DOE Bioenergy Technologies Office (BETO) 2023 Project Peer Review

Identifying Performance Advantaged Biobased Chemicals Utilizing Bioprivileged Molecules

4/7/23
Performance Advantaged Bioproducts

Brent Shanks
Iowa State University

Project Overview

- Biobased products can provide value-added, enhanced performance.
- Given the lack of structure/function relationships between chemical structure and end use performance, there is a need to develop a systematic strategy to identify performance advantaged molecules.

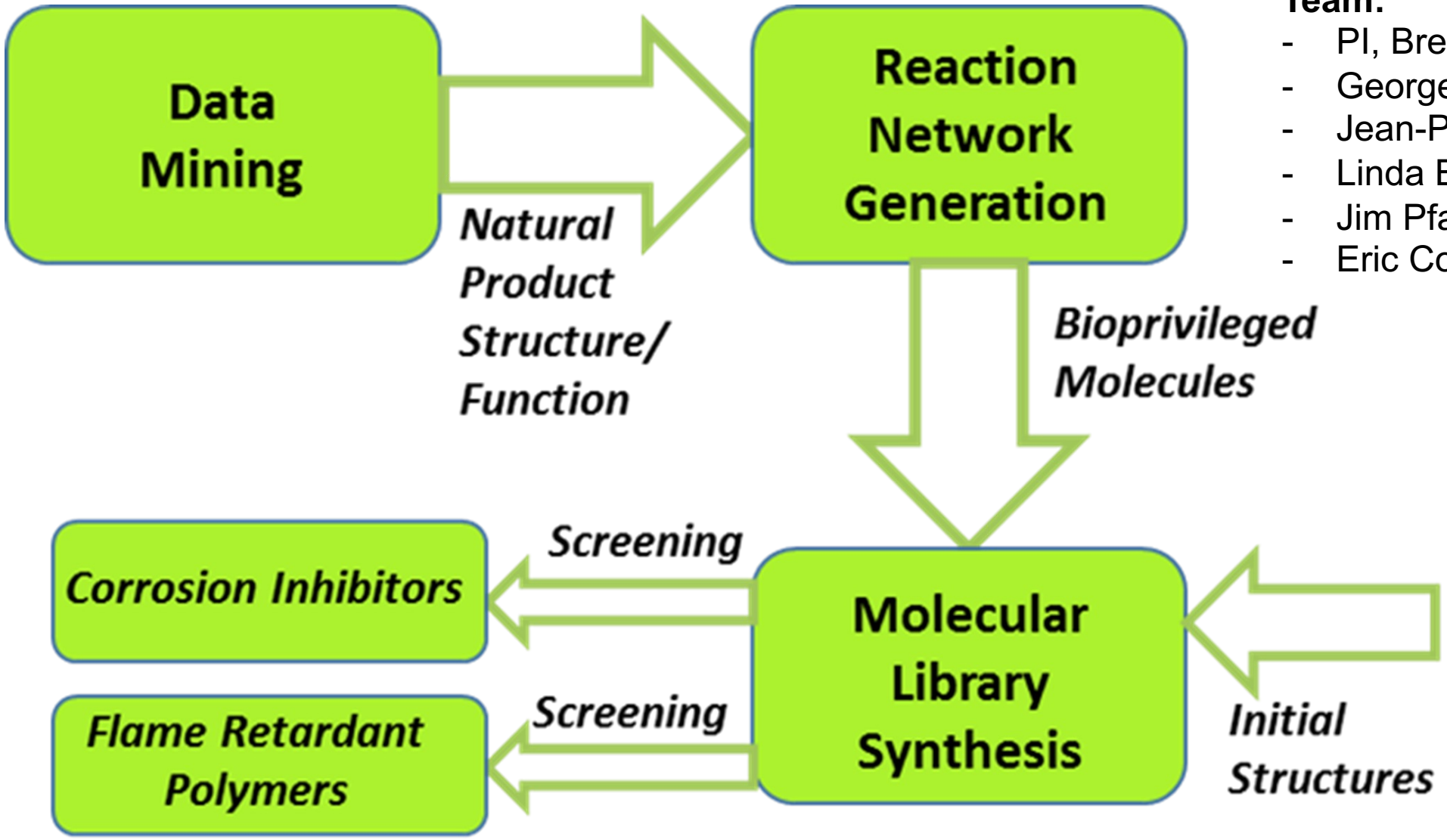
- Primary goals for our project:
  - *Synthesis and characterize ≥ 5 novel molecules with improved performance in the end use application of organic corrosion inhibitors and flame retardant nylon polymers.*
  - *Develop a systematic method for identifying biomass-derived molecules with improved performance in end use applications.*

- Key questions:
  - *Can the chemical literature be used to identify novel biobased chemicals with enhanced end use performance?*
  - *Can key bioderived chemical structures be identified that can be used as the basis for molecular library synthesis?*

