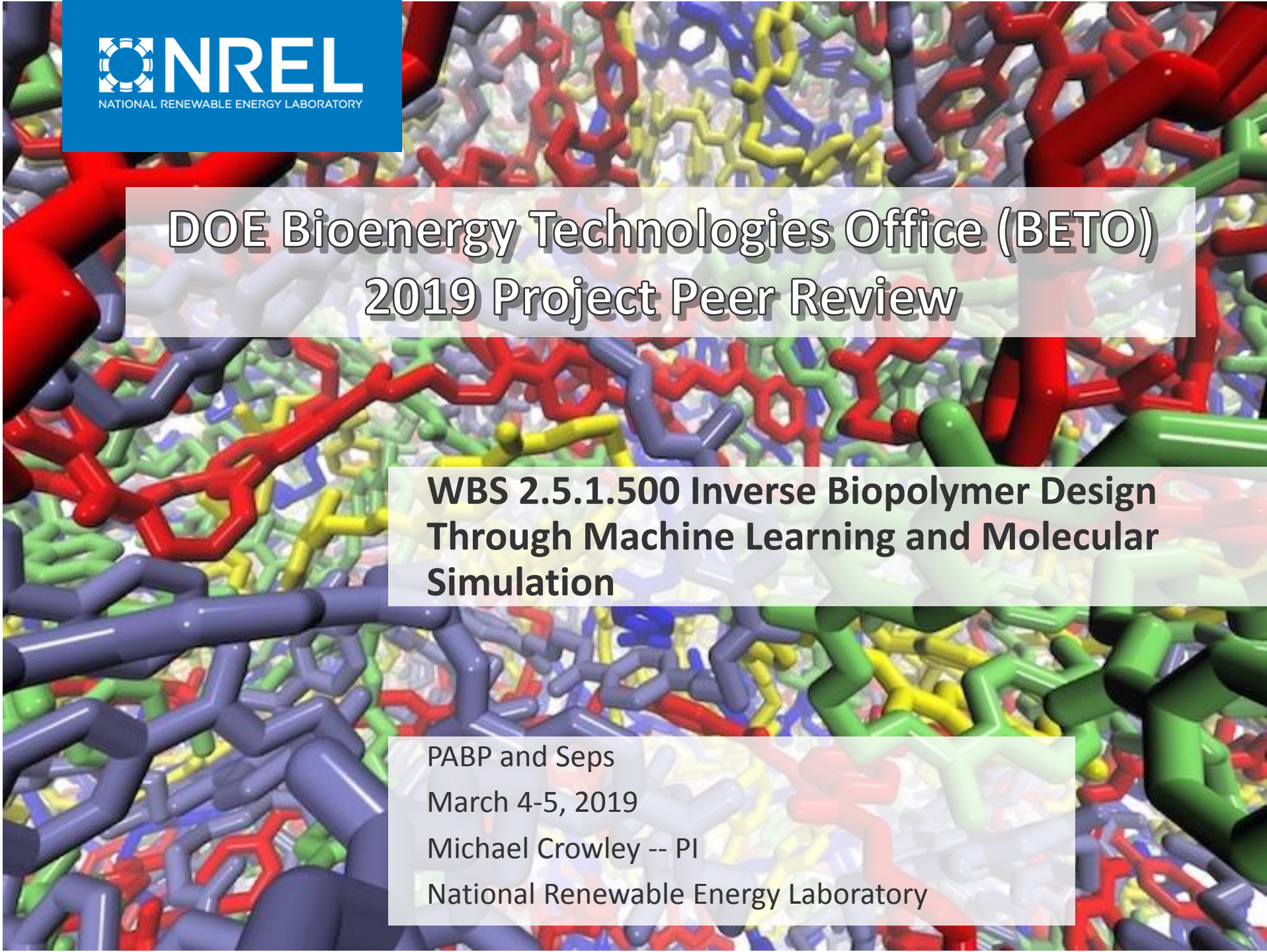


NREL
NATIONAL RENEWABLE ENERGY LABORATORY

DOE Bioenergy Technologies Office (BETO) 2019 Project Peer Review

WBS 2.5.1.500 Inverse Biopolymer Design Through Machine Learning and Molecular Simulation

PABP and Seps

March 4-5, 2019

Michael Crowley -- PI

National Renewable Energy Laboratory

Goal Statement

Goal: Predict performance-advantaged bioproducts using Machine Learning and Molecular Simulation

- Invent new materials, possibly with properties nobody ever dreamed of …
- Determine design principles and structure-function relationships for thermal, mechanical, and
- Leverage chemical variability in sugars, lignin, bio/catalytic intermediates

Outcome: A high-throughput method for predicting polymers for desired performance advantages

- Through prediction, identify 10 new polymer materials that are "performance-advantaged" over petro-derived materials

Relevance: Greatly increase efficiency of designing and producing PABPs, reducing cost and time-to-product.

- Work with industry to ultimately enable new bio-based materials by in-silico high-throughput screening

Flubber
Flying Rubber

Industrial Relevance Example

Needed Polymer property:

Abrasion Resistance
with
Tensile Strength

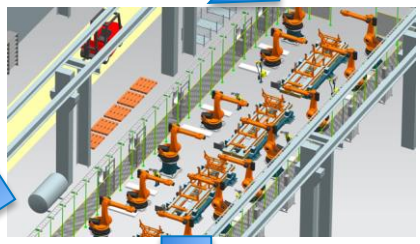


fig. 1 New rope

fig. 2 Used rope

fig. 3 Severely abraded rope

Quad Chart Overview

Timeline

- Start: FY2018
- Merit review cycle: FY2018-2020
- 50% complete of review cycle

	Total Costs Pre Pre FY17	FY17 Costs	FY18 Costs	Total Planned Planned Funding Funding
DOE funded funded	--	--	$400k	$1200k

Partners:

BETO Projects:

Performance-Advantaged Bioproducts via Selective Biological and Catalytic Conversion
Performance Advantaged Bioproducts from Catalytic Fast Pyrolysis
Tailored Polymers Through Rational Monomer Development (LANL)
Analysis in support of novel bio-based products and functional replacements replacements
Lignin-First Biorefinery Development, Biological Conversion of Thermochemical Aqueous Thermochemical Aqueous Streams
Biological Lignin Valorization
Agile BioFoundry
Separations Consortium
Co-optima
HPC NREL

Nat'l labs, universities, companies:

Los Alamos National Laboratory

Barriers addressed

Ct-J Identification and Evaluation of Potential Bioproducts

Computational methods for high-throughput screening and prediction of performance-advantaged polymers

Ct-N Multiscale computational framework accelerating technology

Models must be developed for translating material performance from molecular to industrially relevant scales.

Objective

Produce and characterize PABPs including thermoplastics and thermosets from monomers derived from computational prediction of bio-derived substrates

End of Project Goal

A prediction tool for determining bio-derived polymers with performance advantages resulting in at least 10 **new polymer materials** with performance advantages over what is available from petro-sources.

