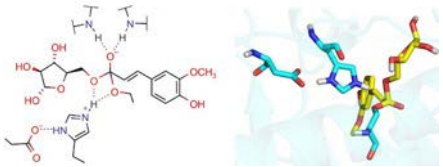
## 1- Molecular Modeling - Brandon Knott

Molecular dynamics
Quantum mechanics



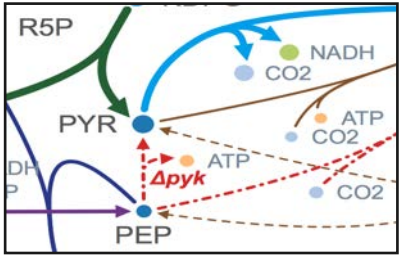
Structure/function
Enzyme design
Molecular processes
Specificity

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

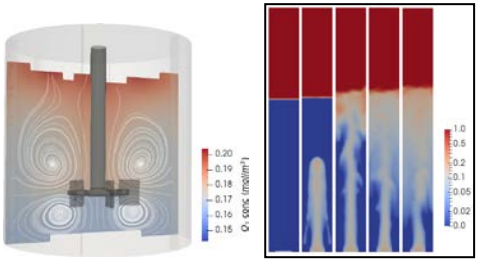



New subcontract with Sophie Barbe at INRA (France) “ML guided multistate modeling for protein engineering” ($60K)

## 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)



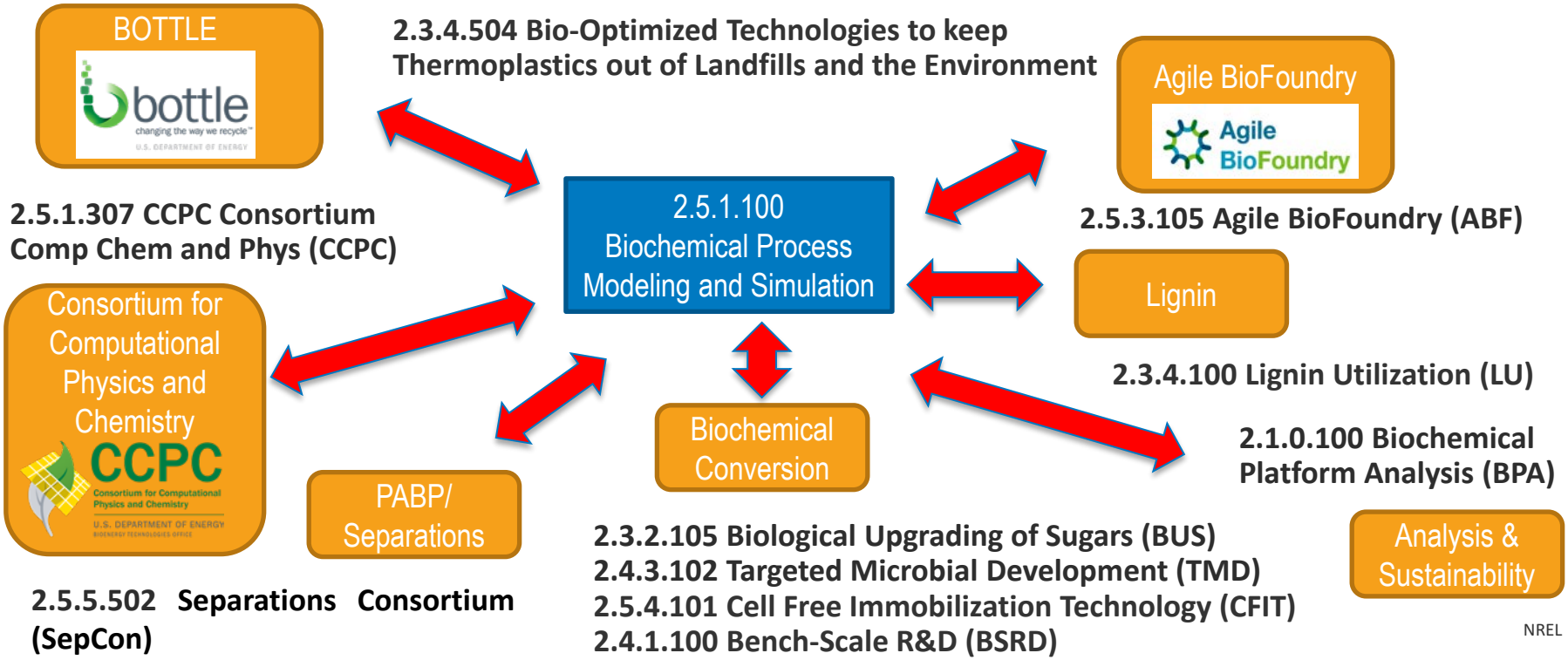
Project split into tasks by modeling type, managed by person with appropriate expertise
Task Managers responsible for:

- **Relevance**
- AOP, Milestones, quarterly reporting according to the guidance of BETO
- **Communication with other projects**
- Tracking go/no-go activities
- Budget management

# 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.

Many interactions across the BETO project portfolio that are **revised at each AOP cycle by meeting with project PIs** to see if modeling is appropriate/tractable and if we have the proper resources.



# 1. Approach: Management

**Objectives:**

- Gain **insight**, discover new approaches and solutions.
- **Guide and stimulate design, experiment, and engineering;** select most promising directions.
- **Accelerate research, provide complementary insight and broaden research space.**

**Approach:**

- Use MultiScale Approach:
  Molecular (Task 1), Metabolic/Cellular (Task 2), and Macroscopic (Task 3) simulation.
- Leverage EERE computer resource: **Kestrel (NREL).**
- 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.**
- Target most **relevant bottlenecks and barriers** in **most BETO-relevant processes.**
- **Go/no-go decisions:** Evaluate available experimental datasets that report enzyme performance (activity, stability, etc.) along with immobilization techniques. Determine whether sufficient data exists to correlate experimentally verified performance with metrics observable in molecular dynamics (MD) simulations with ML.

# 1. Approach: Science

**Challenges:**

- 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.
    - We are recruiting new students including those in MSIs.
    - Placed 2 new subcontracts to ensure continuity in our research.

**Success Factors:**

- Insights achieved, solutions found, unproductive efforts avoided.
- Reduced time to solution: increasing titer, efficiency, speed, performance.
- New routes to advanced fuels and co-products.

# DEIA efforts

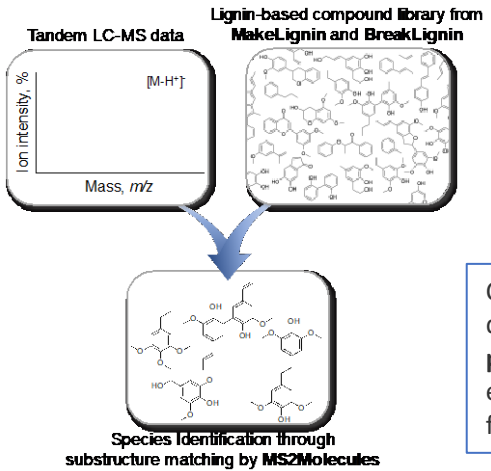
- Developed a presentation series to demonstrate the impact that modeling can have on the sustainable production of renewable SAF and biochemicals. Actively distributed these presentations to Minority Serving Institutions to motivate these students to pursue scientific studies with an emphasis in modeling.
- We applied and were selected to mentor interns for the STAR program (Student Training in Applied Research). NREL's STAR internship program aims to bridge gaps in education, research, and public service to create career pathways for undergraduate students and build long-term partnerships with minority-serving institutions (MSI) to deliver collaborative research opportunities.
- We actively recruited interns from the GEM program. One graduate student joined the program to work on machine vision to automatically analyze experimental results from TEM and CRYO-EM datasets.
- Advertising job openings for student and postdoc in MSIs and other institutions with a population underrepresented in STEM.