1 - Approach

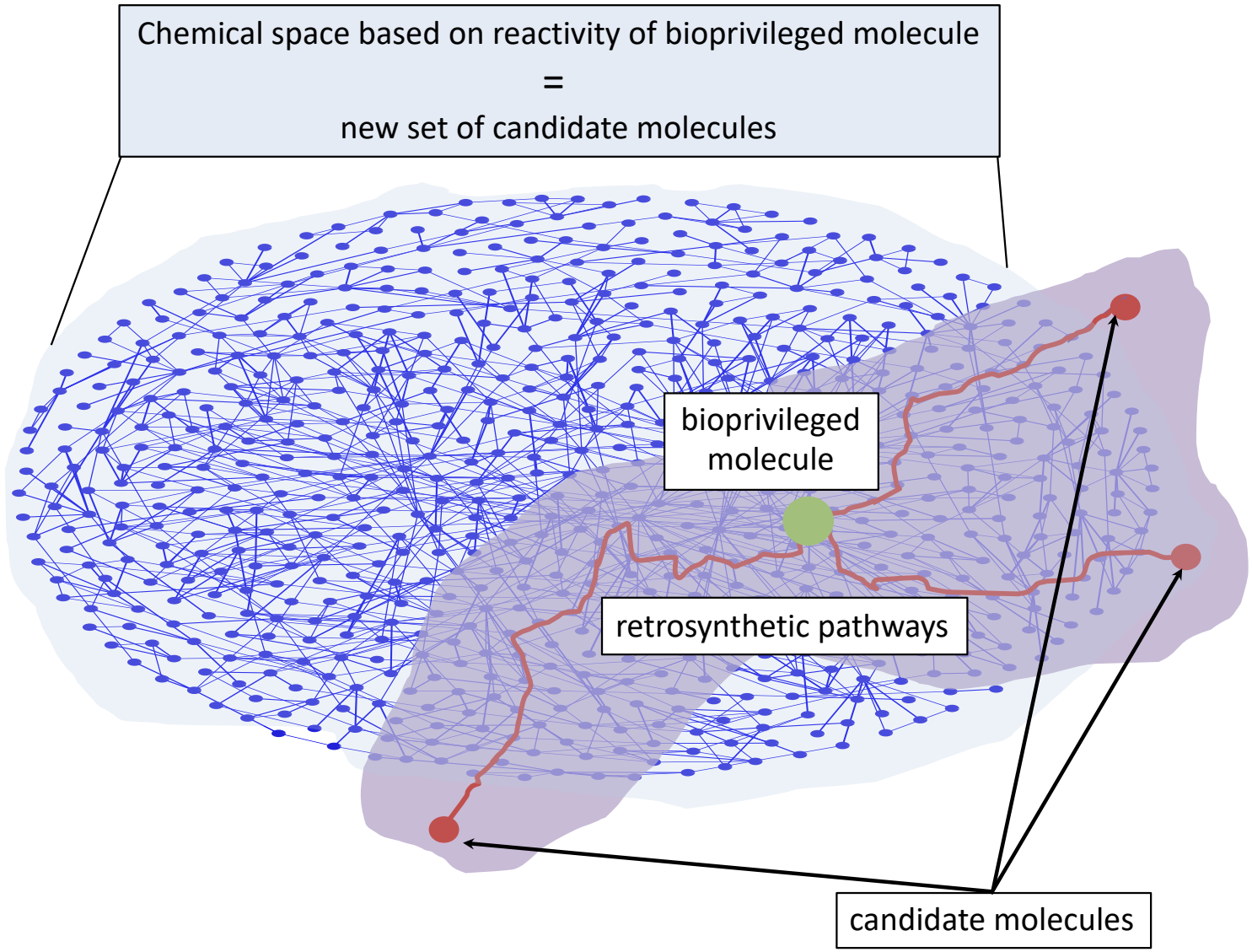
CBiRC
Center for Biorenewable Chemicals

Team:
- PI, Brent Shanks (ISU)
- George Kraus (ISU)
- Jean-Philippe Tessonnier (ISU)
- Linda Broadbelt (Northwestern)
- Jim Pfaendtner (UWashington)
- Eric Cochran (ISU)



Bioprivileged Molecules Strategy

CBiRC
Center for Biorenewable Chemicals



Library candidates

Measuring Success

Center for Biorenewable Chemicals

Key Milestones & Deliverables

Intermediate

- Initial set of at least 5 novel molecules identified
- First set of data mined candidate molecules found

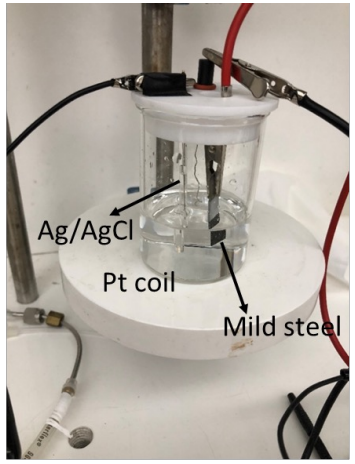
Final

- Data mined candidate molecules synthesized and tested - ≥5 novel molecules identified
- Integrated computational process developed for finding novel chemicals.

Performance Metrics

Center for Biorenewable Chemicals

Corrosion Inhibitors:

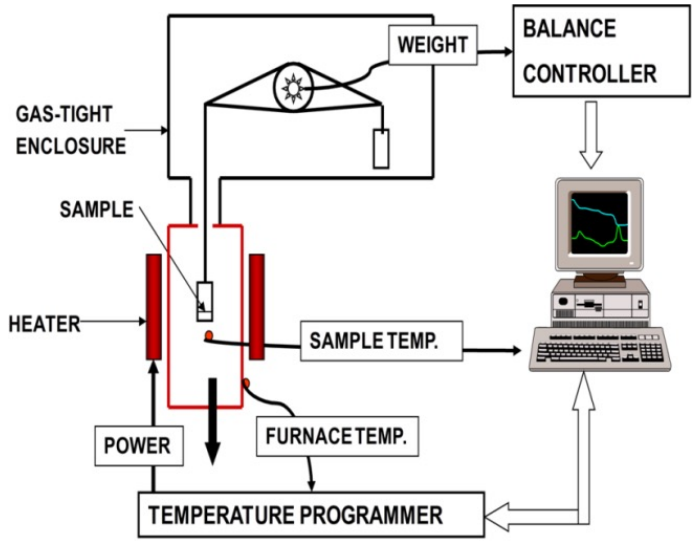


Nyquist Plot

Urotropin Base Cases

0.1 M H_2SO_4	0.3 M HCl
IE% = 24.0	**IE% = 57.0**

Flame Retardant Nylon:



Nylon 6,6 Base Cases

	M_n (kDa)	M_w (kDa)	Char (%)
UPA-0	1.3	2.2	3.7 ± 1.3%
Commercial	14.7	40.9	4.4 ± 0.8%

2 – Progress and Outcomes

CBiRC
Center for Biorenewable Chemicals

Organic corrosion inhibitors:

- *Synthesize and test 1st generation molecular library using triacetic acid lactone as the starting molecule – COMPLETE*
- *Synthesize and test 2nd generation molecular library using 4-hydroxycoumarin as the starting molecule – COMPLETE*

Flame retardant nylons:

- *Synthesize and test nylon 6,6 with 3-hexenedioic acid incorporation – COMPLETE*
- *Synthesize and test nylon 6,6 with tethered moiety monomer incorporation - COMPLETE*

Data mining:

- *Develop chemical data mining protocol with similarity scoring – COMPLETE*
- *Develop method to generate potential chemical structures - COMPLETE*

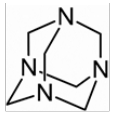
Bioprivileged molecule identification:

- *Using reaction operators, identify sets of C4-C7 candidate bioprivileged molecules - COMPLETE*
- *Curate bioprivileged molecule sets using metabolic pathways - COMPLETE*

Demonstrate integration of computation discovery with subsequent molecular library synthesis - ONGOING

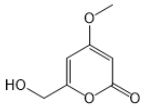
Corrosion Inhibition Efficiencies

Center for Biorenewable Chemicals

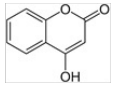
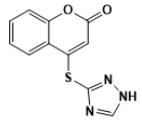
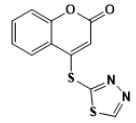
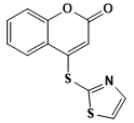
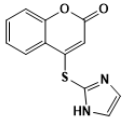
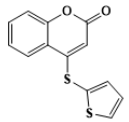
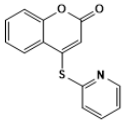
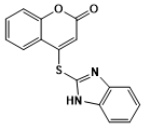
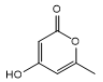
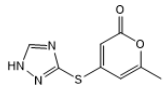
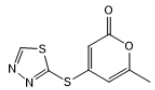
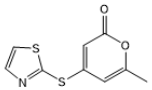
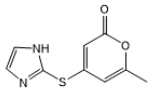
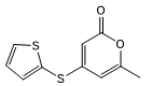
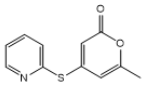
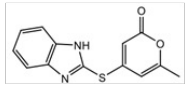
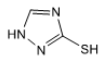
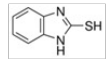


0.1M H_2SO
1 mM inhibitor

IE **24%**

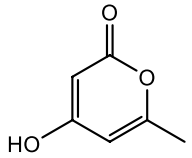


IE13%

								
Label	**4HC**	**C1**	**C2**	**C3**	**C4**	**C5**	**C7**	**C9**
Rct/Ohm	13	478	397	430	269	11	398	133
IE	-23%	97%	96%	96%	94%	-45%	96%	88%
								
Label	**TAL**	**T1**	**T2**	**T3**	**T4**	**T5**	**T7**	**T9**
Rct/Ohm	16	77	73	126	44	229	306	415
IE	0%	79%	78%	87%	64%	93%	95%	96%
								
Label	H_2SO_4	**1**	**2**	**3**	**4**	**5**	**7**	**9**
Rct/Ohm	16	18	19	99	71	NA	145	270
IE	NA	11%	16%	84%	77%	NA	89%	94%