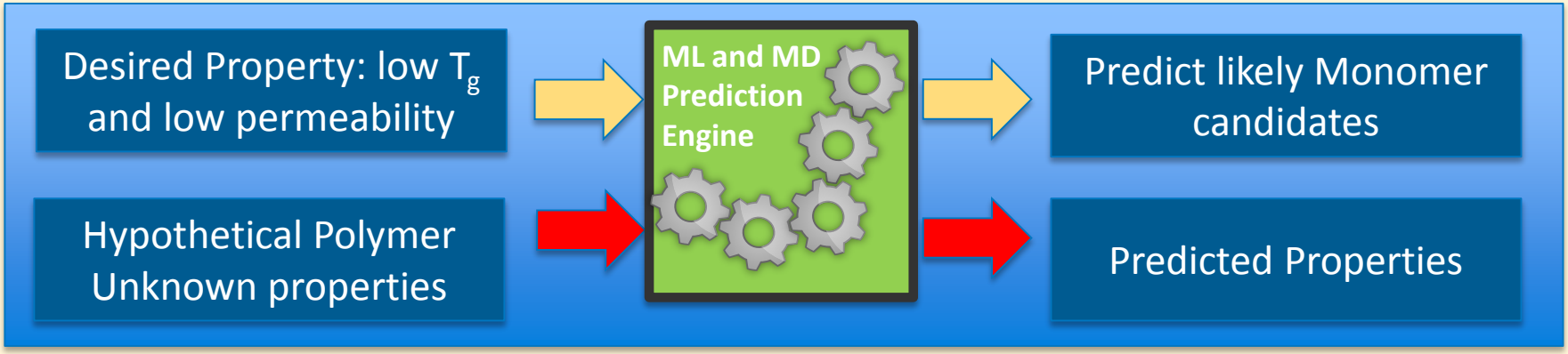
Tool available to industry

Project Overview

History: Novel properties can be predicted through modeling and computation

- Machine Learning (Co-Optima soot precursor prediction)
- Macromolecular simulation (biopolymers, cellulose, lignin)
- Polymer properties (decrystallization, high-T transitions, mechanical properties)

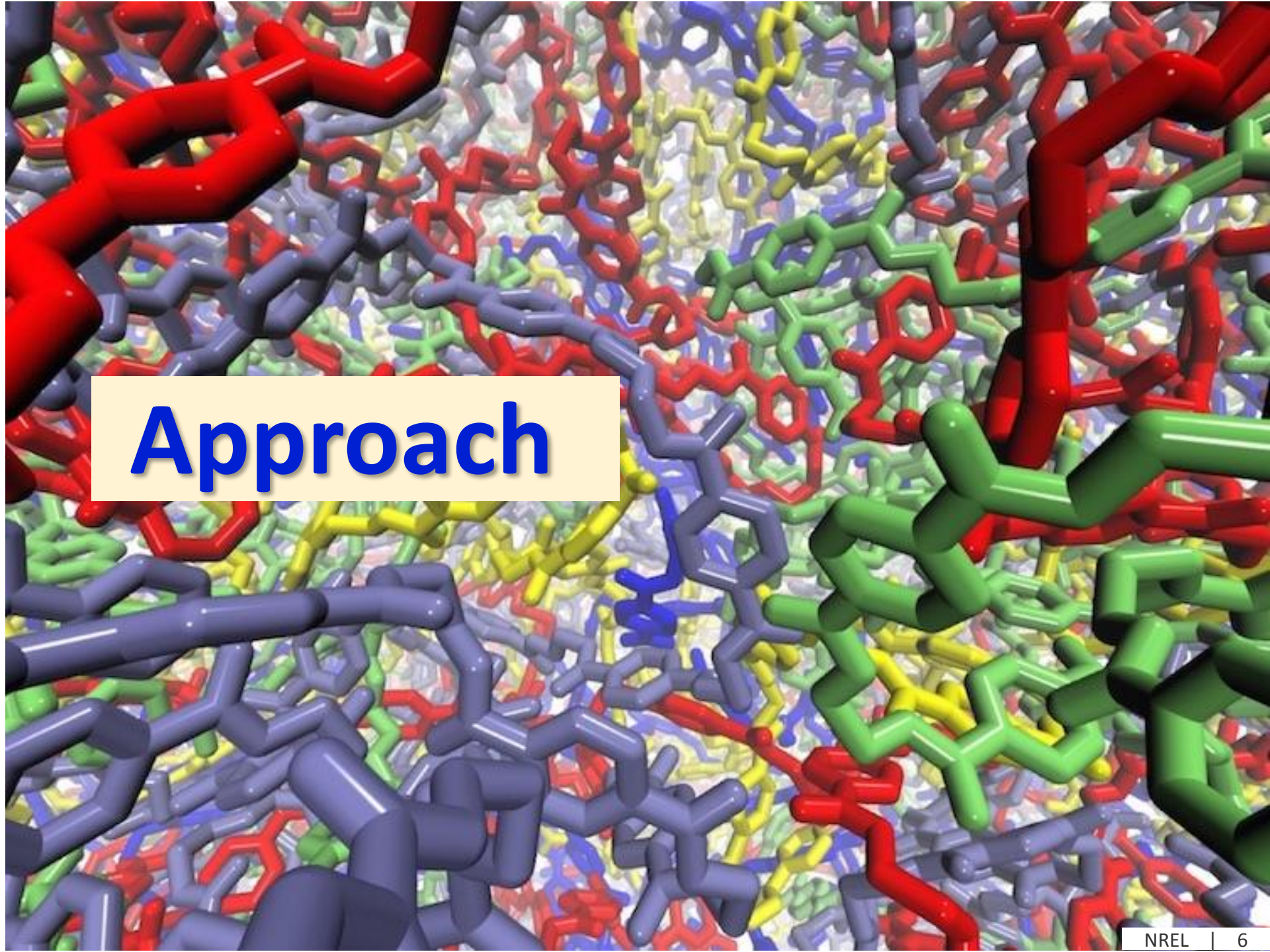
Project Goals: Prediction Engine



- Test Predictions in **Experiment,** understand and predict **Experimental Discoveries**
- Creative advantage: new bio-based intermediates, thermoplastics and thermosets

Context: Harness functionality inherent to biomass

- Industry needs new properties, bio-derived polymers have solutions
- **Machine Learning and Molecular Modeling** can predict polymers with desired properties, and can develop design principles → iterate with experiment


Approach

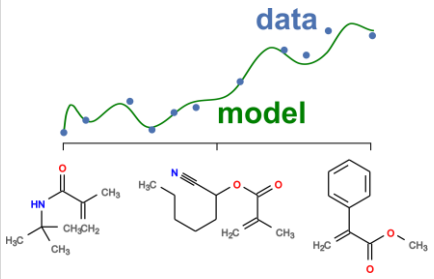
Approach - Management

Project: Inverse Polymer Design
M. Crowley

Machine Learning Task

Peter St. John, Nolan Wilson, Mark Nimlos

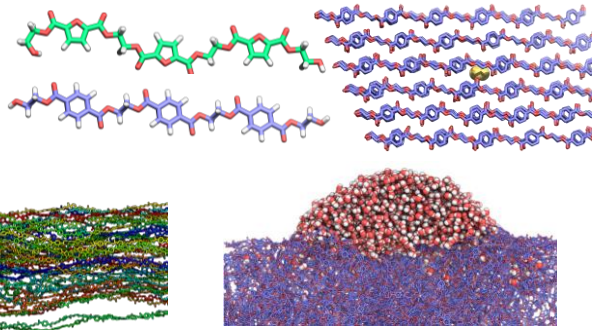
- Polymer Databases
- Rules for Polymer Building from monomers
- Neural Network
- Prediction Engine for high-throughput prediction



Molecular Modeling Task

Brandon Knott, Graham Schmidt

Molecular dynamics
Polymer Physics
Calculate Properties



Management Approach – Collaboration and Meetings

1. **Weekly Task meetings:** progress, problem-solving
2. **Bi-weekly Consortium meetings (three projects together)**:
   1. Performance-Advantaged Bioproducts via Selective Biological and Catalytic Conversion
   2. Performance Advantaged Bioproducts from Catalytic Fast Pyrolysis
   3. Analysis in support of novel bio-based products and functional replacements
3. **Monthly BETO Meetings above plus**
   1. Technical Manager: Nichole Fitzgerald
   2. Tailored Polymers Through Rational Monomer Development (LANL)

