INNOVATION RESEARCH INTERCHANGE
Accelerating Value Creation

A division of the National Association of Manufacturers

IRI 2025 INNOVATORS SUMMIT

May 19–21 / Chicago, IL

INNOVATION RESEARCH INTERCHANGE
Accelerating Value Creation

A division of the National Association of Manufacturers

Accelerating Design through Formulations Modeling at the Molecular Level

Jeff Sanders, PhD, Scientific Lead, Consumer Goods, Schrodinger

IRI 2025 | INNOVATORS SUMMIT

Accelerating Design through Formulations Modeling at the Molecular Level

Jeff Sanders, PhD

May 2025

Schrödinger

Pioneering Digital Chemistry

- 30+ years of innovation
- Over 850 employees worldwide; >40% Ph.D.
- >50% of employees dedicated to R&D
- ~1,785 customers worldwide
- Pipeline of 25+ collaborative and proprietary programs

[redacted]

The world’s most innovative companies use Schrödinger

Panasonic

reckitt

KONICA MINOLTA

L’ORÉAL

SAMSUNG

abbvie

سابك
sabic

SOLVAY

EVONIK
Leading Beyond Chemistry

U.S. AIR FORCE

CAMBRIUM

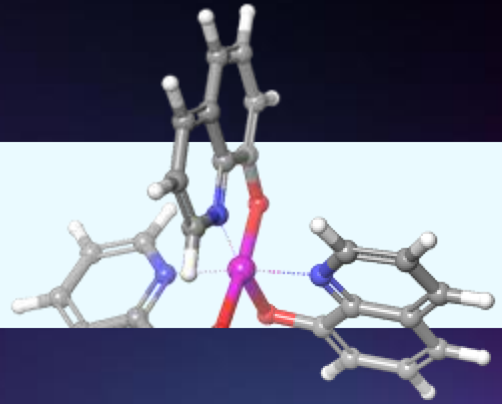
SEPION
TECHNOLOGIES

The future of materials discovery

If all properties can be calculated with perfect accuracy, designing materials would have a **higher success rate**, be **faster** and **cheaper**, and would produce much **higher-quality** materials

“All” synthesizable chemistry ($\sim 10^{10}$ - 10^{80})

Select best candidate system



- ✓ Reactivity
- ✓ Selectivity
- ✓ Solubility
- ✓ Sustainability
- ✓ Redox
- ✓ Kinetics
- ✓ Stability
- ✓ Synthesizability

Physics-based modeling and machine learning

Physics-based modeling

Physics & Machine Learning

Incorporate physics-based information about materials into practical ML models.

Build targeted ML models to expand the impact of physics-based simulations.

Machine learning

Vertical scaling of *in silico* technology to traverse project-specific ideas

Physics and machine learning

PHYSICS + AI generates molecular morphologies and complex structures → **PHYSICS + AI** generates 1,000,000,000s of candidates → **PHYSICS** determines properties of subset of systems (10,000s) to build ML training set → **MACHINE LEARNING** scores 1,000,000,000s of candidates → **PHYSICS** computes properties of top 10,000s candidates → **MAKE & TEST** 10s of systems → **ADVANCE BEST MOLECULE OR MATERIAL**

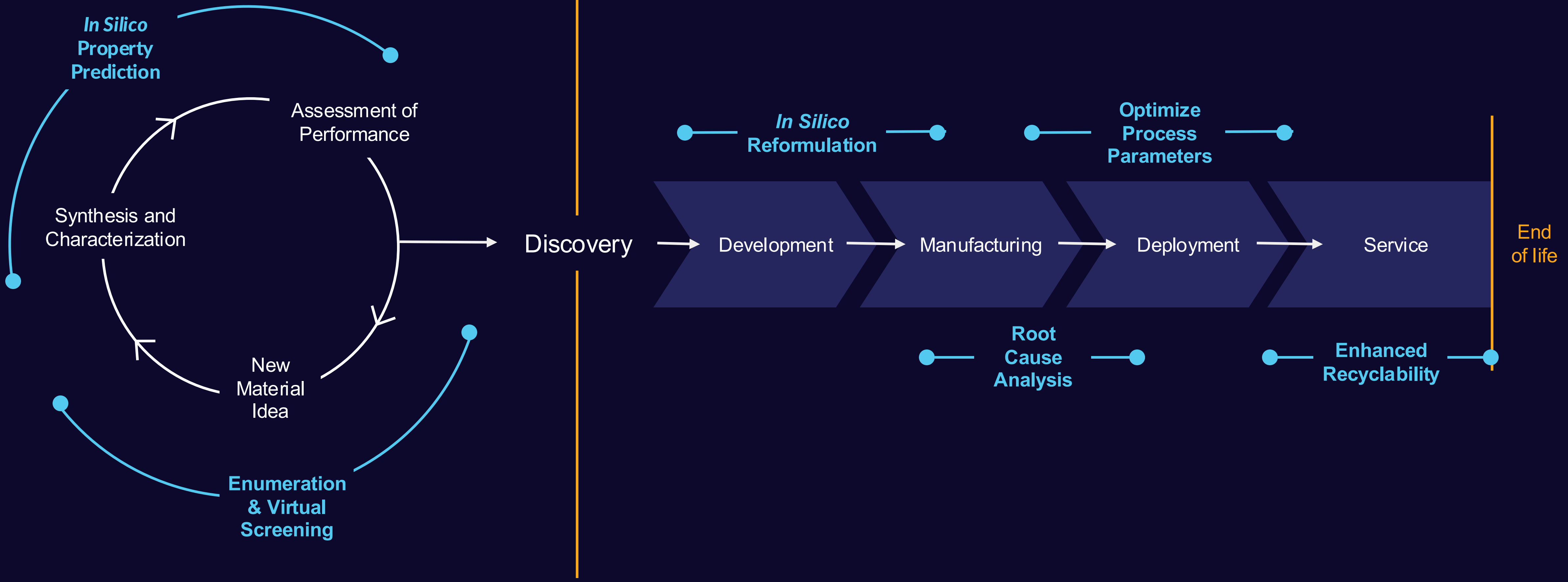
Enabled and accelerated by high performance compute and cloud technology platforms