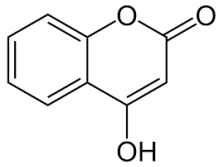
Biological Molecules

1st



TAL

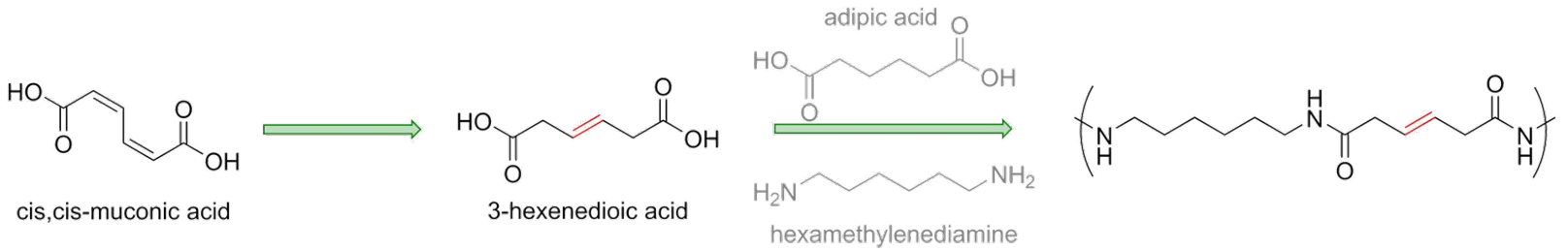
2nd



4HC

3HDA Incorporation in Nylon

UPA-x polymers with blends of x mol % HDA replacement of AA



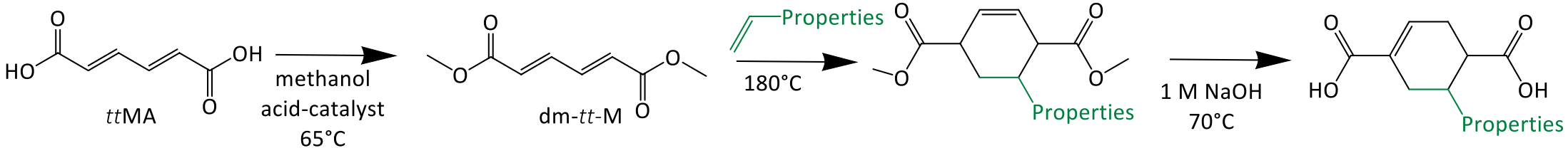
UPA-x polymers

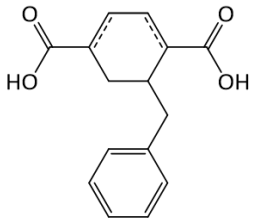
Sample	Char (%)
Commercial	4.4 ± 0.8
UPA-0	3.7 ± 1.3
UPA-5	6.7, 6.7
UPA-10	7.0, 8.3
UPA-20	10.0, 8.9
UPA-30	9.7, 10.5

Initial Functional Derivation

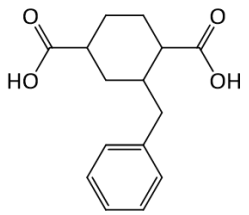
Center for Biorenewable Chemicals

Diels-Alder reaction for tethering



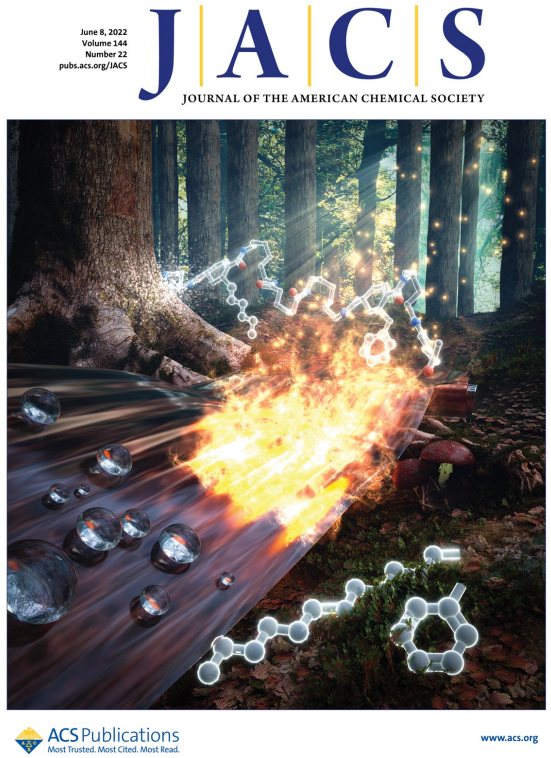


CH1DA-AB



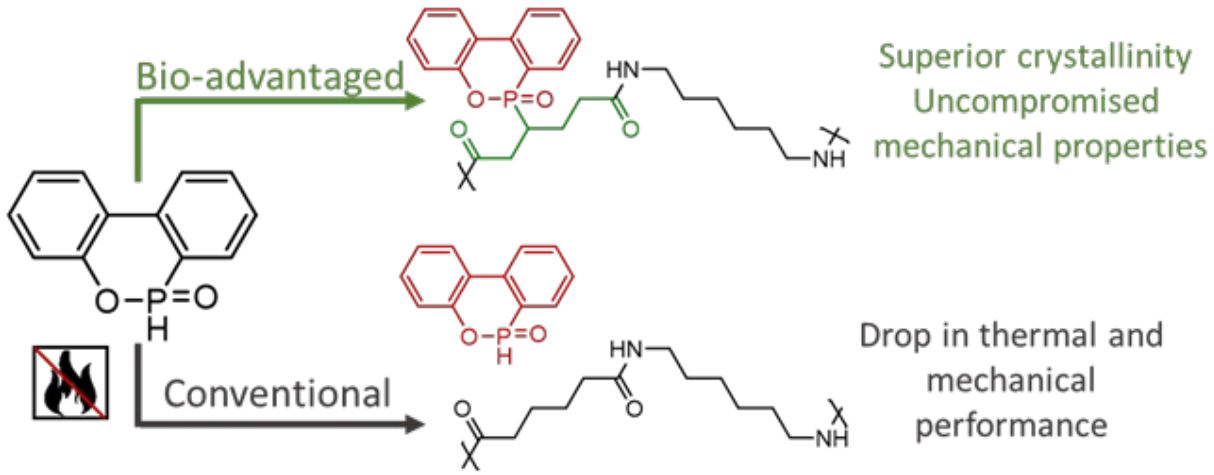
CHDA-AB

Sample	Mn (kDa)	Đ	T_m [°C]	χ_{C_DSC} [%]	$Char_{500°C}$ (%)
Nylon-6,6	34.3	2.0	262.2	26.3	2.9 ± 1.2
CH1DA-AB-25	25.3	2.3	230.7	14.2	3.9 ± 0.4
CHDA-AB-25	10.6	2.4	217.0	7.6	6.1 ± 0.6