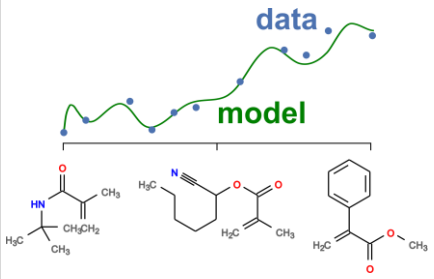
Approach - Management

Project: Inverse Polymer Design
M. Crowley

Machine Learning Task

Peter St. John, Nolan Wilson, Mark Nimlos

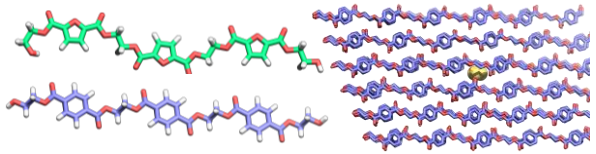
- Polymer Databases
- Rules for Polymer Building from monomers
- Neural Network
- Prediction Engine for high-throughput prediction

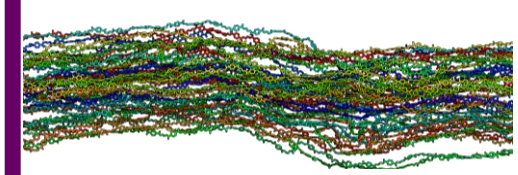


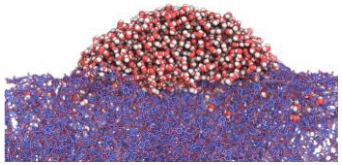
Molecular Modeling Task

Brandon Knott, Graham Schmidt

Molecular dynamics
Polymer Physics
Calculate Properties







Organized by modeling type, managed by expert.
Task Leads responsible for:

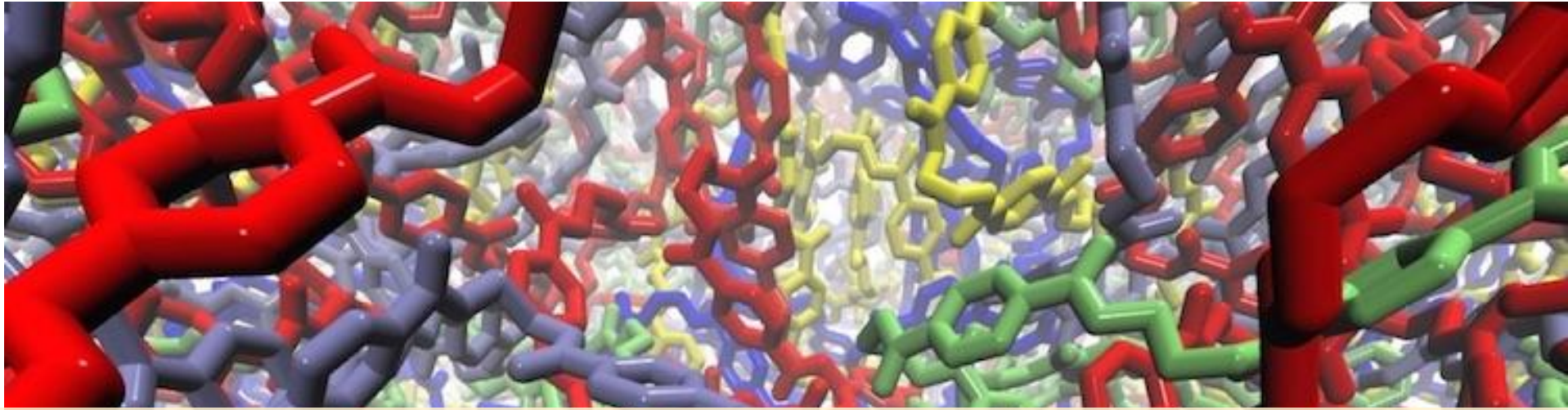
- Relevance to BETO and PABP
- AOP, Milestones, quarterly reporting
- Communication with other projects
- Tracking go/no-go activities
- Budget management.
- Leverage CCPC collaborations using all theory and modeling expertise across laboratories
- Go/no-go decisions to stop ineffective approaches, replace with new approaches that will deliver accurate predictions with high-throughput

Approach - Technical

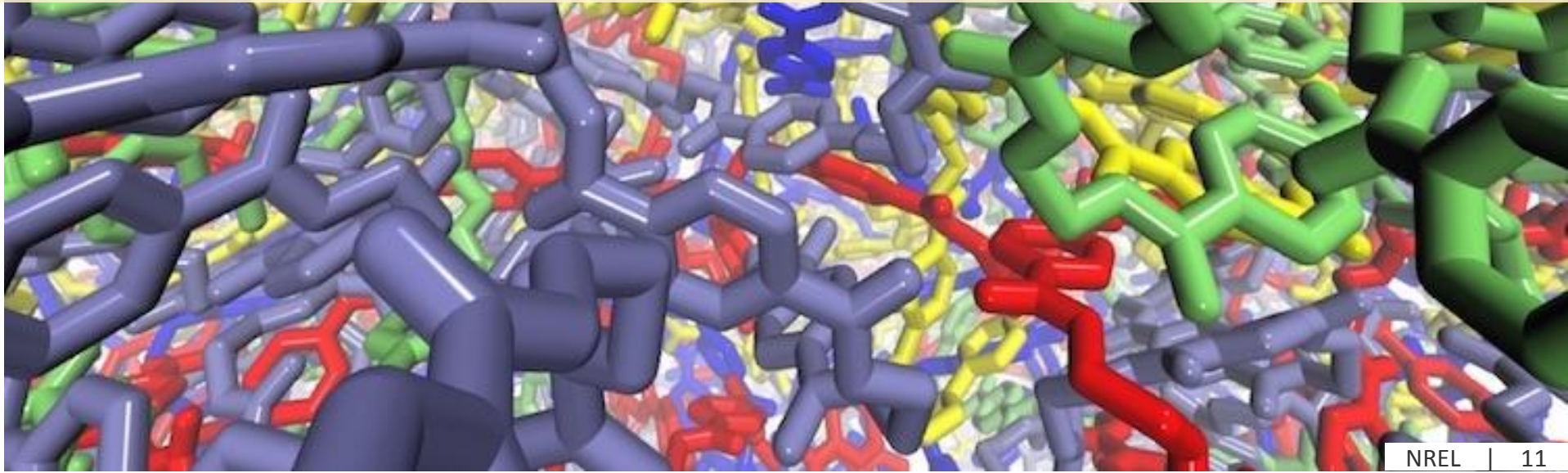
- **Approach: Machine-Learning** and **Molecular Modeling**
  - Build **Polymer Database** with properties from
    1. Literature
    2. Databases
    3. Industrial spec sheets
    4. BETO experimental projects
  - Advance **Neural network architectures** for polymer prediction
  - Design **high-throughput molecular simulation methods** for computing new polymer properties
- **Objective:**
  - **Predict NEW polymers** with performance advantages
  - Gain **insight**, discover approaches and solutions, understand molecular source of properties
  - **Guide and stimulate design, experiment, and engineering;** select most promising directions
  - **Increase research efficiency**
    - concentrate experimental efforts on the systems with highest likelihood of success
    - reducing search space for properties
    - not depending solely on intuition or accident
    - Reduce screening time by orders of magnitude

Approach - Technical

Critical Success Factors	Challenges	Strategies
Machine Learning can predict properties of many classes of polymers and deliver 10 new predicted polymers with performance advantaged properties	Databases are small	Continue literature search for polymers and properties
	Data for many properties is sparse	As above plus supplement missing properties with simulation derived properties
	Incomplete polymer classes represented	Build rules for missing polymer classes
	Neural network architectures for experimental data set scales are not well developed	Continue development and testing of ML approaches
Molecular simulation can augment the properties in the databases and can predict properties of experimentally-derived monomers with unknown properties	Complete design and testing of methods for numerical property determination is non-existent	Design and test for most accurate methods of property calculation, statistical mechanical approaches
	Available Force Fields are unproven, unknown structure for amorphous and many crystalline polymers	Implement high-throughput methods for lookup-type force fields.