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

**RELEVANCE: Characterizing lignin-derived compounds is critical for subsequent upcycling**

**OUTCOME: LigninWrangler, a publicly available user-friendly computational tool to assign molecular formula. Identification of closely related compounds coupling analytics with computation**

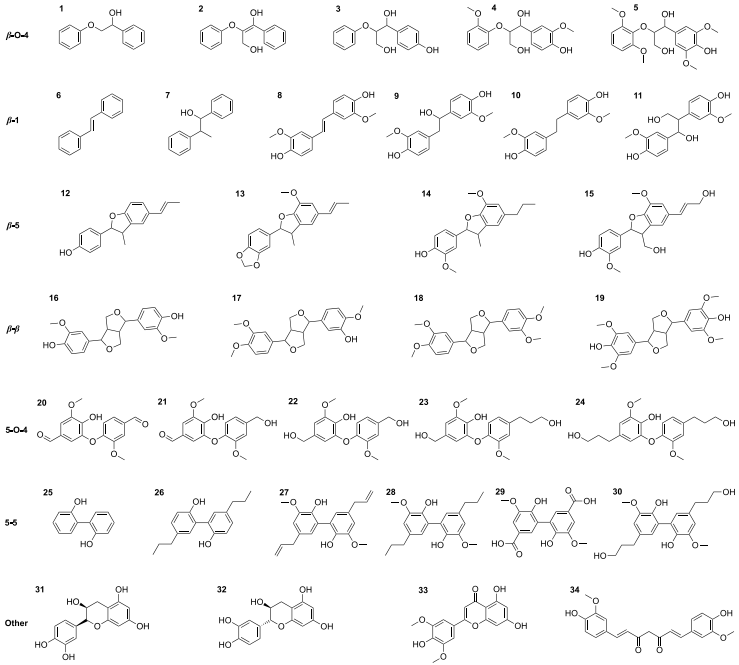
- Lignin heterogeneity is a significant challenge for separations and conversion efforts
- Enhanced analytical technique such as QTOF-ERMS are very promising to identify lignin-derived compounds. However, the interpretation of the data is still difficult.
- To interpret high accuracy mass spec data, our computational tool (**LigninWrangler**) can dramatically help and assign molecular formulae and 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**

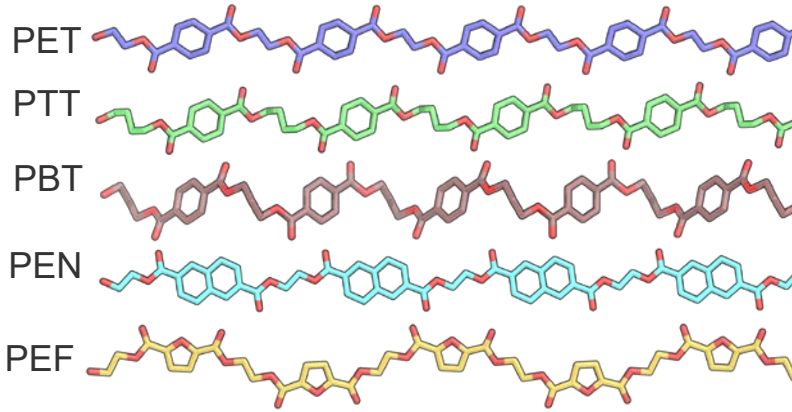
Mayes *et al. in preparation*

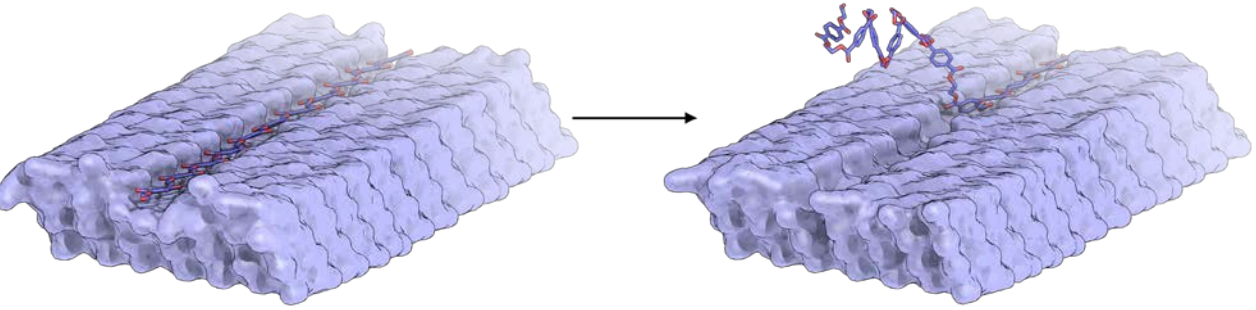
# 2. Progress and Outcomes: Improving plastics decrystallization