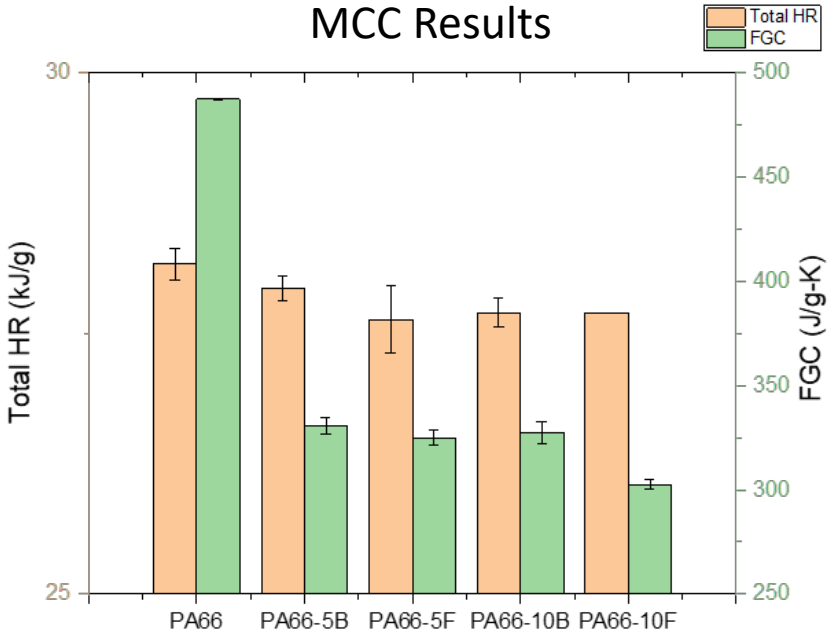
P. Carter et. al. **J. Amer. Chem. Soc.**, *2021,12, 8362-8372*

Michael Addition for 3HDA Functionalization



Sample	DSC Analysis T_m (°C)	DSC Analysis ΔH_m (J/g)	DSC Analysis χ_{c_DSC} (%)	WAXS[a] χ_{c_WAXS} (%)	TGA Analysis $Char_{500°C}$ (%)
PA66	260.5	66.4	26.0	50.8	3.04 ± 1.4
PA66-5B	248.8	50.8	19.9	35.0	4.05 ± 0.6
PA66-5F	247.9	49.5	19.4	38.8	4.04 ± 0.3
PA66-10B	239.7	38.2	15.0	31.8	3.96 ± 0.8
PA66-10F	242.9	46.8	18.3	38.2	4.57 ± 0.2

Samples	Expected DOPO Incorporation (%)	Actual DOPO Incorporation (%)[a]
PA66-5B	6.0	6.1
PA66-5F	6.0	6.2
PA66-10B	12.0	10.9
PA66-10F	12.0	13.7

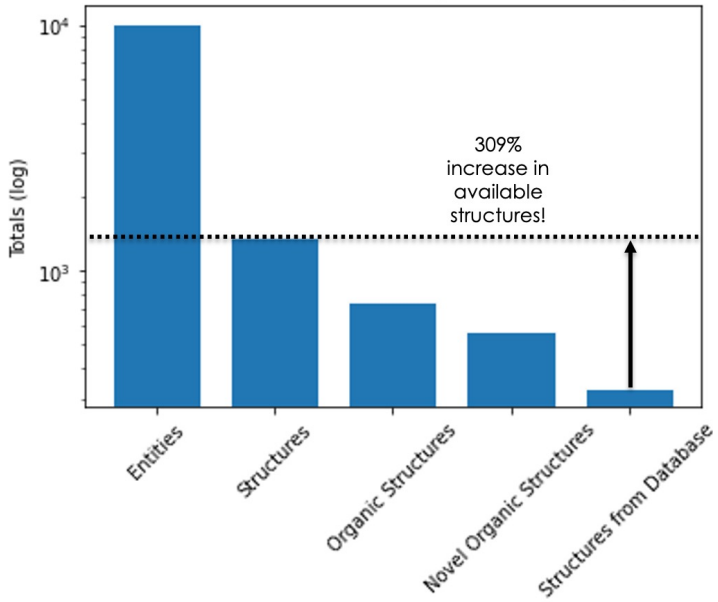


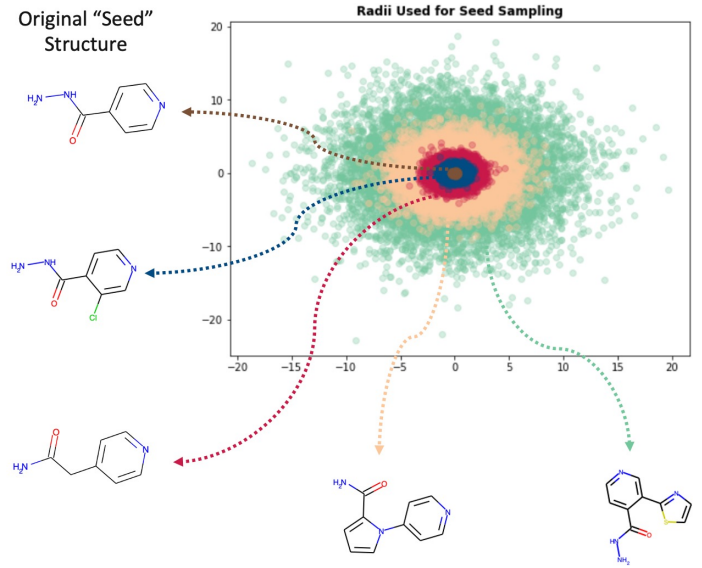
Flame growth capacity (FGC) value, which accounts for both charring and free radical formation inhibition during combustion

Natural Language Screening of Corrosion Inhibitors

CBiRC
Center for Biorenewable Chemicals

- Early in project period we established a large (millions of articles) full text corpus of scientific literature and trained a natural language model based on word2vec:
  - Identify molecules and properties → perform a primitive similarity analysis to ”anticorrosive” → curate new structures
- While the process increased available structures >300% over standard literature anticorrosion databases, it revealed flaws in the approach by which we screen molecules for *similarity*
- We developed an improved approach based on modern molecular representations to better identify “similar molecules” [1]



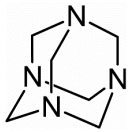
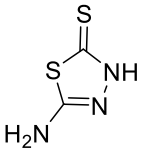
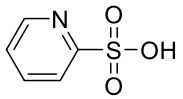
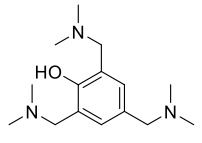
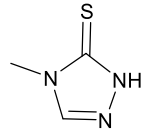


[1]O. Dollar et. al. **Chem. Sci**., *2021,12, 8362-8372 –* ***RSC Editor’s choice article***

Data Mining Molecules

- “Top” 100 molecules identified from data mining

- Selected screening of initial set of strictly literature extracted molecules

0.3M HCl
1 mM inhibitor

IE **57%**

IE 59%

IE 3%

IE 21%

IE 35%

- Further constraints required for candidate molecules

Advancing Ability to Use AI on Chemical Sciences Literature

CBiRC
Center for Biorenewable Chemicals

- One limitation of the first approach is the primitive nature of older language models
- Google scholar reveals this with a simple search for articles on perovskite degradation
- We are applying a concept called *topic modeling* to broadly improve the ability of natural language models to correctly identify molecules and materials within a research article

Google scholar query: "water induced degradation mechanism of ch3nh3pbi3"

Unraveling the Light-**Induced Degradation Mechanisms** of $CH_3NH_3PbI_3$ Perovskite Films
NH Nickel, F Lang, VV Brus… - Advanced Electronic …, 2017 - Wiley Online Library
… of **water** is a byproduct of the light-**induced degradation** process19 … act as catalysts for the **degradation mechanism** by promoting the … The reaction of the light-**induced degradation** in the …
☆ Save 99 Cite Cited by 98 Related articles All 2 versions