Technical Accomplishments
Progress and Results



Machine Learning and Databases

State of the art before project start

- Polymer databases sparsely populated
- Databases incompatible with deep learning
- Polymer descriptors not appropriate for prediction of properties
- Databases had multiple, sometimes conflicting entries
- Neural networks needed to be developed for property prediction

Achievements for FY18

- Multiple incompatible databases merged and curated
- Database expanded from literature and experiment
- Chemical descriptors designed
- Automated rules for converting monomers to all possible polymers
- Transfer and message-passing methods developed to enhance small database sizes
- Full-scale Machine-Learning methods and software constructed and tested

Database Development

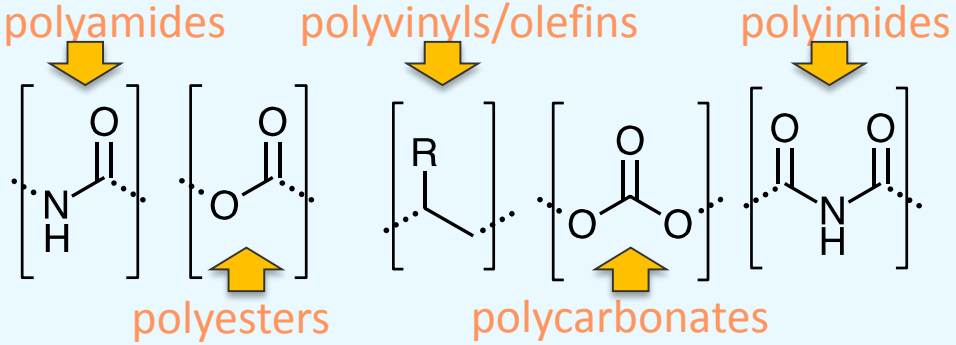
Parameter	No. of Polymers
Glass Transition	554
Density	227
Permeability: O_2 ,CO_2,N_2,H_2O	171, 126, 128, 36
Melt Temperature + Boolean	170
Modulus	77

Database:

Starting with no cohesive database, built and expanded a functional polymer database

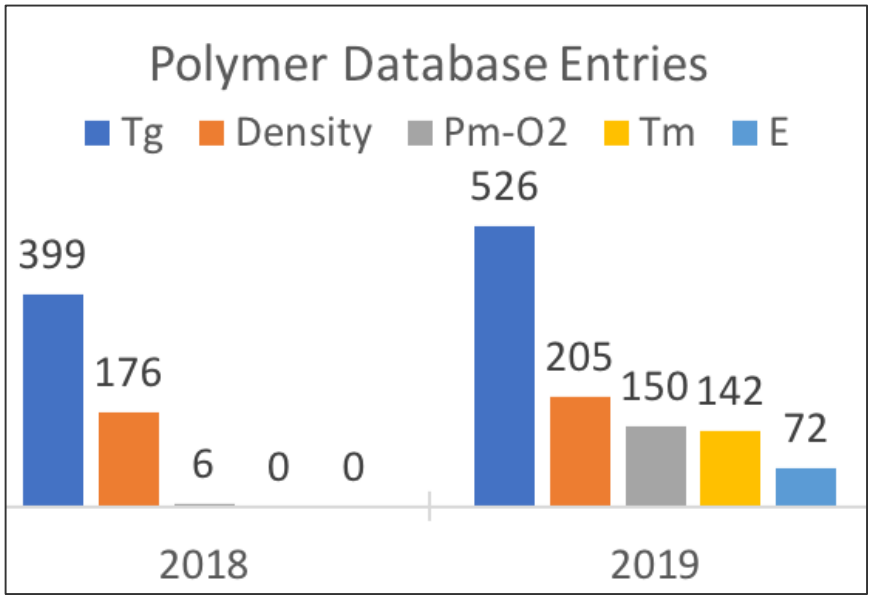
Continued expansion

Polymer Classes



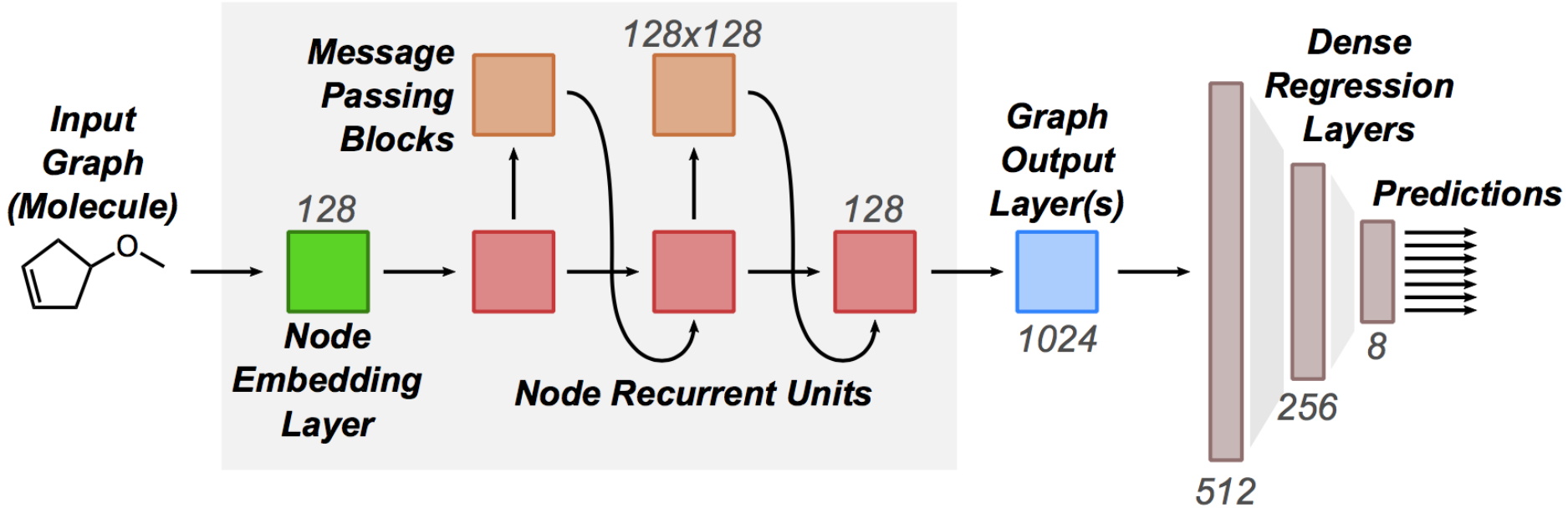
Five polymer classes are currently represented in the database, more to be added.