Computational Chemistry

Modeling impact on materials design

Design Cycle

Lifespan



Benefits of leveraging digital technology

Less

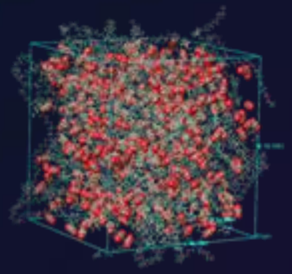
- **Time** to insights and target solutions
- **Cost** to optimize materials development process
- **Experimental synthesis and testing** of materials with undesirable properties
- **Distance** between teams and expertise areas

More

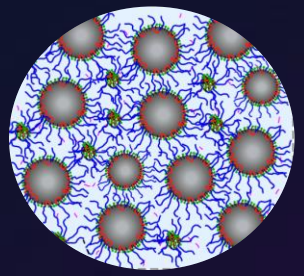
- **Hypotheses** to test
- **Access** to chemical space
- **Optimization** of multiple property parameters at the same time
- **Dynamic collaboration** in the design process
- **High-quality** materials with desired performance and properties

Bottom-up approach to meeting consumer needs

Size

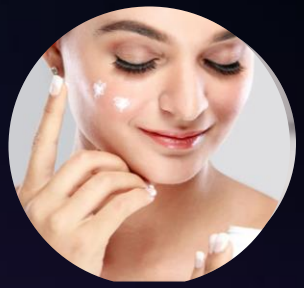


Molecular
Physiochemical properties



Microstructure
Morphology

Macroscopic
Performance



Product
Consumer needs

Simulate

In action

↑

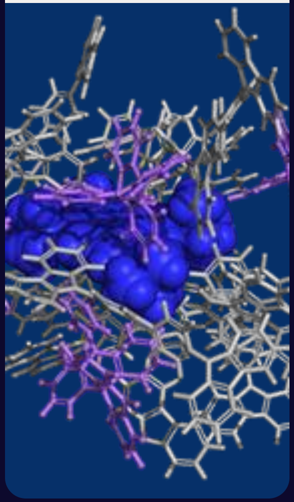
Formulation

↑

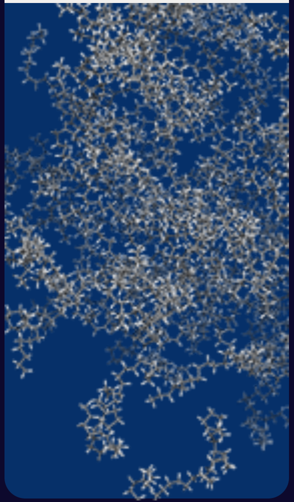
Ingredients

Solutions for all applications

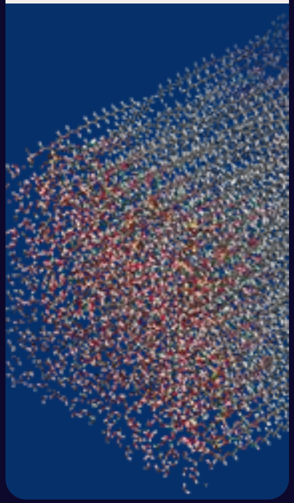
Organic Electronics



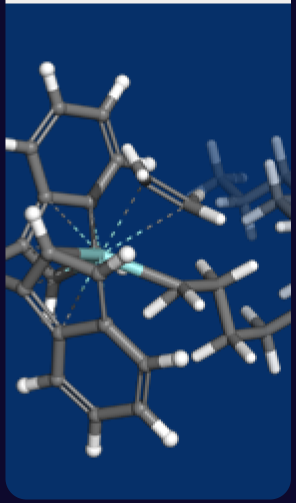
Polymeric Materials



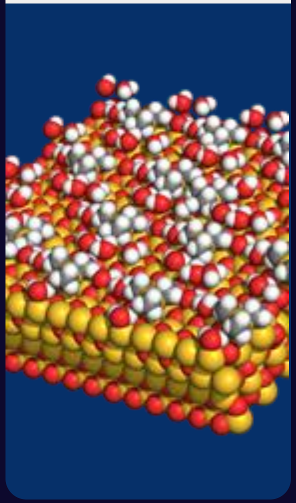
Consumer Packaged Goods



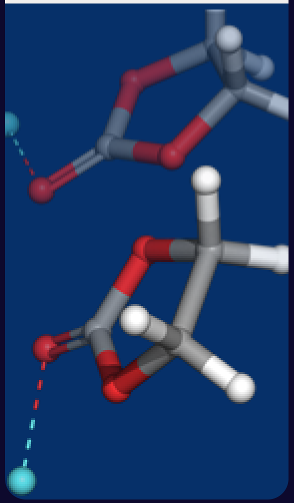
Catalysis & Reactivity



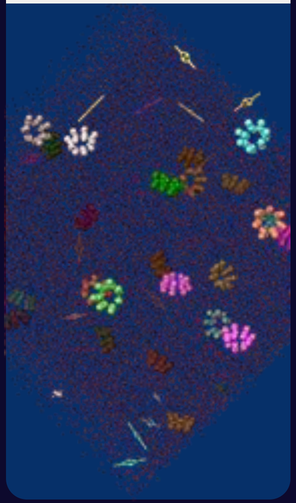
Semiconductor



