HELPING CHEMICAL INDUSTRY IN ITS GREEN AND DIGITAL TRANSITION

NEXTMOL

Dr. Monica de Mier, CEO

 | 28.09.2023

Jornada Innovaplásticos 2023

About us

NEXTMOL

We are an awarded startup company specialized in **computational chemistry**.

We offer digital solutions to **accelerate the R&D** of best-in-class and more sustainable **specialty chemicals** based on **molecular modeling** and **artificial intelligence**.

Spin-off company of:

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

With the support of:

REPSOL Fundación

aws

NVIDIA INCEPTION PROGRAM

Funding and awards:

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

ACCIÓ
Generalitat de Catalunya

CDTI
INICIATIVA neotec

e-dci
European Extreme Data & Computing Initiative

enisa

European Innovation Council

SEAL OF EXCELLENCE
European Commission

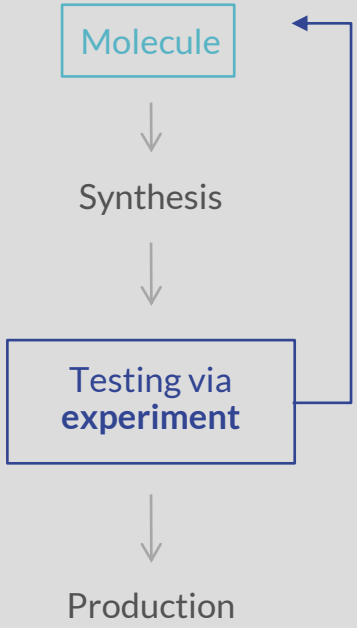
SOUTH SUMMIT

PREMIOS EMPRENDEDOR XXI

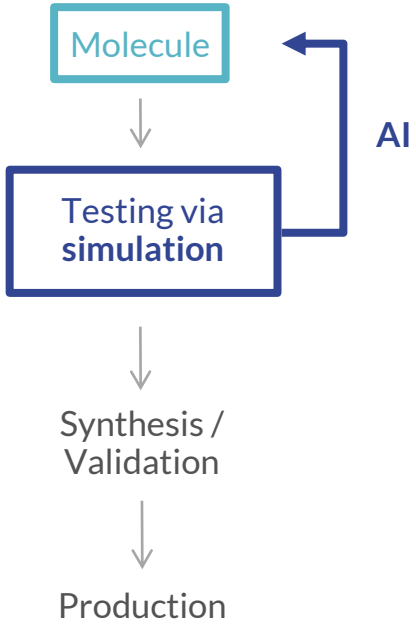
ISO 27001
ISOQAR REGISTERED
alcumus.com

NEXTMOL

Traditional Lab



Lab with Nextmol

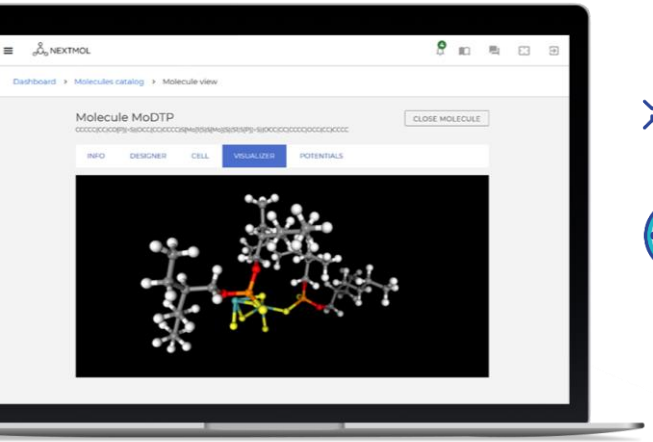


Computational lab for chemical R&D

- Molecules characterized directly on the computer
- Automated idea-testing loop
- Only the most promising candidates go to the lab
- Guides and complements the experimental work

Computational and green lab for chemical R&D

- Reduced synthesis and waste
- Rapid replacement of raw materials, ingredients or formulations by green alternatives (biobased, biodegradable, etc.)
- Targeted functionalities and properties