**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. Guidance for plastics upcycling strategies.**

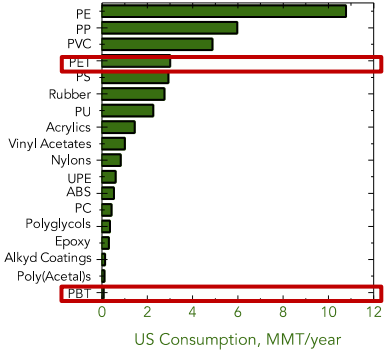
- Molecular dynamics based procedure to compute the free energy to decrystallize a single chain.
- Commercial soft drink bottles are about 16%, but PET, polyolefins, and nylons can be > 50%.
- Next steps will target non-aqueous solvents, and other polymer classes (polyolefins, nylons).



Polymer chain in crystal surface

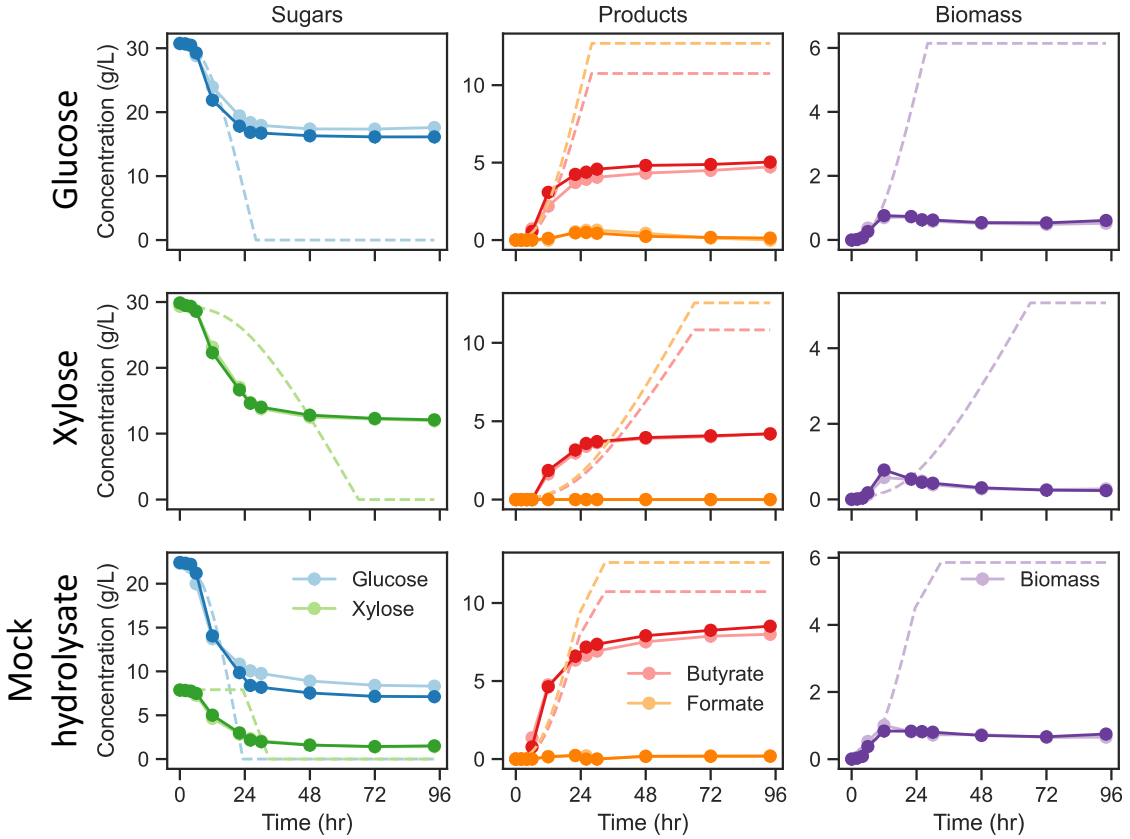
Polymer chain extracted from surface




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

# 2. Progress and Outcomes: Enabling butyric acid production in *C. thermobutyricum*

- Developed a model of substrate uptake and cell growth in *C. thermobutyricum* for the production of butyric acid.
- Model captures early stages of exponential growth but misses presence of product inhibition.
- Simulations suggest extractive fermentation or adaptive evolution to improve butyrate tolerance are potential improvements for increasing titers and yields