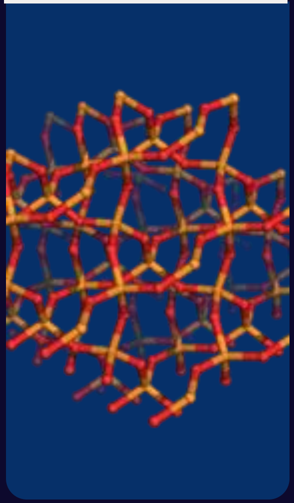
Energy Capture & Storage



Pharmaceutical Formulation & Delivery



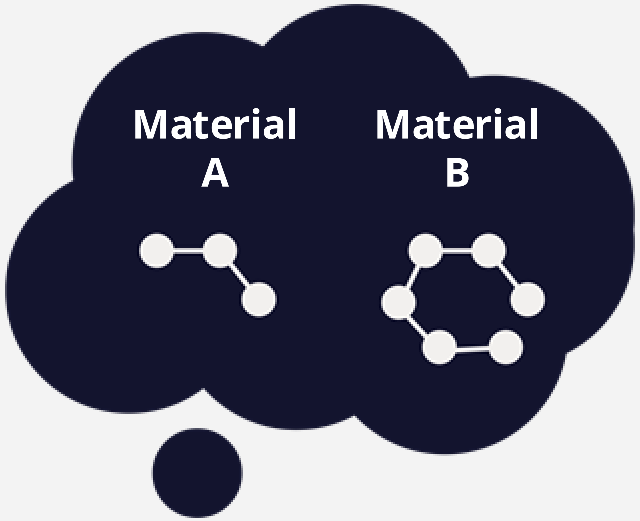
Metals, Alloys & Ceramics



Machine Learning (ML):
statistical models that computers use to perform a task

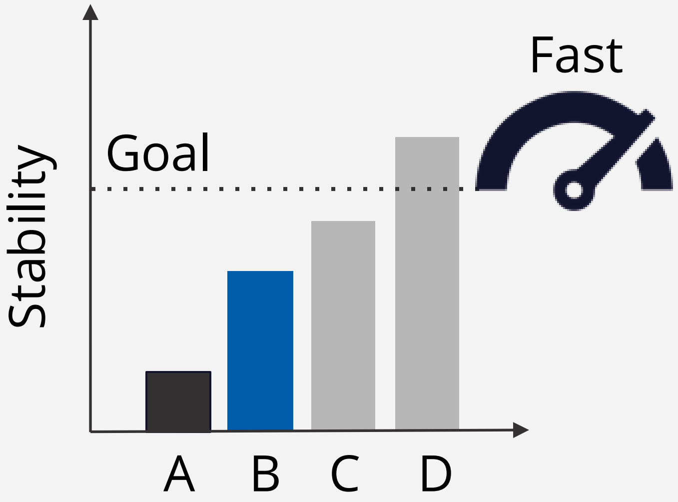
ML-accelerated materials design

Idea Generation



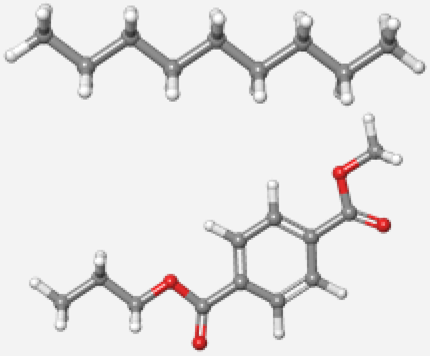
Machine Learning

Prediction



Quantitative Structure Property Relationship (QSPR) Modeling

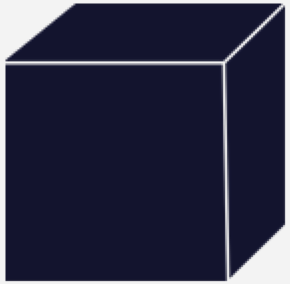
Structures



Features or Descriptors

$x_1 , x_2 \dots x_n$

ML Model



Property

Regression task; e.g. density

Classification task; e.g. solid or liquid phases

How do we select the best machine learning model?

Descriptors

Cheminformatic Descriptors
- Molecular
- Polymeric
- …

Cheminformatic Fingerprints
- Linear
- Radial
- Dendritic
- …

Inorganic Descriptors

3D Descriptors

…

ML Algorithms

Partial least squares regression (PLS)

Best subset multiple linear regression (MLR)

Kernel-based PLS (kPLS)

Principal components regression (PCR)

Naive Bayes classification

Ensemble recursive partitioning

…

Properties

Glass transition temperature (T_g)

Solubility

Recyclability

ΔE_{ST}

Reaction rate

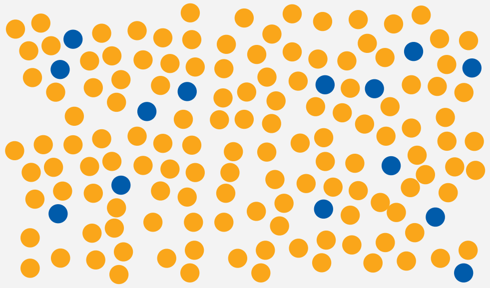
Sweet or not sweet

…

AutoQSAR: customizable and easy to visualize

Training Options

Determine training/test split (default 75-25)