Polymers

Surfactants

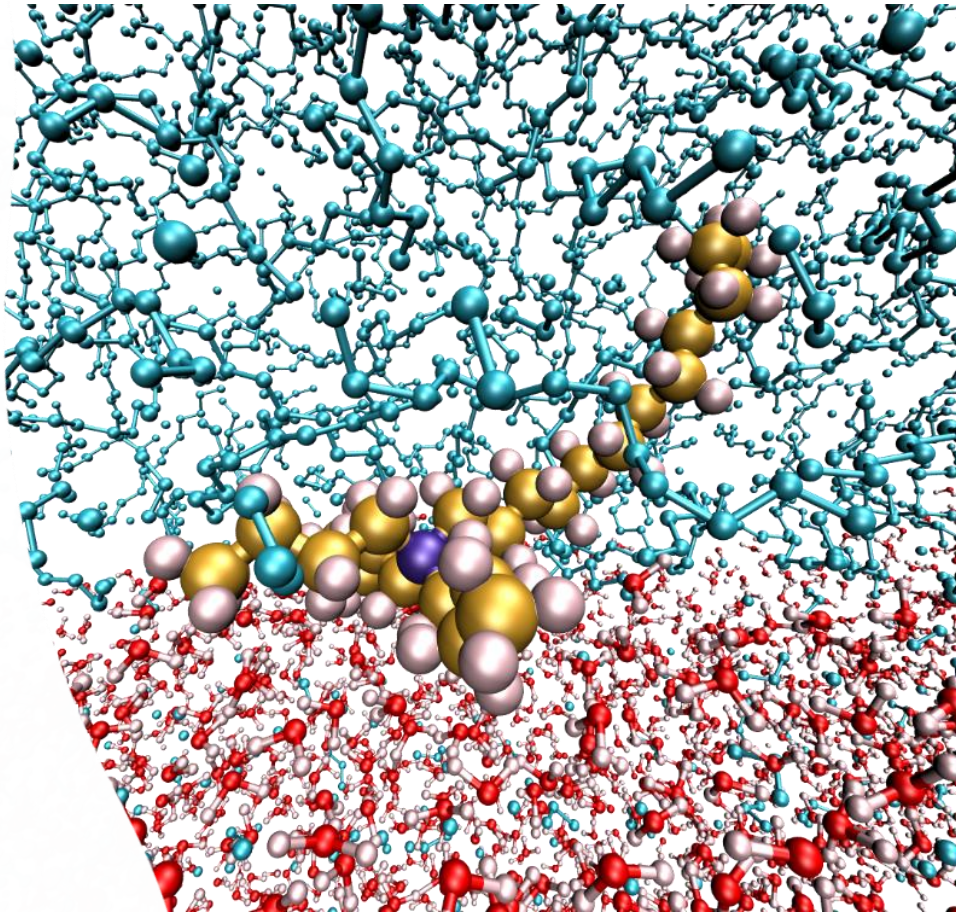


Image: Anti-agglomerant additive, © Nextmol

NEXTMOL

Use case:

Biodegradable polymers for hair care applications

Polymers – a key ingredient of hair care products

- Polymers are ubiquitous in hair care products: Viscosity modifiers, emulsifiers, “hold” in hair gels and sprays, conditioners to improve combing, etc.
- Polymer in liquid formulations are ingredients used in millions of products, but they cannot be recovered and thus create a large pollution problem.
- **Goal:** substitute established hair care polymers with biodegradable alternatives that exhibit the same (or even better) performance.