Defining connectivity of monomers in each class is complicated



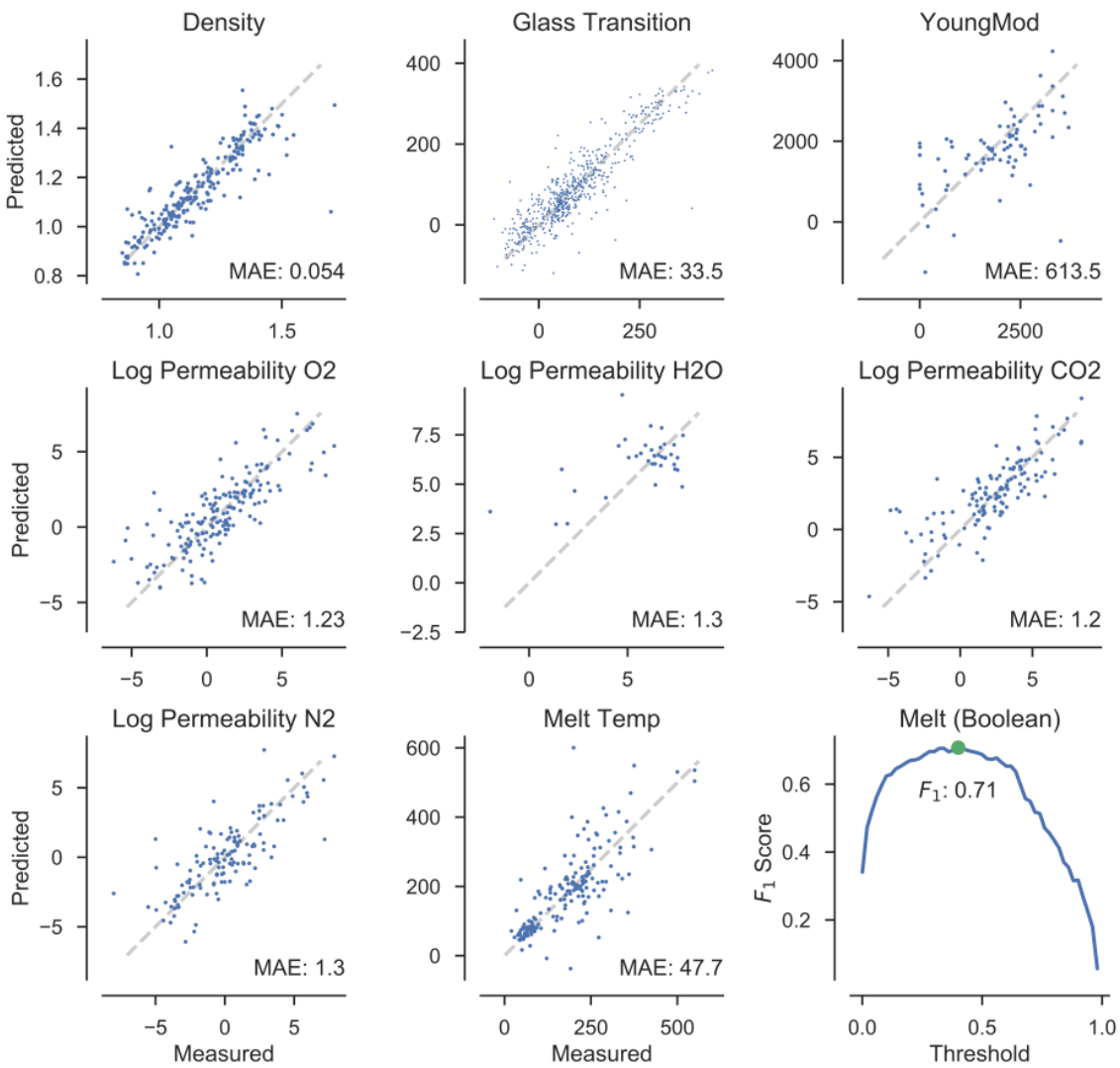
Machine Learning

- Developed machine learning model to simultaneously predict 8 polymer properties from monomer structures



Machine Learning

- Developed machine learning model to simultaneously predict 8 polymer properties from monomer structures
- Prediction performance is improved as more data is available



Prediction performance on held-out data during 10-fold cross-validation

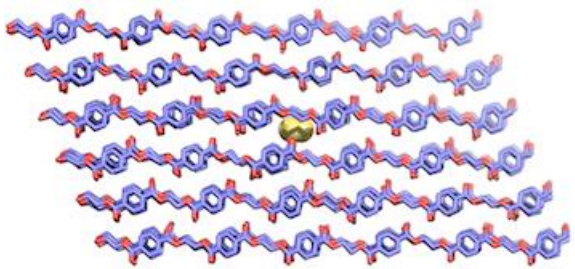
Molecular Simulation—Models

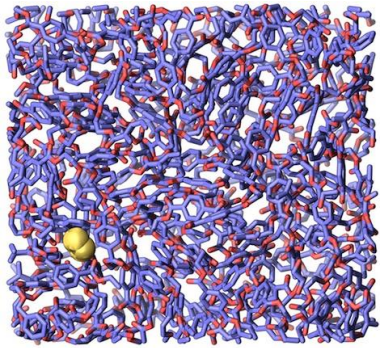
State of the art before this project:

- Force fields unproven, not easily transferable
- Amorphous polymer system setup ill-defined
- Weak connection to experiment

Achievements for FY18

- Force fields for PET, PEF, Nylons, and bio-inspired variants
- Structure builders (most polymers are semi-crystalline)
  - Crystalline (bulk, multiple faces)
  - Amorphous (three approaches: polymerize, random placement, melting)



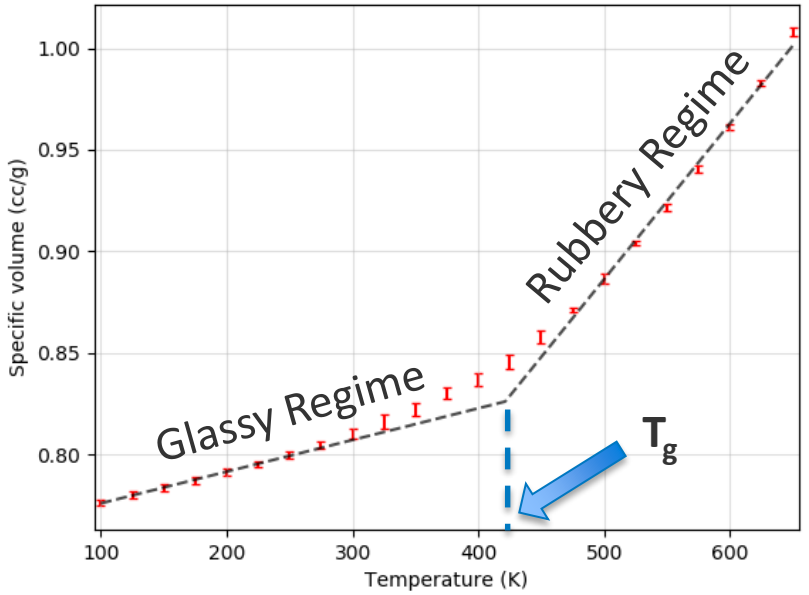


Molecular Simulation—Aromatic Polyesters

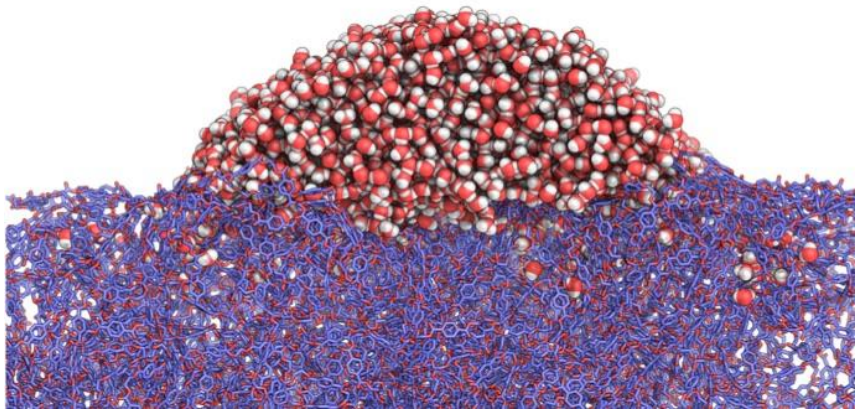
- Created model builder for amorphous and crystalline polymers
- Force Fields developed for PET, PEF
- Developed methods for calculating:

T_g	Density	Melting Point
Strength Moduli	Contact Angle	Permeability
Diffusion	Entanglement	Tortuosity

Glass Transition Temperature



Contact Angle (wetting capacity)

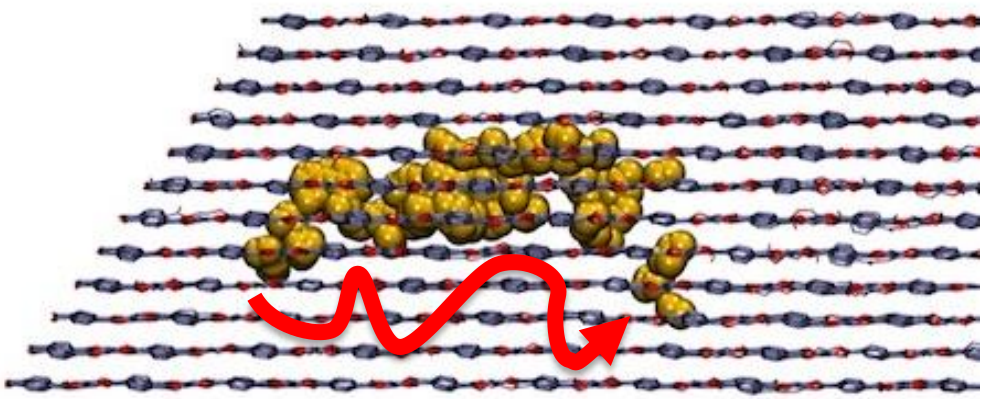


Molecular Simulation—Aromatic Polyesters

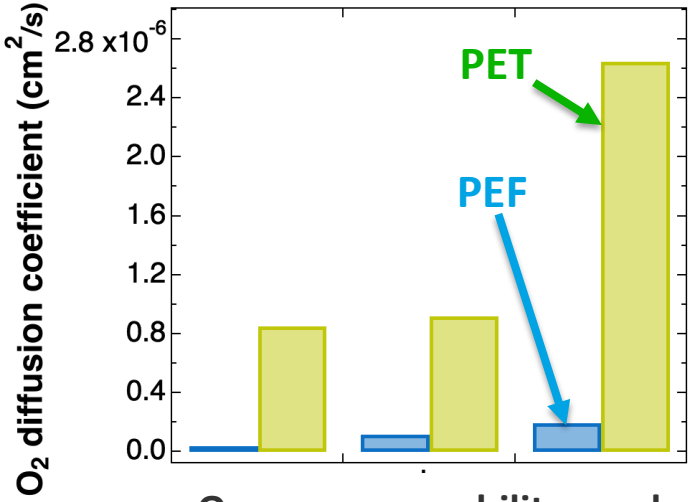
- Created model builder for amorphous and crystalline polymers
- Force Fields developed for PET, PEF
- Developed methods for calculating:

T_g	Density	Melting Point
Strength Moduli	Contact Angle	Permeability
Diffusion	Entanglement	Tortuosity

Gas Diffusion and Permeability



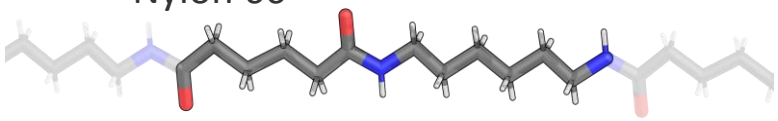
Path of diffusing oxygen in PET



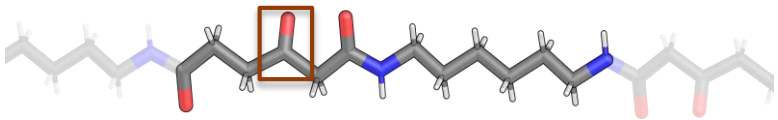
Oxygen permeability much higher in PET than PEF

Molecular Simulation—Nylons

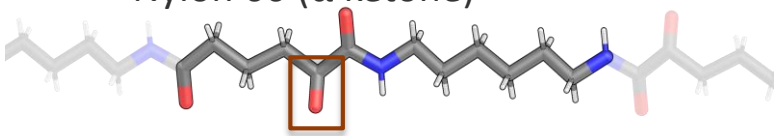
Nylon 66



Nylon 66 (β ketone)

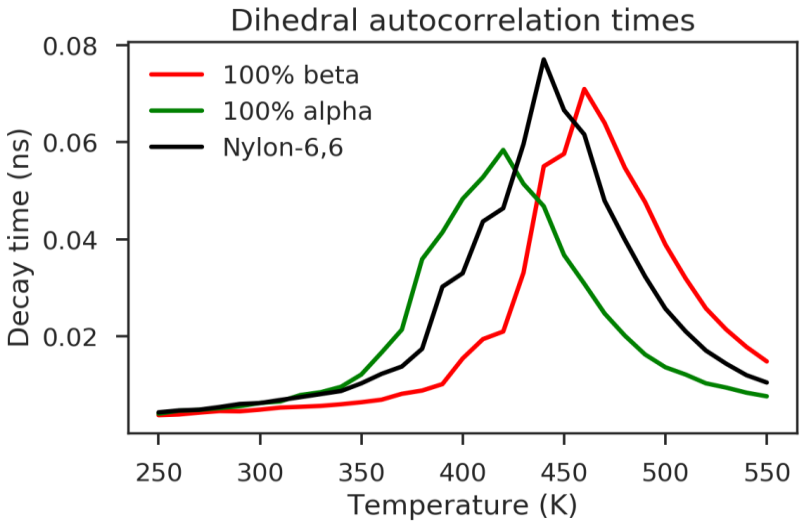


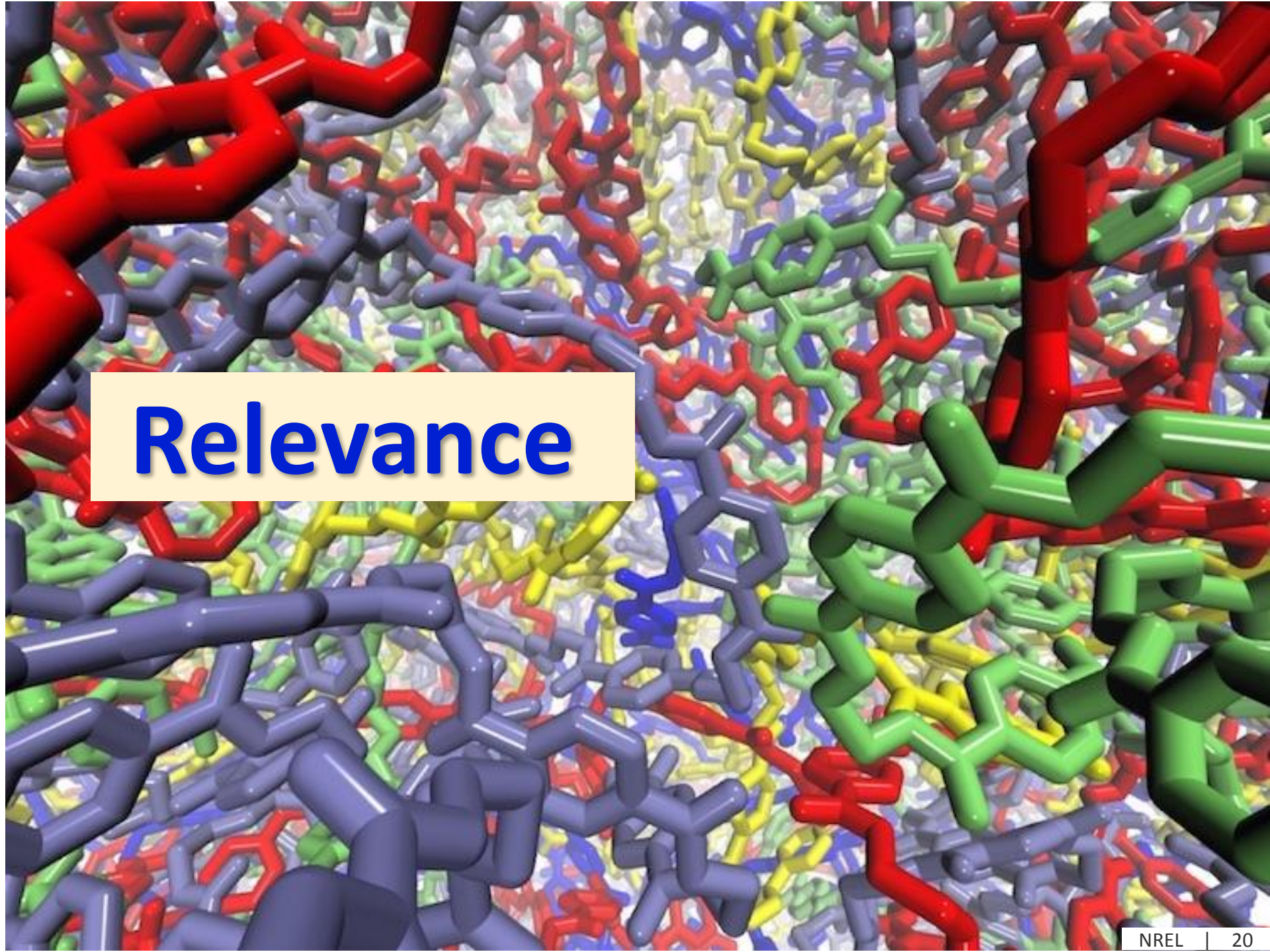
Nylon 66 (α ketone)