DFBA trajectories (dashed lines)
Experimental results (solid lines)

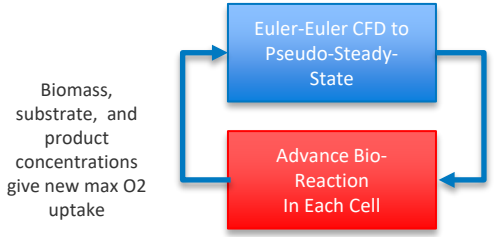
# 2. Progress and Outcomes: Enabling modeling of fermentation scale up (2,3 BDO)

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

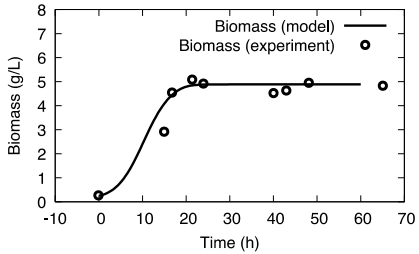
**Outcome: Improved subcycling methodology and minimal surrogate metabolic model to make overall model faster.**

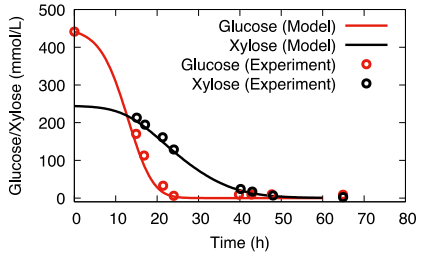
- Improved reduced surrogate developed from previous metabolic modeling efforts.
- Reaction model for sugars to BDO
  - Species: Microbes, Glucose, Xylose, acetoin, glycerol, BDO, oxygen.
- Captures important effects
  - High $O_2$ -> more acetoin, less BDO
  - Low $O_2$ -> more glycerol, less BDO

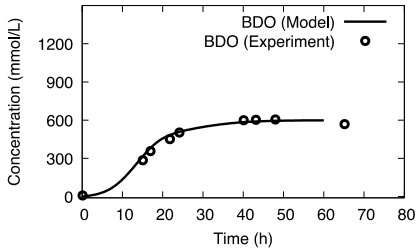
Developed a novel reactor subcycling methodology to circumvent fermentation and diffusion time scale mismatch capture batch dynamics over a long batch reaction.