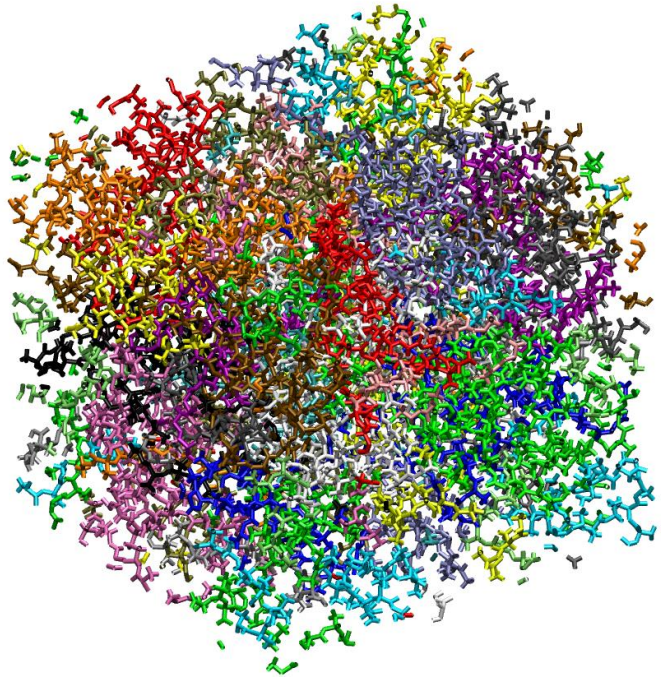
https://www.rsc.org/policy-evidence-campaigns/environmental-sustainability/sustainability-reports-surveys-and-campaigns/polymers-in-liquid-formulations-plfs/

Molecular modeling

The first step is to generate the polymer from its monomers:

- Inputs to provide consist of the SMILES of the repeating units and the capping groups.
- Copolymers can be either random or block, controlled by respective specification of the ratios and block sizes of the monomers.
- The final density of the generated polymer blends is then reached after a short MD equilibration.

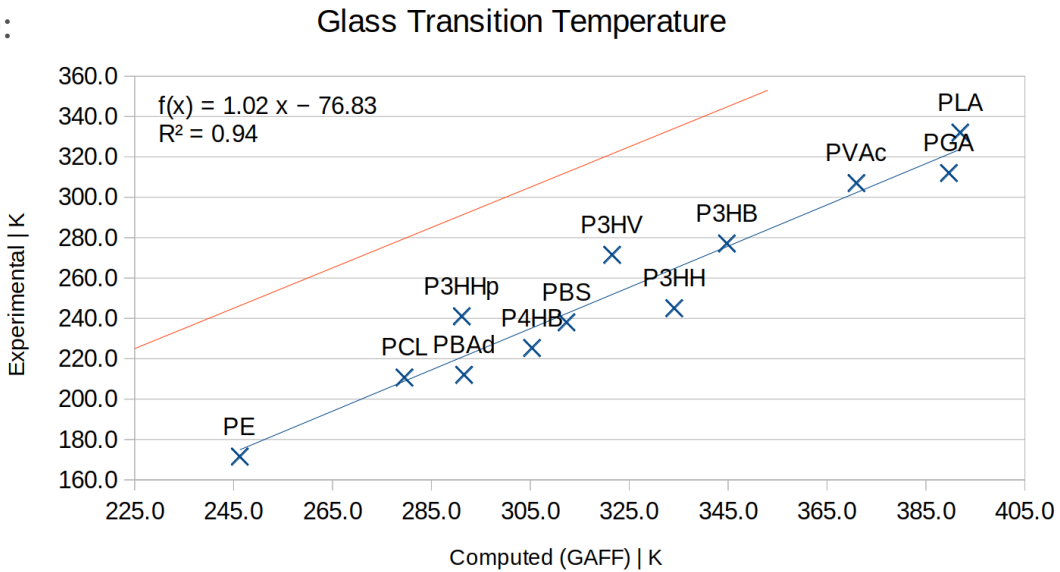
Density (g/cm^3)	Polymer Database	Nextmol builder
Poly(4-hydroxybutyrate)	1.22-1.28	1.283



Computational characterization

We characterize the polymers by calculating computational physico-chemical descriptors of polymer blends via Molecular Dynamics (MD):