Determine number of models to keep (default 10)

Single Model M_i
Or
Consensus $\Sigma_i M_i$

View Model Reports

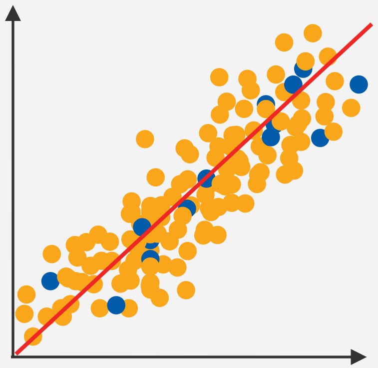
Scores and rankings:

Model scores

Training and test set R^2

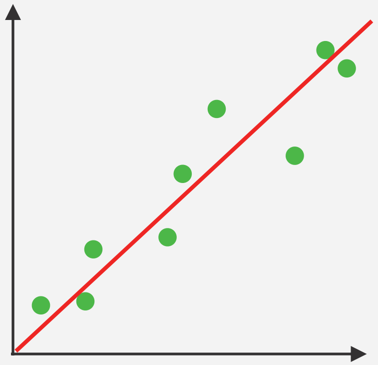
Test set RMSE

Scatter plot with color-coded training and test set



Make Predictions

Prediction with the *best model* or *ensemble* prediction from the top N models

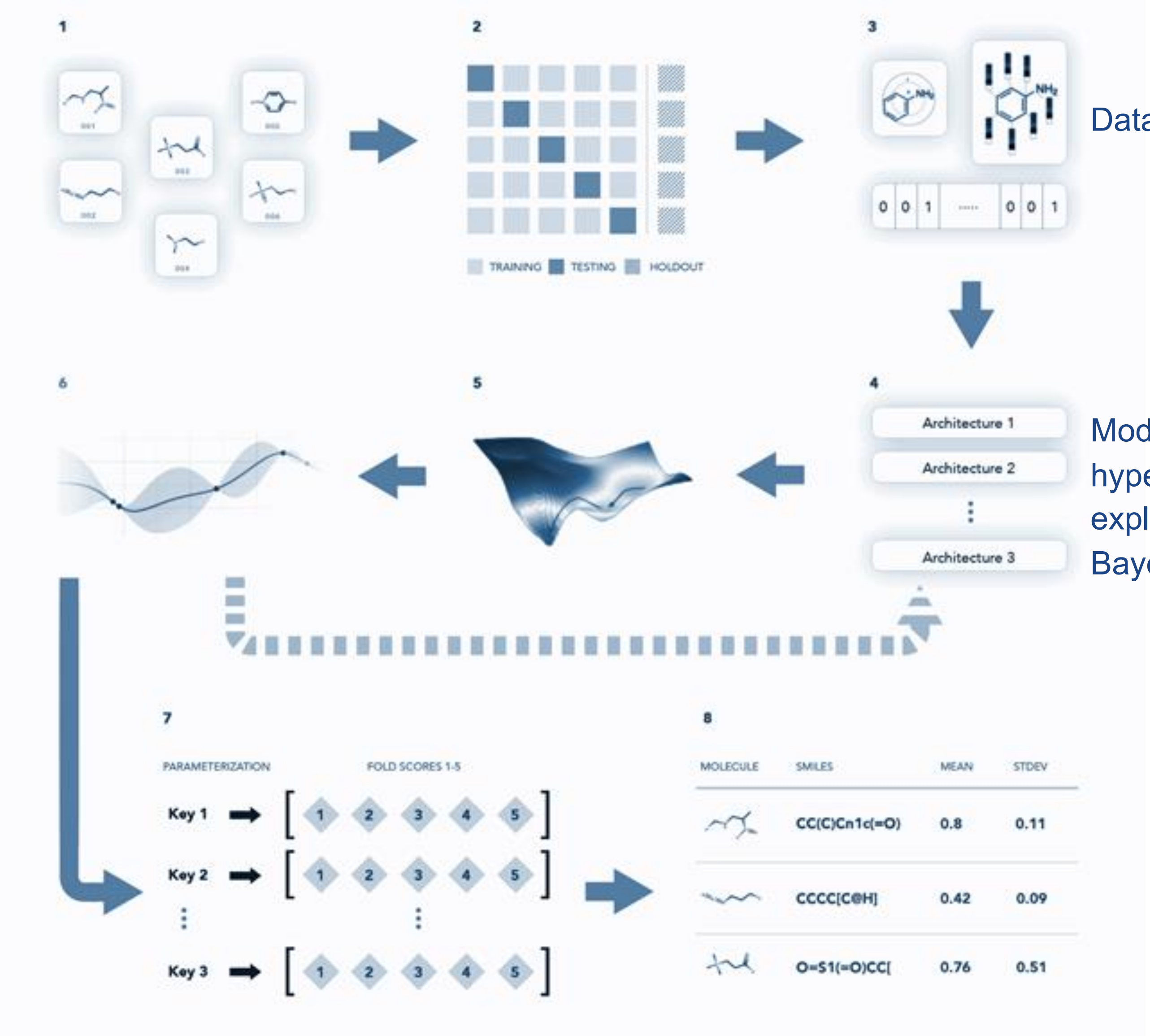


Domain applicability (score and alert)