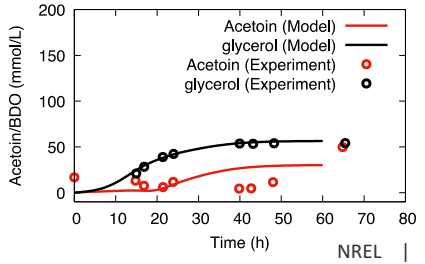



Reaction model compared to experimental data



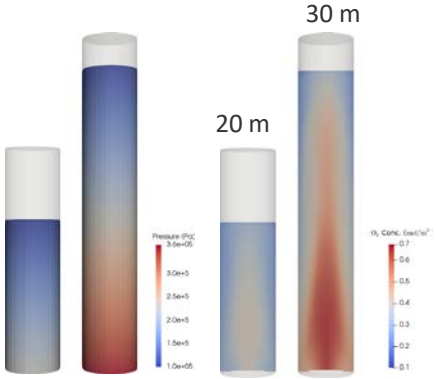



# 2. Progress and Outcomes: Enabling 2,3 Butanediol (BDO) production at scale

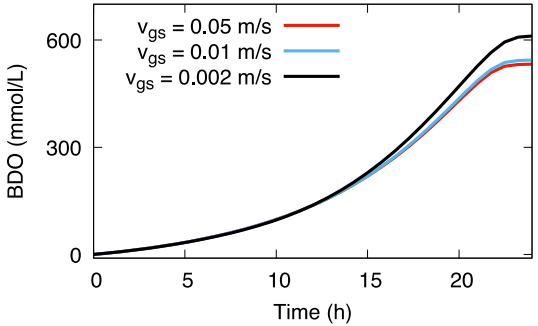
**RELEVANCE: Predicting and driving the scale-up of fermenters for 2,3 BDO production. Inform techno-economic analysis.**

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

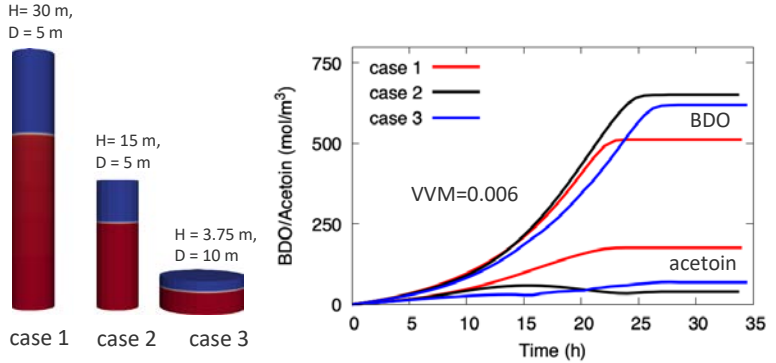
- Coupling multiphase flow with microbial bioreaction
  - Disparate time scales: reactions ~ hours, fluid-dynamics ~ 100 sec
  - Subcycling scheme: steady-state fluid-solves interleaved with reaction update




Higher $O_2$ conc. In taller reactors




Higher BDO at lower aeration rates




Optimizing reactor geometry for BDO production

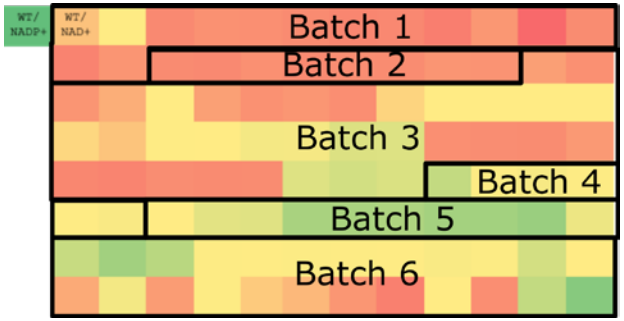
- Lower aeration rates and shorter bubble columns favor higher BDO production
  - Taller columns (case 1) leads to higher O2 concentration at the bottom triggering more acetoin and less BDO production
  - Pond like reactor (case 3) leads to lower radial mass transfer
  - Intermediate case (case 2) results in highest BDO yield (20% improvement)

# 2. Progress and Outcomes: Using Machine Learning to improve key enzymes

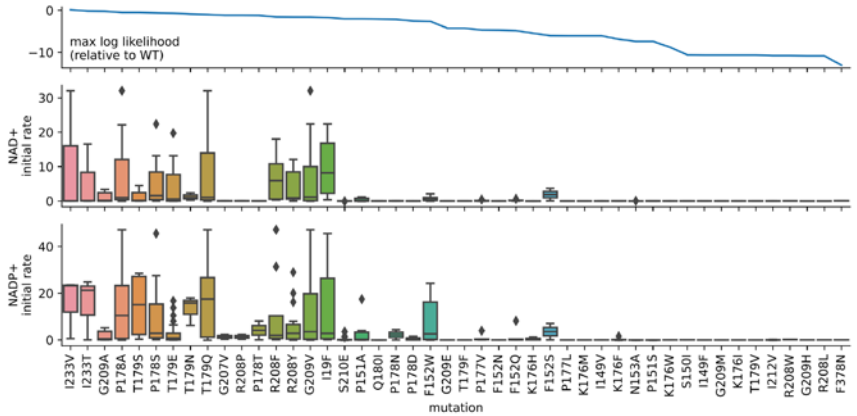
**RELEVANCE: Improve key enzymes for *in vivo* or cell-free systems**