→ degradation mechanism, not water induced

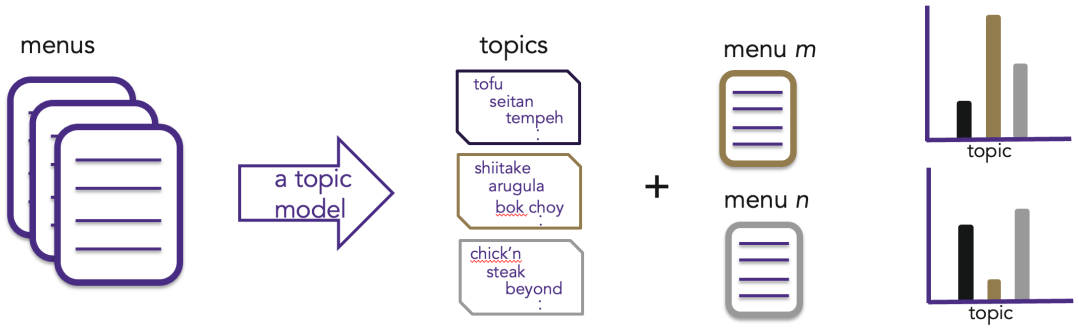
Nanoscale Studies at the Early Stage of **Water-Induced Degradation** of $CH_3NH_3PbI_3$ Perovskite Films Used for Photovoltaic Applications
J Llacer, D Moerman, O Douheret… - ACS Applied Nano …, 2020 - ACS Publications
… This work also aims to understand the early stages of **water-induced degradation** at the surface of those films. Through high-precision exposure to small amounts of **water** vapor, we …
☆ Save 99 Cite Cited by 3 Related articles All 6 versions

→ water induced degradation, but not mechanism

Unraveling the **Water Degradation Mechanism** of $CH_3NH_3PbI_3$
C Zheng, O Rubel - The Journal of Physical Chemistry C, 2019 - ACS Publications
… **degradation** is obviously important for the endurance. There are plenty of publications investigating the moisture-**induced degradation** … the liquid **water degradation mechanism** becomes …
☆ Save 99 Cite Cited by 32 Related articles All 4 versions

→ water induced degradation mechanism

Concept of topic modeling as applied to text searching of menus

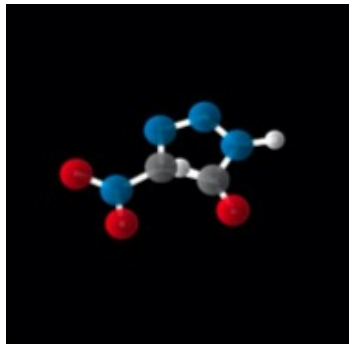


Goal-Directed Generation

Center for Biorenewable Chemicals

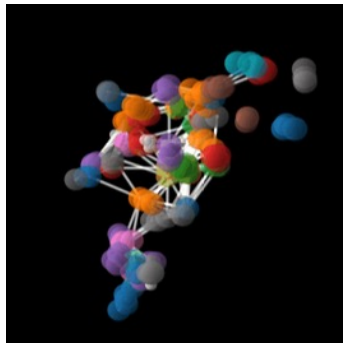
New algorithms are needed to generate candidate molecules that match desired properties. This gets vary hard for “real world” molecules…

State of the art 3D generative models can produce valid small molecules…

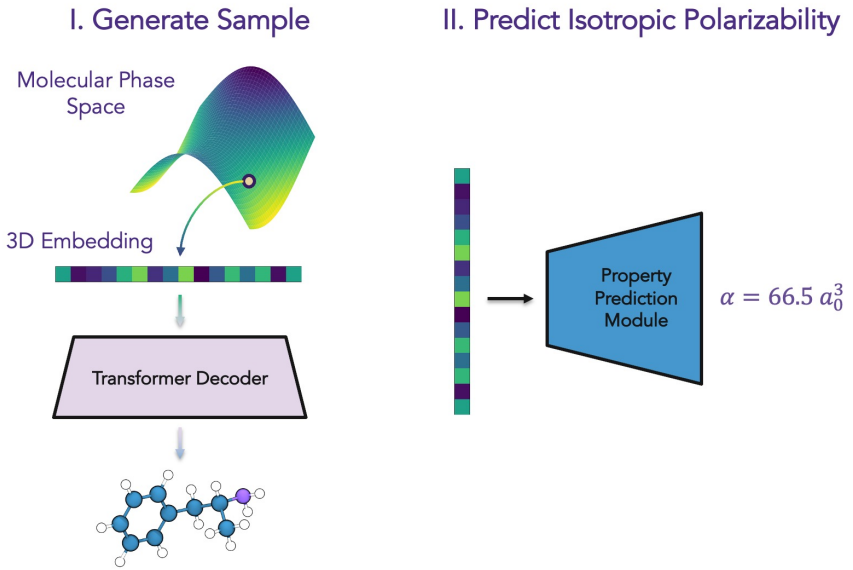


Hoogeboom et. al, 2022

… but often fail when tasked with generating larger drug-like molecules



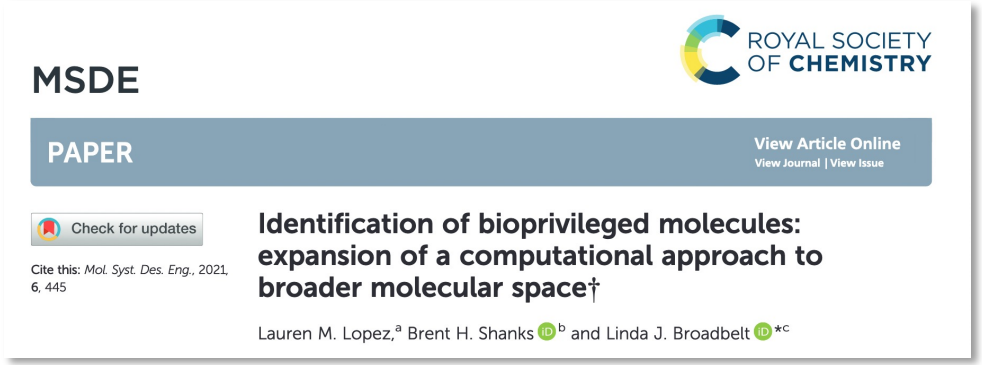
New approach improves 3D reconstruction of new molecules while simultaneously providing improved searching for desired properties



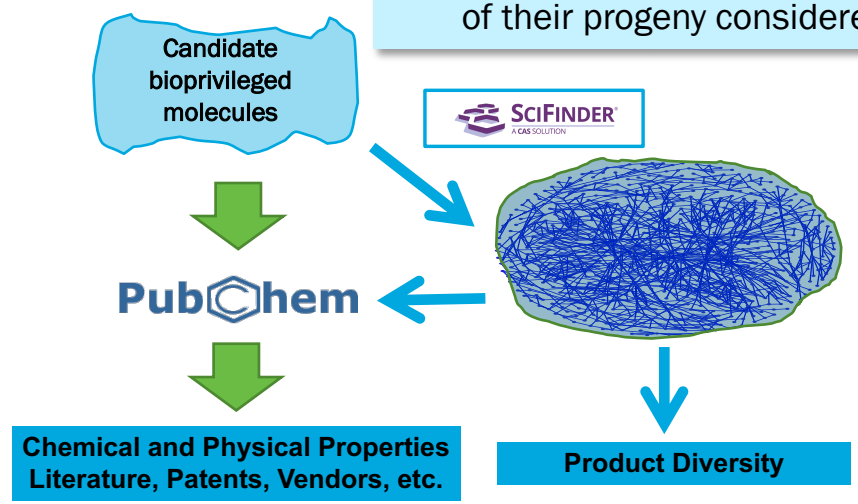
Not discussed here: corrosion inhibition was a very challenging topic for this approach due to the lack of computable physical properties that reliably correlate to inhibition efficiency