Structure similarity

Comparison to training set

DeepAutoQSAR: automatic selection of best ML models



Data splits, featurization

Model architecture, descriptor and hyperparameter combinations explored and optimized via Bayesian Optimization

Models Sampled

- Dense Neural Network
- Random Forest
- XGBoost
- TorchGraphConv
- GCN
- GraphSAGE
- GIN
- TopK
- SAGPool
- EdgePool
- GlobalAttention
- Set2Set
- SortPool

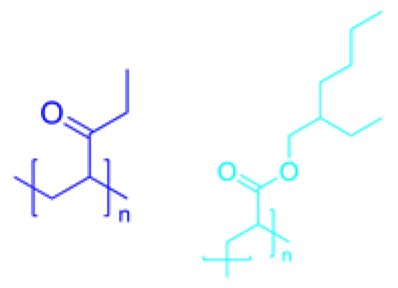
Consensus Model

- Prediction = an average of the predictions for 3 best models
- Uncertainty = SD across the predictions

For more details, see DeepAutoQSAR White Paper:

Extending DeepAutoQSAR: formulation machine learning

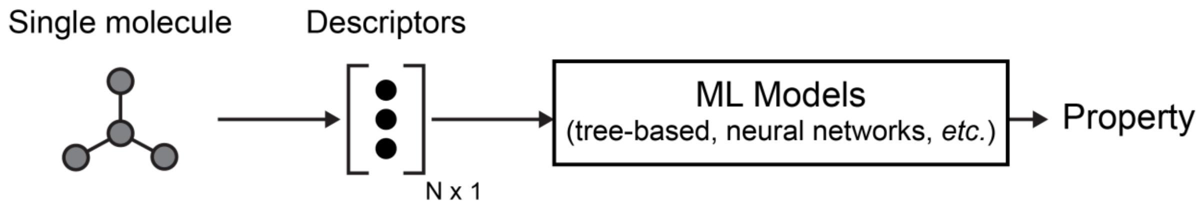
Example: Predict Copolymer T_g



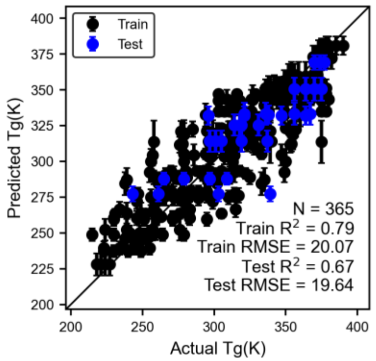
76%   24%

Glass transition temperature (T_g) = 264 K

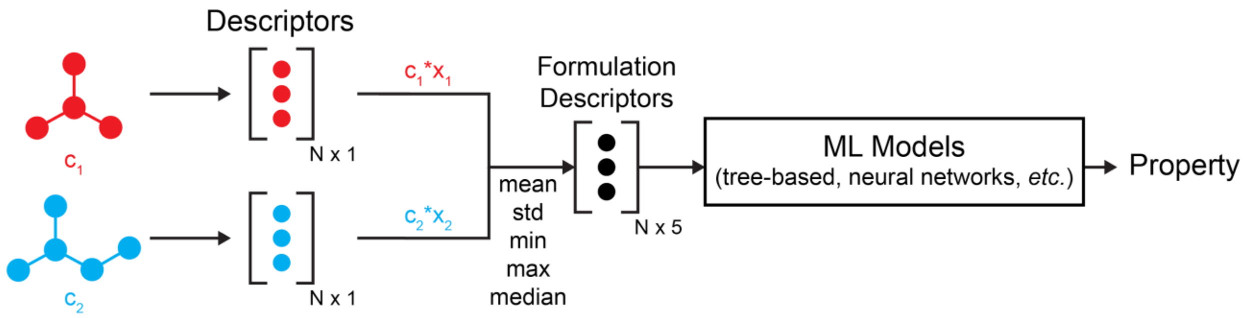
DeepAutoQSAR (single molecule): Encode mixtures as a single SMILES and compositions as additional features



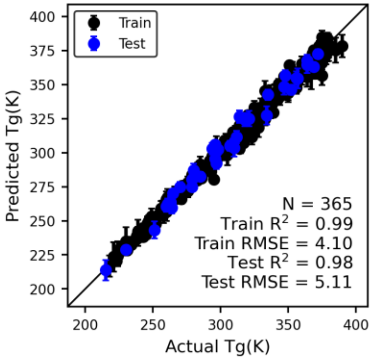
Performance on copolymer system*



Formulation Machine Learning Approach



Performance on copolymer system*



Formulation machine learning enables accurate screening of ingredients and compositions

Formulations ML: statistical models that incorporate ingredients and composition

Formulations are fundamental to our daily lives

Everyday consumer products
(Shampoos, perfumes, plastic)

Energy Storage
(Electrolytes in batteries)

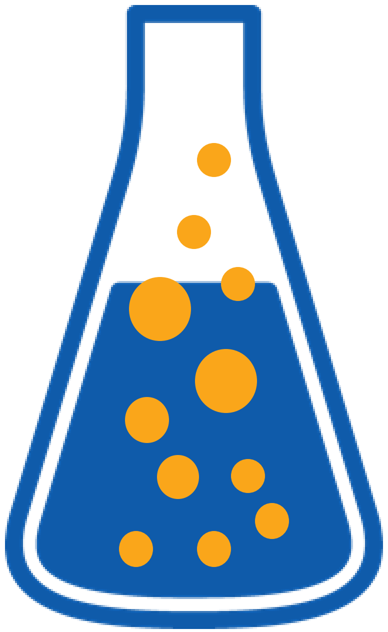
Formulations
Complex, multicomponent mixtures prepared based on a composition