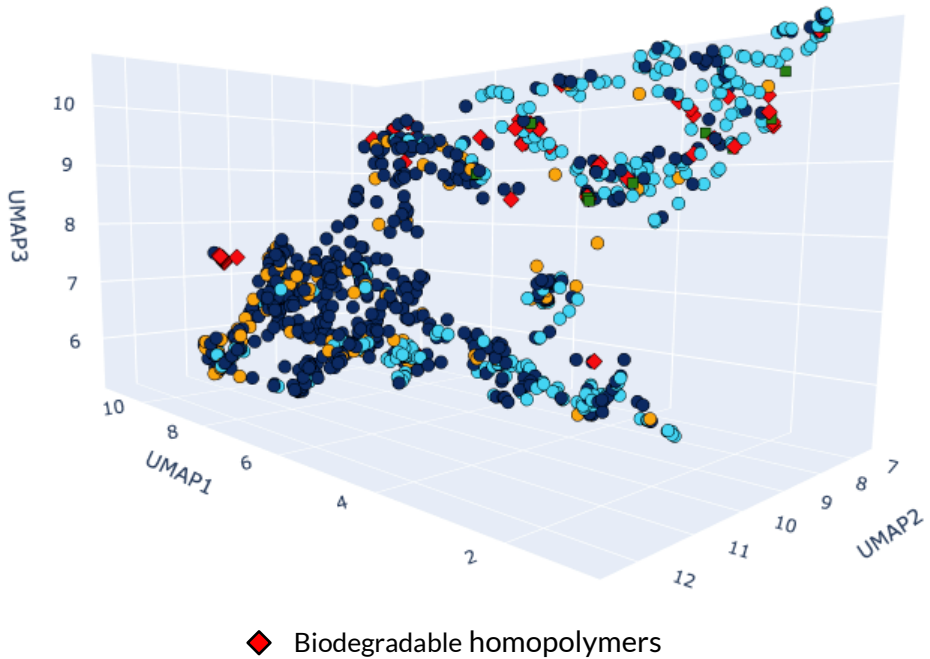
- Properties: Glass transition temperature, Hansen solubility parameters, persistence length, radius of gyration, …
- Good agreement with experimental reference values – we can rank the polymer blends against each other.
- Possibility to perform a virtual high-throughput screening.



Selecting biodegradable polymers via ML

Combining ML and MD allows us to focus on the most interesting candidates:

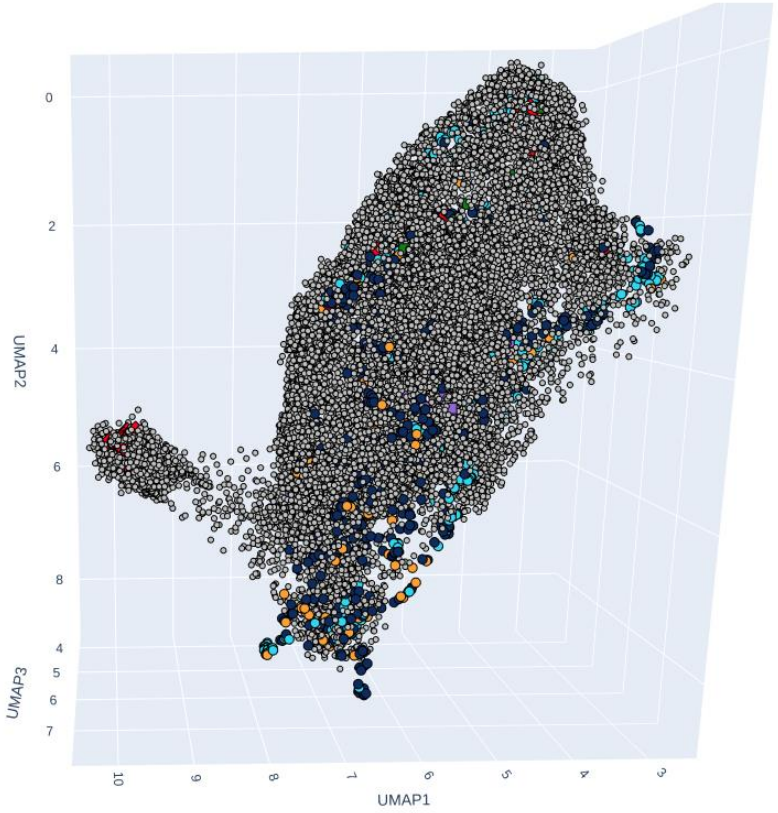
- We collected a dataset of 1,044 homopolymers from the literature.
- We analyzed the chemical space: molecular descriptors and dimensionality reduction.
- We selected those polymers that are close to the known biodegradable ones to contribute to our ML models → Focus of the subsequent MD calculations.
- These homopolymers of interest will be combined into copolymers for property tuning.