Identification of Bioprivileged Molecules: C4-C7

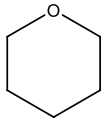
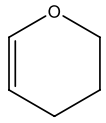
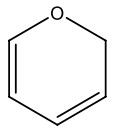
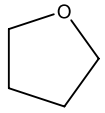
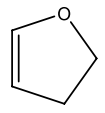
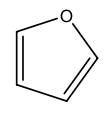
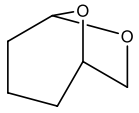
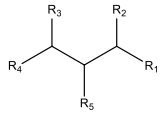
CBiRC
Center for Biorenewable Chemicals



Molecules' properties and diversity of their progeny considered

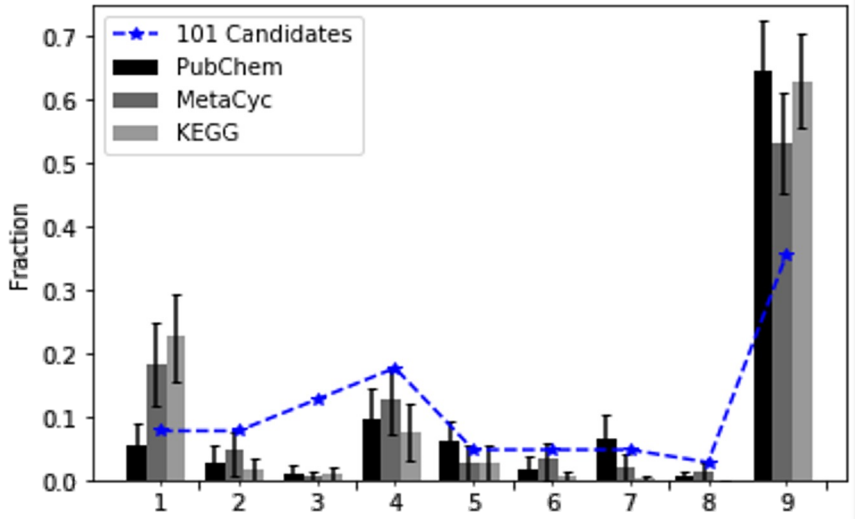


Molecules classified into motifs using machine learning

Motif 1	Motif 2	Motif 3
		
Motif 4	**Motif 5**	**Motif 6**
		
Motif 7	**Motif 8**	**Motif 9**
		

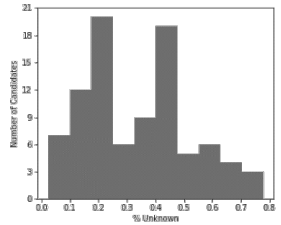
Molecules are more "biological" than chemical

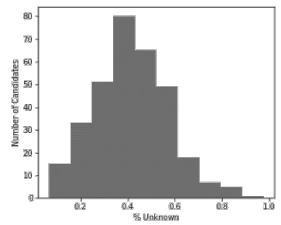
Comparison of Motif Distribution of New 101 Candidates to Three Databases



Framework discovers new molecules and pathways to them

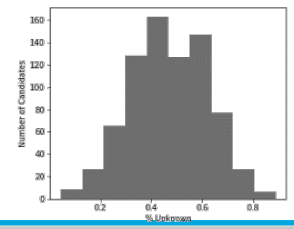
C4

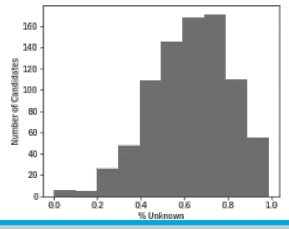




C5

C6



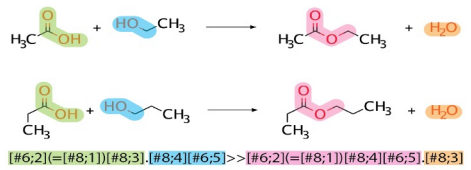


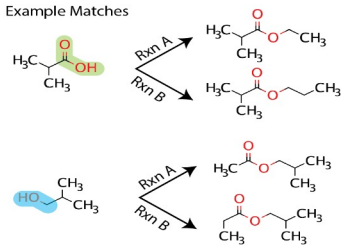
C7

Interrogating the "Bio" in Bioprivileged

CBiRC
Center for Biorenewable Chemicals

SMARTS operators encoding known biochemistry





1. Select Starting Organisms

2. Generate Network

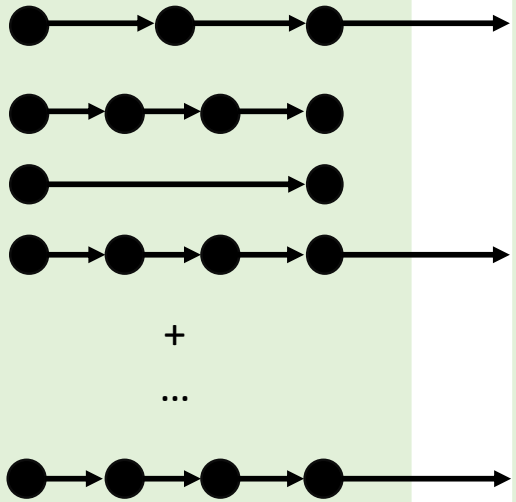
3. Identify Pathways

4. Analyze and Rank Candidates

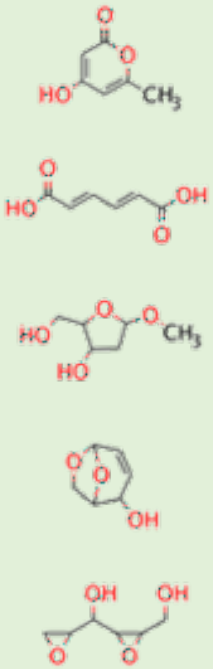
GEMs

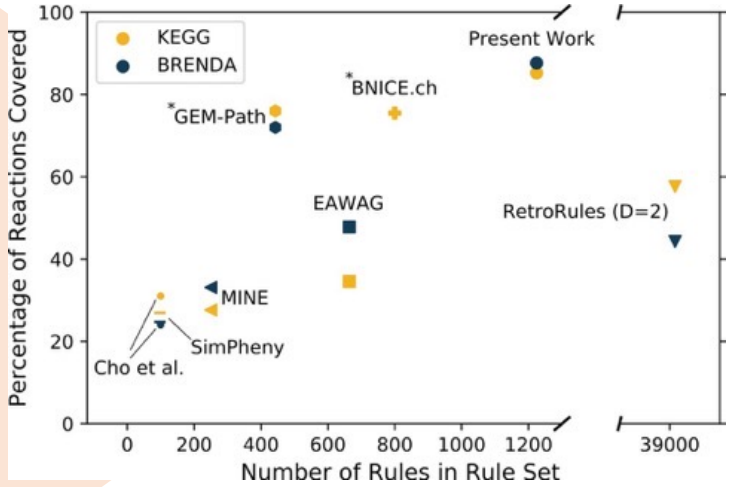
E. coli
S. cerevisiae
P. putida
C. ljungdahlii