Pharmaceutical formulation
(Medicine, drugs)

Oil and gas
(Gasoline, lubricants)

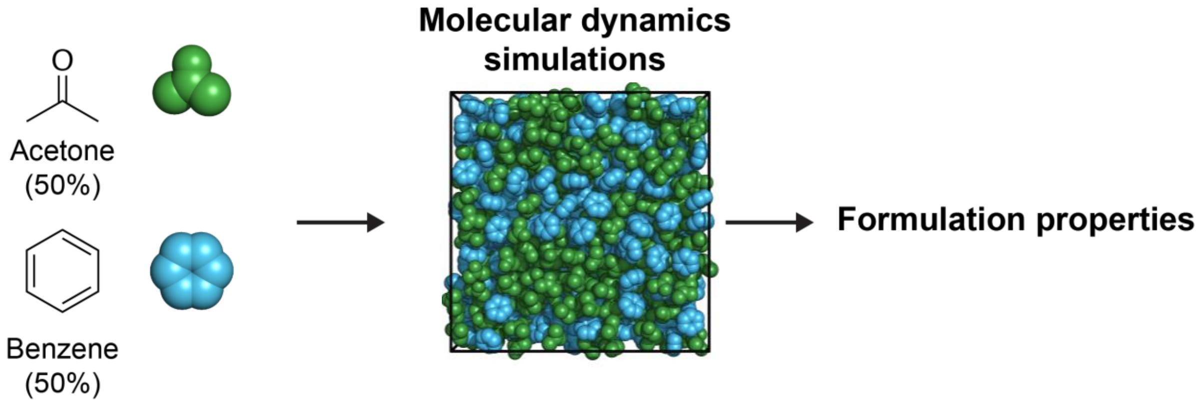
Computer-aided formulation design

Experiment
(slow + expensive)



Molecular Dynamics (MD) Simulations

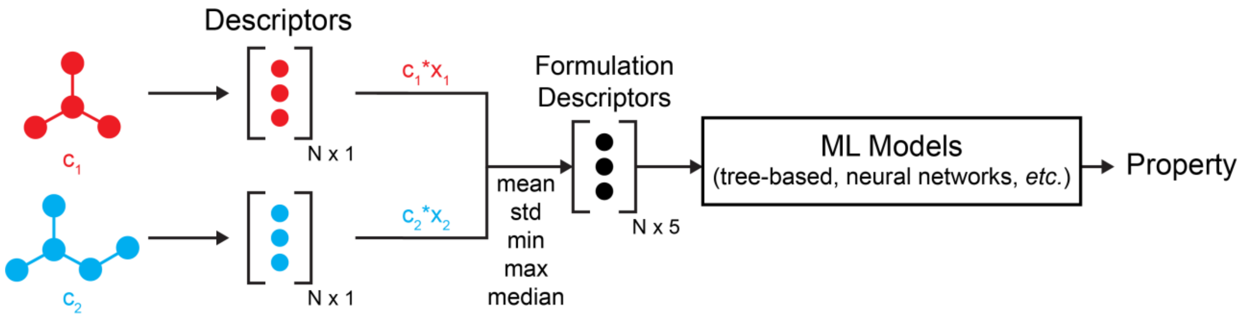
MD simulates all interactions between ingredients in a formulation



- ☑ Mechanistic insight
- ☑ Extrapolates well
- ✗ No reactions
- ✗ Expensive (~hrs)

Machine Learning (ML)