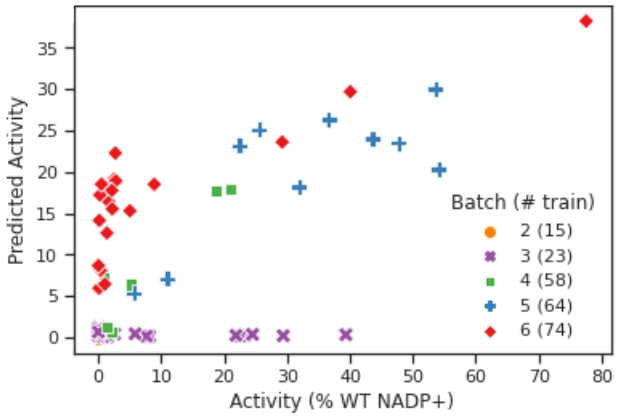
**OUTCOME: Demonstrate utility of machine learning (ML) approach from sparse experimental datasets**

- GapN essential for cell-free biocatalysis, but $NADP^+$ cofactor labile, difficult to recycle, and restricts choice of pathways.
- Achieved >80% WT activity by hand-crafting mutants with MD simulations over 6 rounds of experimental batches
- Evolutionary-scale language models can guide enzyme engineering toward functional mutations even with sparse experimental data.




- Model successfully predicts relative $NAD^+$ activity after training with 10s of experimental examples



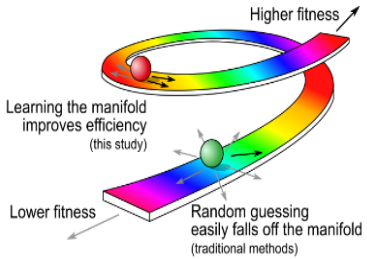



# 2. Progress and Outcomes: Enabling Machine Learning-based Directed Evolution

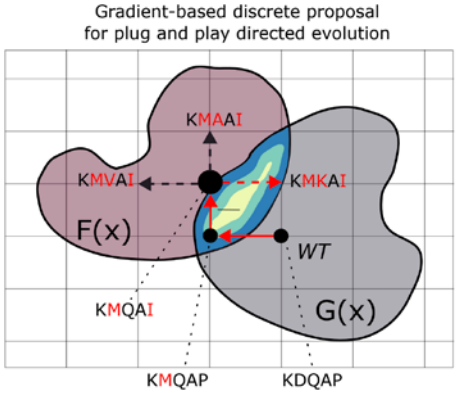
## RELEVANCE: Develop tools to facilitate ML-assisted enzyme engineering

## OUTCOME: ML-based directed evolution platform

- 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



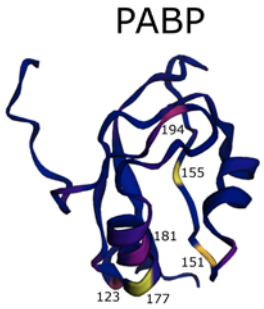

Hie et al. *bioRxiv 2022*

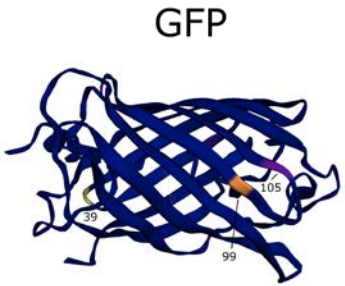


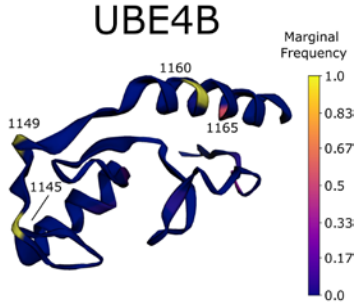

Product of experts
$P(x) = F(x)G(x)$

| 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)

# 3. Impact

## Reduce Cost of Research and Time-to-Solution

- Computational tools to couple with lignin analytics for high mass accuracy species identification.
- Plastics decrystallization guiding deconstruction strategies for circular economy.
- **Enhanced machine learning models to improve enzyme engineering.**
- 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

## Technology Transfer

- 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 and publications (15 since last peer review, 2 patents, 2 software release).