Biodegradable polymers

Polymer informatics with copolymers

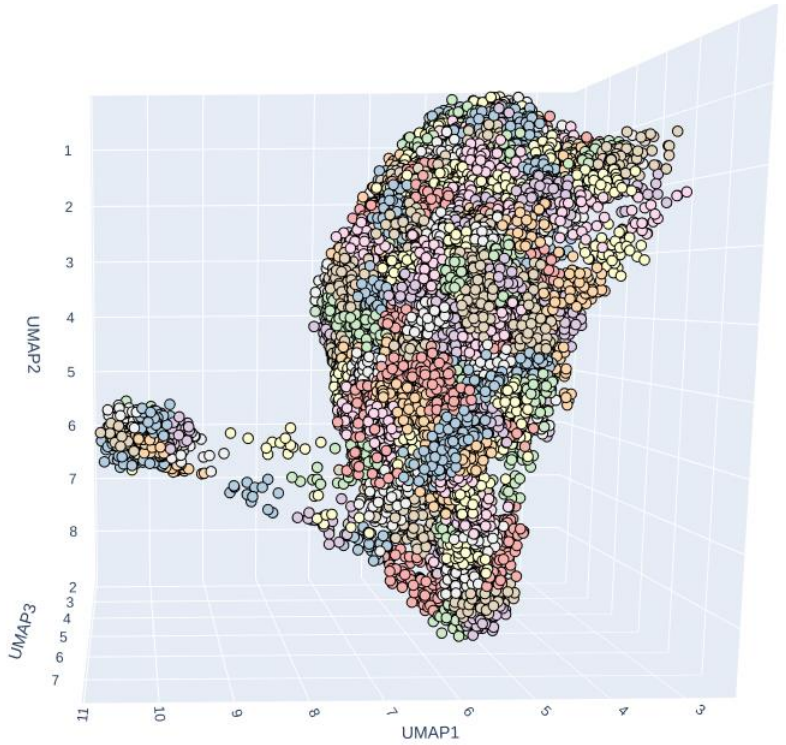
- 20,000 hypothetical copolymers have been generated by binary combinations of 67 biodegradable monomers at different ratios.
- These hypothetical copolymers are represented in grey in the figure. They permit to densely cover the chemical space of interest.
- Evaluation of their physico-chemical properties, such as Tg, will permit to train predictive ML model in order to accelerate the identification of promising candidates.



Polymer informatics with copolymers

As this is a huge number, it is necessary to prioritize the sampling of the chemical space that is relevant to our problem. For that purpose, we have adopted the following strategy :

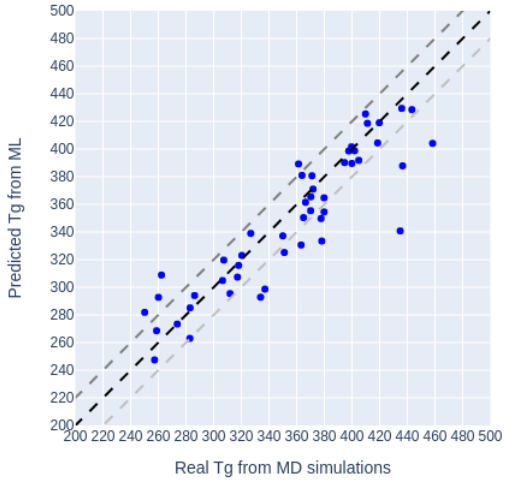
- Perform an unsupervised K-means clustering to identify 500 cluster of similar structures.
- Select the closest copolymer compositions to the cluster centroids.
- The selected structures are built, and their physico-chemical properties are evaluated via MD simulations.