- Simulations reveal the enhanced thermal properties are due to increased rigidity in the polymer backbone that delays the glass transition

- Experiments indicate a significant increase in glass transition temperature of Nylon 66 when bio-derived β keto adipic acid is utilized rather than adipic acid.
- MD simulations demonstrate this increase, which is not achieved in α ketone case (consistent with experiment)




Relevance

Relevance

Contribution to Bioeconomy and Industry:

- Provide **Prediction Tool** to Industry
- Publicly accessible **Polymer Databases**
- Making available high-throughput **Property Calculator** to augment Machine Learning
- **Speed selection and design** through high-throughput screening
- Design and test new polymers with higher probability of getting to market
- Targets specific needs of manufacturing *(desired polymer properties)*

Within BETO mission:

- Research and develop high-performance biofuels, **bioproducts,** and biopower.
- Create **high-value performance-advantaged bioproducts** to enable more economic biorefineries.

Relevance

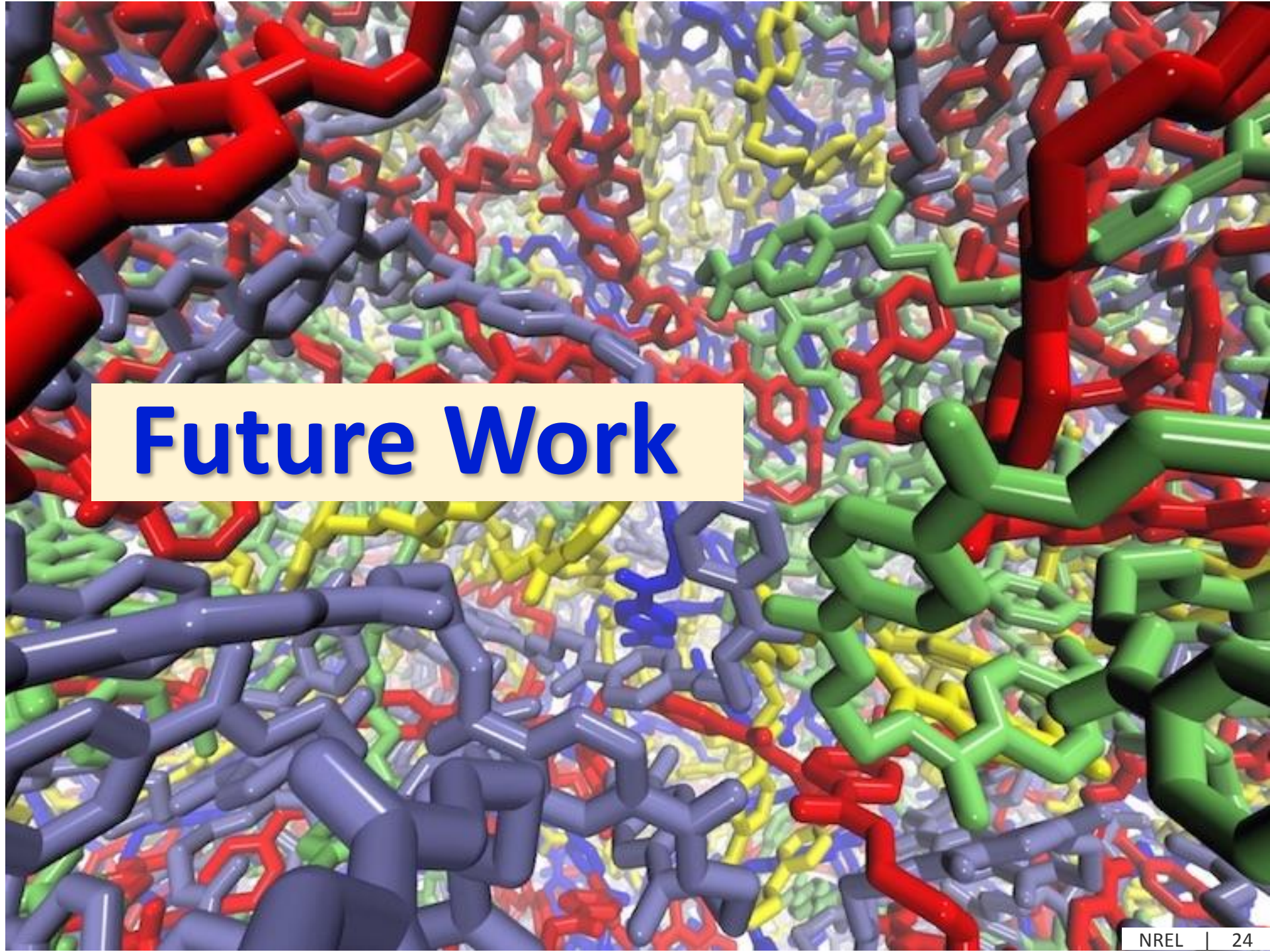
Stakeholder Outreach and Engagement:

- Provide suggested solutions to **inadequate properties** based on **structure-function relationships**
- Speed industrial development and deployment of new bio-based materials in the BioEconomy by:
  1. Providing polymer database and neural network for predictions
  2. Making available high-throughput property calculator to augment Machine Learning

Relevance

Scientific Discovery:

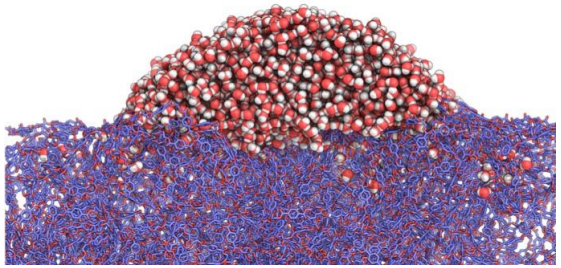
- **Highly Complementary with experiment:** Quickly down-selects candidates, predicts
- **Conceive and discover materials** with as yet unknown properties that are otherwise discovered through chance or extensive, time-consuming combinatorial approaches.
- **Elucidates structure-function relationships** for novel biobased polymers using high-throughput computational methods
- Discovers and designs biobased polymer materials with **unique performance attributes** *(not available from other sources)*
- Provide suggested solutions to **inadequate properties** based on **structure-function relationships**