Pickaxe
NetGen



Top 1k Zhou

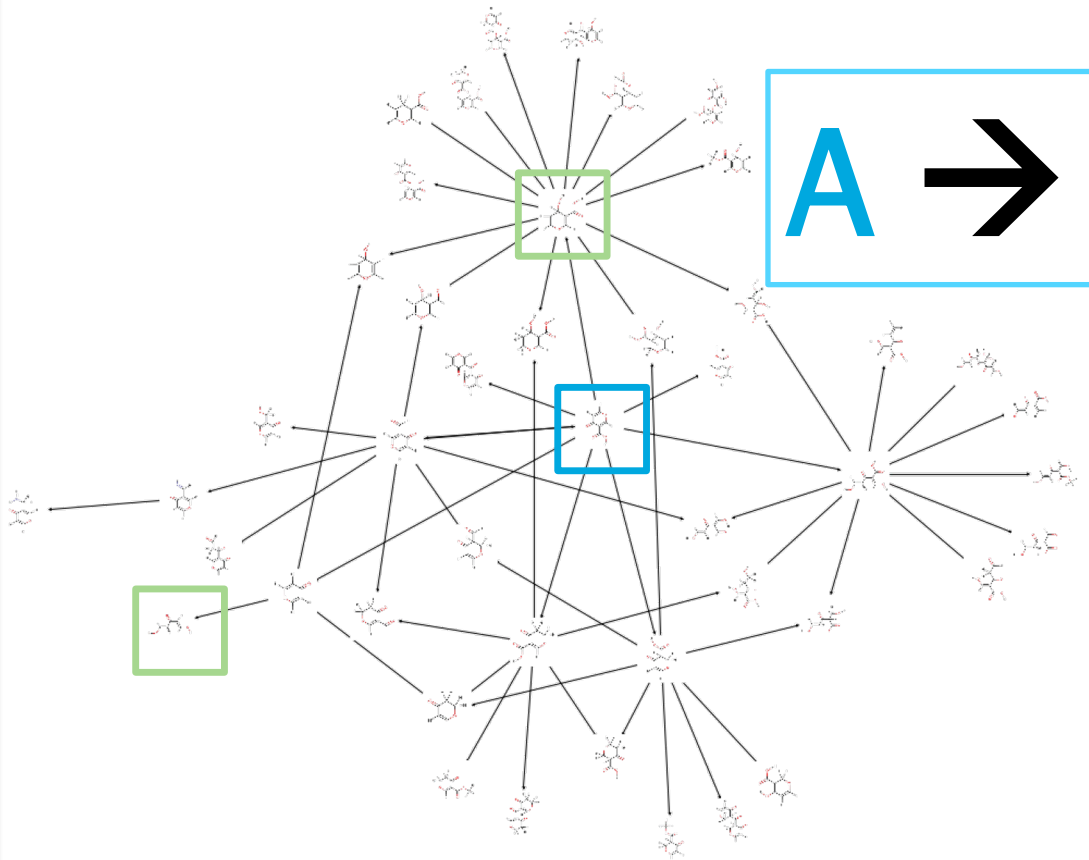




Available via GitHub

Connecting Bioprivileged Molecules to Corrosion Inhibitors

CBiRC
Center for Biorenewable Chemicals



- **Identify Starting Compound "A" (Bioprivileged Molecules)**
- **Identify Ending Compound "B" (Corrosion Inhibitors)**
- **Connect A to B (Reaction Pathways)**

46,777 molecules were generated using variation autoencoders (VAE) from a pool of 542 known corrosion inhibitors

Need a prioritization approach for pathway generation

Can we partition these molecules into classes to reduce complexity of pathway exploration?

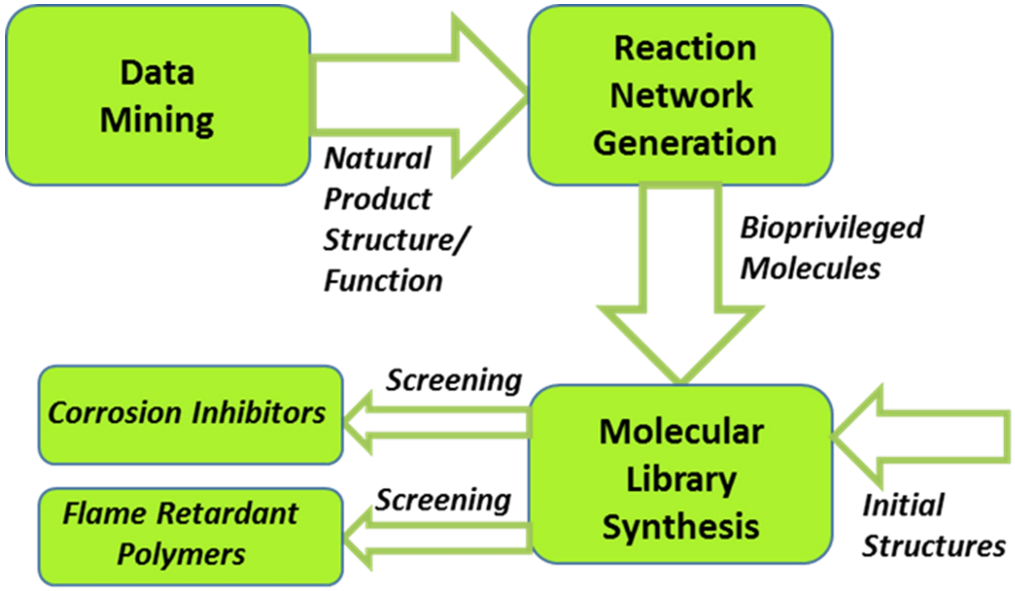
Generate centroids of each cluster → Convert centroids into pseudo-fingerprint → Identify defining features of cluster

3- Technology Impact

CBiRC
Center for Biorenewable Chemicals

- Technology areas impacted:
  - Novel molecules with improved performance in the end use application of organic corrosion inhibitors and flame retardant nylon polymers
  - Development of a systematic process for identifying biomass-derived molecules with improved performance in end use applications.

- Specific outputs to date:
  - ≥ 12 novel biobased chemicals were identified and screened to demonstrate at least 10% improvement in performance properties. *Extended testing is ongoing*.
  - *Publically disseminated computational framework for systematically identifying useful biobased chemicals*.



Discovery of performance advantaged molecules methodology

Summary

- Synthesized and validated novel performance advantaged biobased chemicals
  - *Performance advantaged organic corrosion inhibitors and flame retardant nylons found from original bioprivileged molecules (triacetic acid lactone and muconic acid)*
  - *Extended testing is being performed on several of the most promising molecules.*
- Literature data mining integrated with AI to generate candidate molecules
  - *Important advancements in accounting for the structure of chemical species*
  - *Ongoing work to constrain the developed chemical space*
- Further development of computational discovery of bioprivileged molecules
  - *Incorporation of a broader range of reaction operators*
  - *Linking chemical reaction network with biological reaction network*

- *Work ongoing to close the loop between computational framework and molecular library synthesis/testing*

Quad Chart Overview

Timeline

- *10/1/18*
- *5/30/23*

	FY22 Costed	Total Award
DOE Funding	*$2,053,500*	*$2,500,000*
Project Cost Share *	*$601,810*	*$625,000*

TRL at Project Start: 1
TRL at Project End: 2

Project Goal

Advance two technology areas; a) novel molecules with improved performance in the end use application of organic corrosion inhibitors and flame retardant nylon polymers and b) development of a systematic process for identifying biomass-derived molecules with improved performance in end use applications.