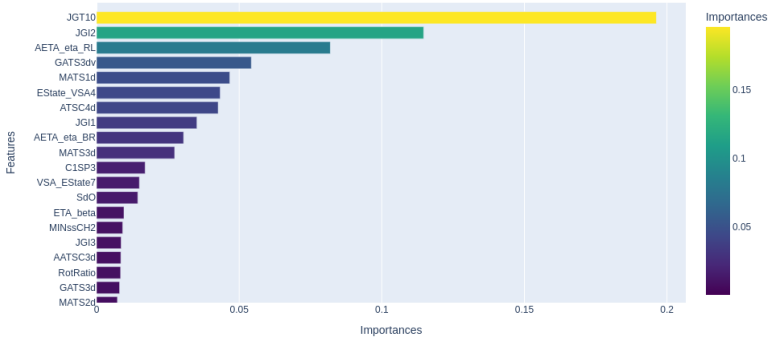


Machine Learning model

- We have trained supervised Machine Learning models on 500 polymers for which properties had been calculated via Molecular Dynamics simulations of 4 replicas:
  - Homopolymers, including hyaluronic acid, cellulose, starch, alginate and pectin
  - Copolymers, including copolymers with polysaccharides
- Properties: Glass transition temperature, Hansen solubility parameters
- 2D molecular descriptors, Random Forest
- 90/10 train-test split

Mean absolute error of 18 K and a R^2 of 0.8



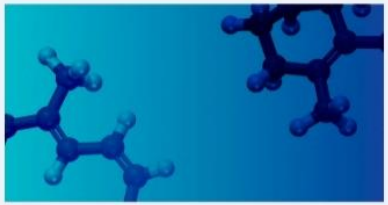


Most important features

Biodegradable polymers

NEXTMOL

Logout



Polymers

SMILES of monomer #1

O(C(*)(C(*)))

SMILES of monomer #2

*OC(CCCCCC)CC(*)(=O)

Ratio of monomer #1

0.50 − +

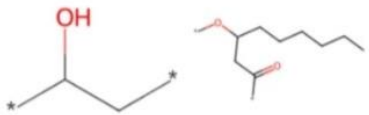
Model 1: Machine Learning prediction of Tg and HSP

The main outcome of these models is the prediction of the glass transition temperature (Tg).

The following solubility parameters are also predicted:

- Hildebrand: obtained from the total cohesive energy of the system
- Dispersion: obtained from the Lennard-Jones vdW contributions to the cohesive energy
- Electrostatic: obtained from the Coulombic contributions to the cohesive energy Use the sidebar to input the polymer of your interest.