Future Work

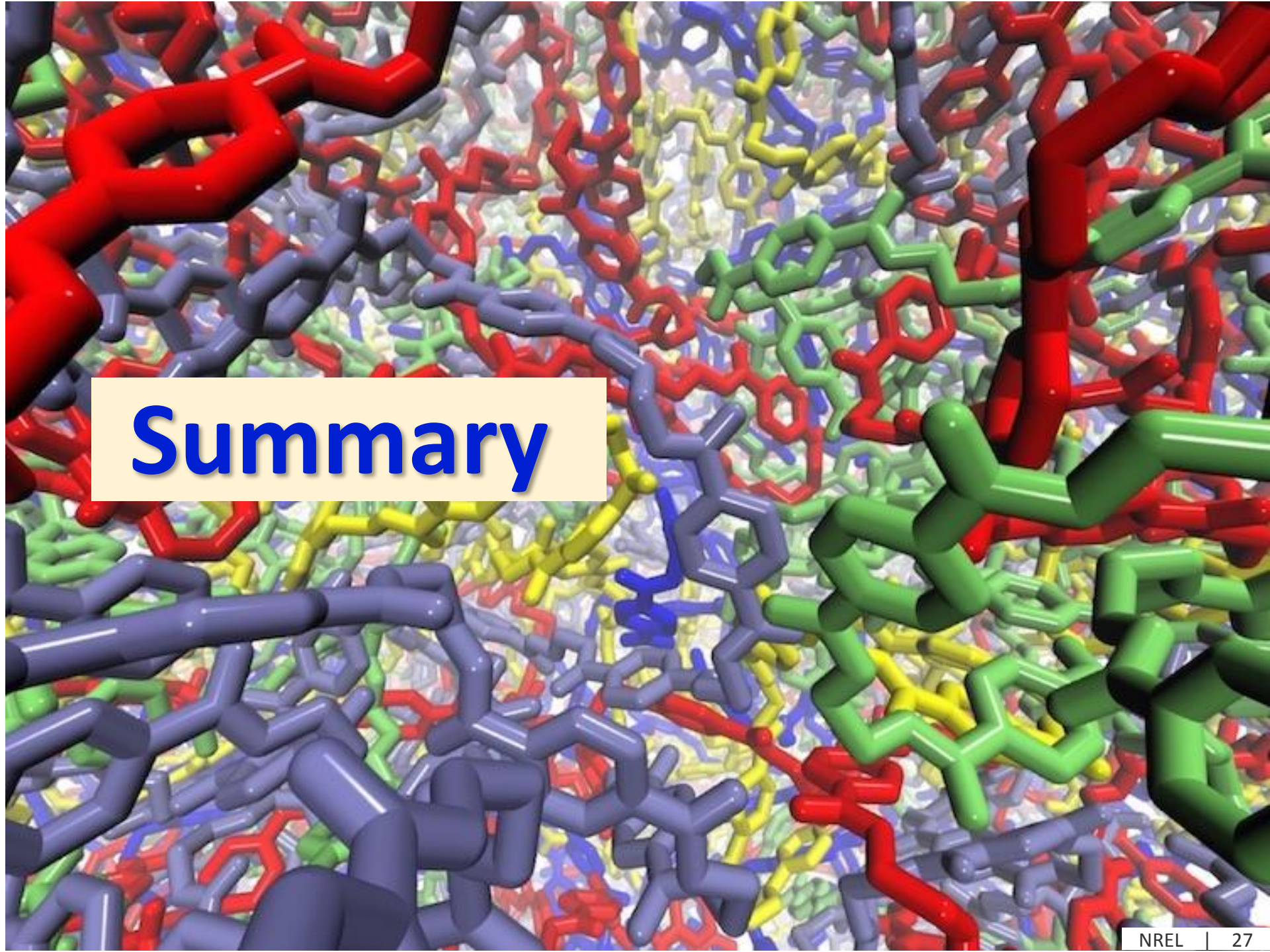
Future Work

- **Expand database** and rules to more classes of polymers
  - Increase **reliability and fidelity** of predictions
- Expand database for more properties and more entries for each property
- Search for bio-derived performance-advantaged products
  - Curate database of potential biologically-derived monomers from publicly available resources (i.e. Pubchem, MetaCyc, KEGG)
  - Polymerize monomers with polymer rules, screen for valid polymers
  - Predict polymer properties using machine learning models
- Add block copolymers and branching
- Provide predictions for PABP to be tested experimentally

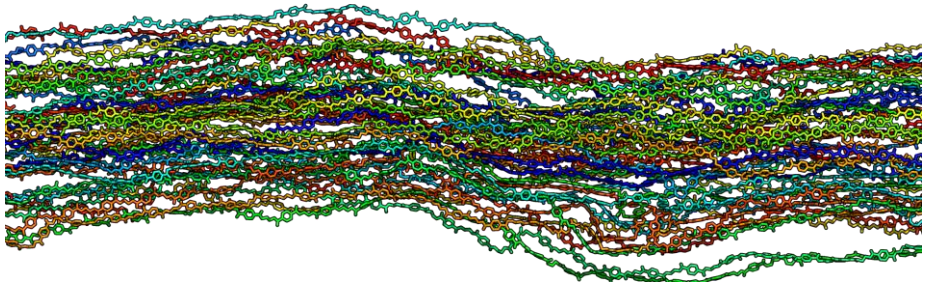
Future Work



- Implement High-Throughput Property Prediction
  - Fast and accurate Force Field determination
  - Refine structure prediction (amorphous and crystalline)
  - Automated building molecular systems, running simulations, analysis, and producing data for Task 1.
- *In-silico* testing of predictions from Task 1 for polymers with performance-advantaged properties
- Refine existing and add new property-calculation methods
- **Structure-function** relationships developed in conjunction with experimental tasks
- Add block copolymers and branching
- Deliver predictions for PABP to experiment for testing and make available to industry


Summary

Summary



Overview

Project designed to develop capability for

- Design of bio-based, performance-advantaged polymers
- Prediction of polymers with specific properties
- Prediction of properties of polymers from suggested monomers

Approach

Close collaboration within the PABP consortium to enhance and accelerate discovery, go from prediction to production, and from scientific understanding to design principles

Progress/Achievement

Achieved construction of Polymer Databases, Machine-Learning Prediction, Molecular Dynamics models, property calculation methods

Relevance

Relevant to the bioeconomy, speed of PABP to market, valorization of biofuels, needs of materials manufacturing, delivers both new bioproducts and tools for design and selection.

Future Work

Plans to expand databases, deliver real targets for testing, speed high-throughput screening and prediction, deliver user-friendly and accurate prediction tool

Machine Learning and Databases
Peter St. John
Nolan Wilson
Mark Nimlos

Molecular Simulation
Brandon Knott
Graham Schmidt

Experimental Collaboration
Nicholas Rorrer

Thank You

www.nrel.gov

Michael.Crowley@nrel.gov

NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.

Publications/Presentations

Presentations

Peter St. John, Caleb Phillips, Nolan Wilson, Mark R. Nimlos, Travis W Kemper and Ross E Larsen, *End-to-End Learning for Prediction of Optoelectronic Properties of Organic Photovoltaic Polymers* AIChE 2018

Papers

Peter C. St. John, Caleb Phillips, Travis W. Kemper, A. Nolan Wilson, Michael F. Crowley, Mark R. Nimlos, Ross E. Larsen *Message-passing neural networks for high-throughput polymer screening.* Submitted to *International Conference on Machine Learning* 2019

Response to Reviewers' Comments 2017

This is the first review of this project