# 3. Impact

## Reduce Cost of Research and Time-to-Solution

- Computational tools to couple with lignin analytics for high mass accuracy species identification.
- Plastics decrystallization guiding deconstruction strategies for circular economy.
- **Enhanced machine learning models to improve enzyme engineering.**
- 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

## Technology Transfer

- 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 and publications (15 since last peer review, 2 patents, 2 software release).

# 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 resource

## Impact:

- **Reduced cost of research and time-to-solution**
- **Provided NEW insights**
- Enabled technology transfer and scientific dissemination

## Progress and Outcomes (highlights):

- LigninWrangler, a computational tool for identification of closely related lignin-derived compounds
- Guidance for for plastics upcycling strategies of semicrystalline commodity plastics
- Natural language processing-based techniques 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.

# Quad Chart Overview (for AOP Projects)

## 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 30g/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
Michael Crowley
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 Lignin Valorization (BLV)
- Biological Upgrading of Sugars (BUS)
- Lignin Utilization (LU)
- Targeted Microbial Development (TMD)
- Low Temperature Advanced Deconstruction (LTAD)
- 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)

# Thank You

**www.nrel.gov**

NREL/PR-2700-85595

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.

# Past Responses to Previous Reviewers' Comments

Q: Bringing together metabolic model dynamic process modeling and reactor design is a great ideas and will help reduce the timeline from concept to product. The project will benefit from focusing on one of the 3 tasks instead of working on all 3 together?

A: Our history has shown that the combination has been very successful in all three tasks and allows us to provide a multiscale approach to experimental bottlenecks from reaction mechanism to industrial scale reactors. Often these scales overlap in a question and require expertise at each scale to work together such as in the BDO work. This project does not require the direct attention of the PI in all three tasks. Each task lead has the responsibility of concentrating on achieving the goals of each task and in collaborating with experimental efforts in BETO.

Q: The models are impressive but there seems to be a lack of validation of the models maybe because the data necessary for direct validation is difficult to obtain.

A: This is a good point. It is always the most satisfying and constructive to get full validation in an effort to refine models. In BETO efforts, time is often at a premium so we find the fastest way to get the best answer we can to down-select out the worst candidates and to provide design principles at the cost of rigorous, validated, refined models. Since this approach has proven useful, we will be refining and validating as time and experimental effort permit. Also, direct validation may be difficult in some cases but several of the predictions have proven valid or lead to new thinking that improved the process (*e.g.* amino acid mutations, knockouts, sparging rate).

# New: Responses to Previous Reviewers' Comments

- All catechism aspects are critical and appear to be getting logical and reasonable consideration. The progress and outcomes have shown direct benefits to a multitude of projects. 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. I would envision most PIs will see the conspicuous value within the Biochemical Process Modeling and Simulation team and look to engage in these strategies and tools to enhance their own outcomes. It is great to see the creation and deployment of tools, not only for research purposes or for BETO directly, but for broader, multiple industry use such as the publicly available reactor models and new computational approaches. 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. Great project, well done!
- Management plan is adequate and requires good communication with other groups that can benefit from the tools being developed. Mitigation and risks are not addressed adequately in the review update. Several approaches have been deployed in several areas to assist with model prediction. It is not clear that this effort is adequately funded to deliver timely information. It is also not clear whether the team gets involved directly in some of these projects to fine-tune the models while gaining insights from hands-on experimental work. Modeling outputs and projections can be improved if inputs are validated by engaging modelers directly in the research. The relevance to BETO's mission is excellent, as there is a need for timely delivery in a cost-effective manner. Gaining insights and improving technology transfer for scale-up industrial opportunities will require active involvement and will be dependent on personnel and resource availability. Having personnel in the mix with industrial experience will ensure more success upon transfer. Progress and outcomes are provided for several

# New pubs

- 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.
- 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.
- 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."