Predictions



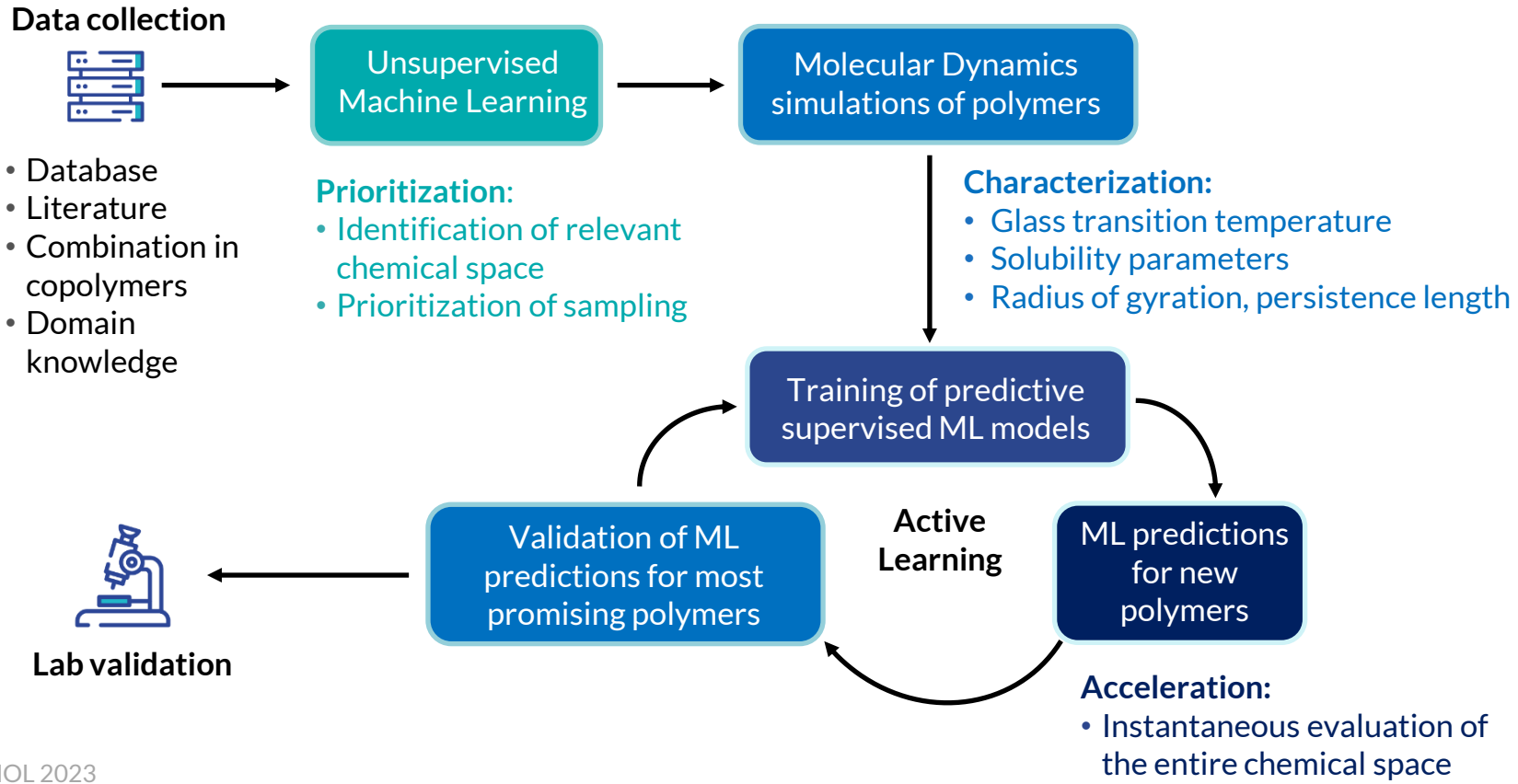
Ratio monomer #1: 0.5 | Ratio monomer #2: 0.5

Prediction

		ratio 2	Applicability Domain	vs AD threshold	rescaled Tg [K]	Tg [K]	Hildebrand [J/cm³]^(1/2)	Dispersion [J/cm³]^(1/2)	Electrostatic [J/cm³]^(1/2)
0	(=O)	0.5	inside	0.99	223.73	323.09	22.12	18.39	9.71

Download polymer predictions as CSV

Summary



NEXTMOL

Thank you!

www.nextmol.com
monica.demier@nextmol.com
linkedin.com/nextmol

Application to hypothetical polymers

- We analyzed the chemical space of a dataset of one million hypothetical polymers (PI1M) and compared with the known biodegradable ones.
- The biodegradable polymers are concentrated in two specific regions.
- 8,621 (out of 1 million) hypothetical polymers are located close to known biodegradable polymers and will therefore be studied in more detail.
- Acceleration of more than x100 compared to a brute force analysis of all polymers.