End of Project Milestone

- *Initial set of at least 5 novel molecules demonstrated*
- *Data mined molecules synthesized and tested – at least 5 novel molecules identified*
- *Integrated computational process developed for finding novel chemicals*

Funding Mechanism

FOA - Development of technologies able to contribute to the production of price-competitive biofuels and bioproducts: Performance Advantaged Bioproducts.

Project Partners

- Northwestern University
- University of Washington

Additional Slides

Responses to Previous Reviewers' Comments

- If your project has been peer reviewed previously, address 1-3 significant questions/criticisms from the previous reviewers' comments which you have since addressed
- Also provide highlights from any Go/No-Go Reviews

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

Patents and Publications

Patent Applications:

- Tessonnier, J.P., Hadel, J.E., Dell'Anna, M.N., Carter, P., Ganesbom, Shanks, B.H. and Cochran, E.W., "Base-Catalyzed Isomerization of Muconic Acid Derivatives," United States Utility Patent application, **63/080,959,** 2021.
- Tessonnier, J.P., Shanks, B.H., Cochran, E.W., Hadel, J.E., Dell'Anna, M.N. and Carter, P., "Isomerization of Cyclohexenedicarboxylic Acid and its Derivatives," United States Utility Patent application, **63/080,964,** 2021.
- Shanks, B.H., Kraus, G.A., Bradley, W., Huo, J. and Podolak, K., "Heteroaryl-Thio-Substituted Pyrones," United States Utility Patent application, **17/504,307**, 2021.
- Tessonnier, J.P., Cochran, E.W., Shanks, B.H., Gansebom, D., Abolmohammdi, S. and Forrester, M.J., "Polyesters and Polyamides and Their Preparation Through In Situ Hydration of Trans-3-Hexenedioic Acid," United States Patent application, 2021

Publications:

- Huo, J., Bradley, W., Podolak, K., Ryan, B., Roling, L.K., Kraus, G.A., Shanks, B.H., "Triacetic Acid Lactone and 4-Hydroxycoumarin as Bioprivileged Molecules for the Development of Performance Advantaged Organic Corrosion Inhibitors" *ACS Sustain. Chem. Eng*., 2022, **10**, 11544-11554. DOI: 10.1021/acssuschemeng.2c02940
- Carter, P., Trettin, J.L., Lee, T.-H., Forrester, M.J., Shanks, B.H., Tessonnier, J.-P., Cochran, E.W., "A Bio-Enabled Platform to Access Polyamides with Built-In Target Properties," *J. Am. Chem. Soc*., 2022, **144**, 9548-9553. DOI: 10.1021/jacs.2c01397
- Abdolmohammadi, S., Gansebom, D., Goyal, S., Lee, T.-H., Kuehl, B., Forrester, M., Lin, F.-Y., Hernandez, N., Shanks, B., Tessonnier, J.-P., Cochran, E.," Analysis of the Amorphous and Interphase Influence of Comononomer Loading on Polymer Properties towards Forwarding Bioadvantaged Copolyamides," *Macromolecules*, 2021, **54**, 7910-7924. DOI: 10.1021/acs.macromol.1c00651
- Dollar, O., Joshi, N., Beck, D.A.C., Pfaendtner, J. "Attention-based generative models for de novo molecular design," *Chemical Science*, 2021, **12**, 8362-8372. DOI: 10.1039/D1SC01050F (Edge Article)
- Dollar, O., Joshi, N., Beck, D.A.C., Pfaendtner, J. "Giving Attention to Generative VAE Models for De Novo Molecular Design," *Theoretical and Computational Chemistry*, 2021. DOI: 10.26434/chemrxiv.13724629.v1
- Lopez, L.M., Shanks, B.H., Broadbelt, L.J., "Identification of bioprivileged molecules: expansion of a computational approach to broader molecular space," *Molecular Systems Design & Engineering*, 2021, **6,** 445-460 – selected by the Editors as a HOT MSDE article. DOI: 10.1039/D1ME00013F
- Ashraf, C., Joshi, N., Beck, D.A.C., Pfaendtner, J. 'Data Science in Chemical Engineering: Applications to Molecular Science', *Annual Review of Chemical and Biomolecular Engineering,* 2021, **12,** 15-37. DOI: 10.1146/annurev-chembioeng-101220-102232
- Carraher, J.M., Carter, P., Rao, R.G., Forrester, M.J., Pfennig, T., Shanks, B.H., Cochran, E.W., Tessonnier, J.P. "Solvent-driven isomerization of cis,cis-muconic acid for the production of specialty and performance-advantaged cyclic biobased monomers," *Green Chemistry*, 2020, **22**, 6444-6454. DOI: 10.1039/D0GC02108C

Presentations

- "Developing Robust Strategies for Advancing the Next Generation Biobased Chemical Products," Shanks, B.H., Braskem, virtual seminar, September, 2022
- "Promising Biobased Organic Corrosion Inhibitor Libraries Derived from Pyrones," Kraus, G.A., Shanks, B.H., Huo, J., Podolak, K., AMPP Annual Conference + Expo 2022, San Antonio, TX, March, 2022.
- "A bio-enabled platform to access polyamides with built-in target properties," with Carter, P., Trettin, J., Shanks, B.H., Tessonnier, J.-P., Cochran, E.W., American Chemical Society Spring Meeting, San Diego, CA, March, 2022.
- "Isomerization and Michael Addition Chemistry for the Production of Bioenabled, Flame Retardant Nylon 6,6 Monomers," Hadel, J., Shanks, B.H., Cochran, E.W., Tessonnier, J.-P., American Chemical Society Spring Meeting, San Diego, CA, March, 2022.
- "Discovery of Novel Compounds and Pathways through Identification of Bioprivileged Molecules," Lopez, L.M., Shanks, B.H., Broadbelt, L.J., Annual Meeting, American Institute of Chemical Engineers, virtual, November, 2021.
- "Novel biobased cyclic diacids for polyamides with tailored functional properties," Carter, P., Trettin, J., Shanks, B.H., Cochran, E.W., Tessonnier, J.-P., American Chemical Society Spring Meeting, virtual, April, 2021.
- "Developing performance-advantaged organic corrosion inhibitors utilizing bioprivileged molecules," Huo, J., Bradley, W., Podolak, K., Kraus, G., Shanks, B.H., 43rd Symposium on Biomaterials, Fuels and Chemicals, virtual, April, 2021.
- "Developing Performance-Advantaged Chemicals Utilizing Bioprivileged Molecules: Organic Corrosion Inhibitors" Huo, J., Bradley, W., Podolak, K., Ryan, B., Roling, L.K., Kraus, G.A., Shanks, B.H., Annual Meeting, American Institute of Chemical Engineers, virtual, November, 2020.
- "Synthesizing Novel Biobased Molecules for Enhanced Polymer Properties," Trettin, J., Carter, P., Shanks, B.H., Tessonnier, J.-P., Cochran, E.W., Annual Meeting, American Institute of Chemical Engineers, virtual, November, 2020.
- "Developing Novel Chemicals through Bioprivileged Molecules," Shanks, B.H., Center for Biocatalysis and Bioprocessing Annual Meeting, Iowa City, IA, October , 2019.
- "Identifying Performance Advantaged Biobased Chemicals Utilizing Bioprivileged Molecules," Shanks, B.H., Huo, J., Kraus, G.A., Tessonnier, J.-P., Cochran, E., Broadbelt, L., Pfaedtner, J., Beck, D., Biomass to Biobased Chemicals and Materials Gordon Conference, Newry, ME, July, 2019.