ML allows you to map structure and composition to property

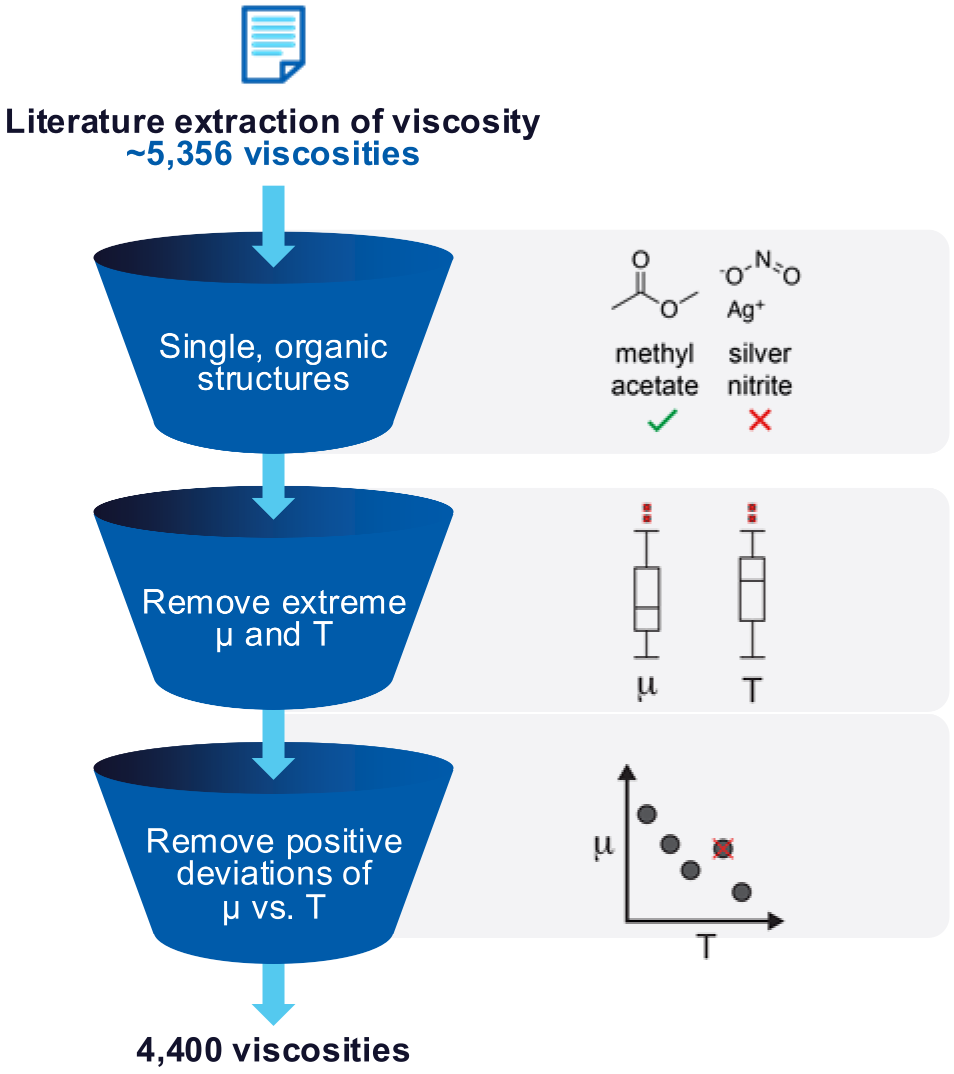


- ☑ Efficient
- ☑ Fast screening (~s)
- ✗ Requires data
- ✗ Extrapolation may be poor

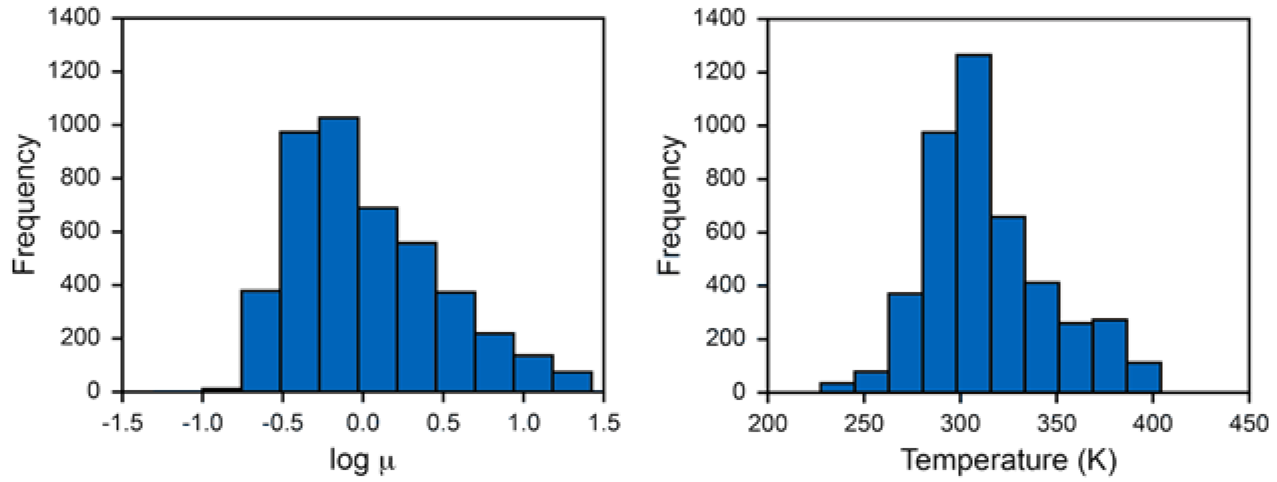
Schrödinger

Example - Impact of MD-derived simulation descriptors for predicting viscosity

Viscosity dataset for machine learning models



Distribution of viscosity and temperature

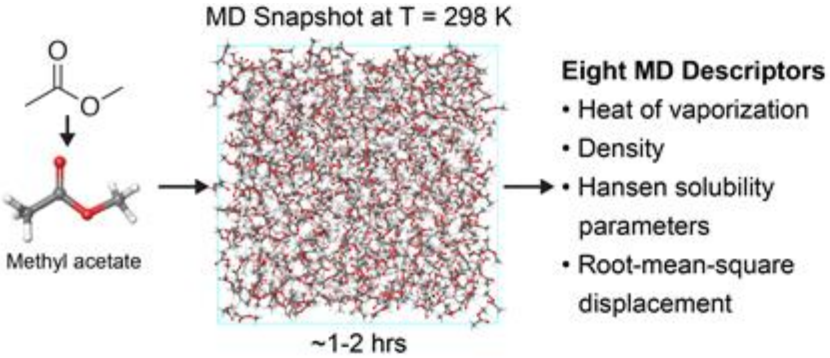


Dataset summary:

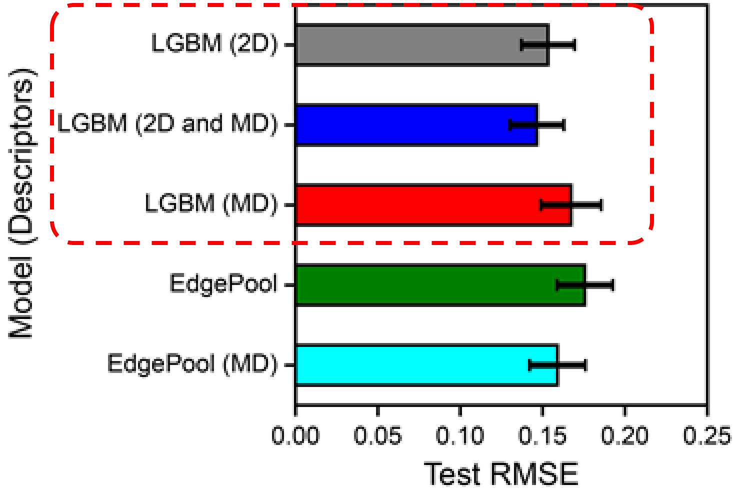
- **1,005** unique molecules
- Atomic elements of {H, C, N, O, F, Si, P, S, Cl, Br, and I}
- Viscosity between **0.10** to **26.52 cP**
- Temperature is between **227 K** to **404 K**

Impact of MD-derived simulation descriptors on viscosity

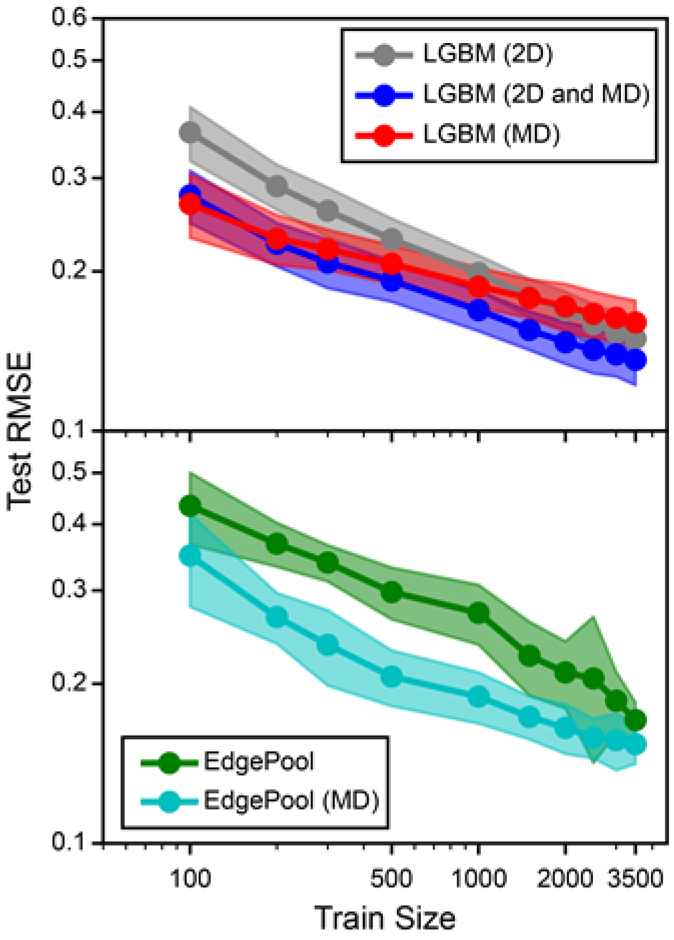
Generation of MD descriptors

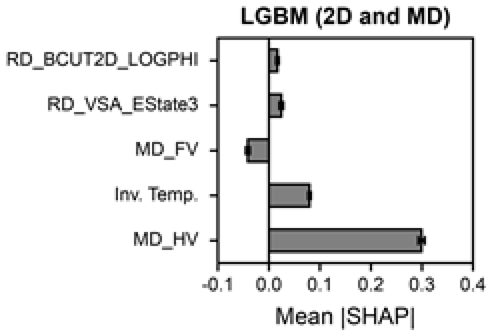


Performance with large dataset



Learning curve with and without MD descriptors





- MD descriptors increase accuracy, especially for small training size (<1,000)
- MD descriptors capabilities can be extended to formulations/mixtures and polymers

Case Study 1 – Machine Learning models for formulations development

Predicting large-scale miscibility

Challenge

Solvent miscibility is complex, especially in ternary or higher-order systems due to non-additive interactions

Solution

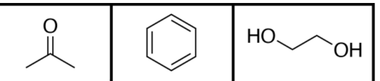
ML can capture complex interactions, making it well-suited for predicting miscibility in multicomponent systems

Results

Physic-based Formulations Modeling of over 30K mixtures result in highly accurate models

Impact

Miscibility prediction of new multicomponent mixtures can be predicted in seconds



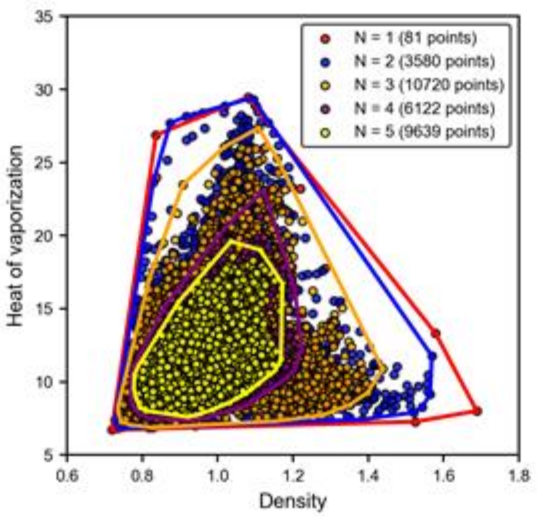
Miscibility	Acetone	Benzene	1,2-Ethanediol
Acetone	-	miscible	miscible
Benzene	miscible	-	immiscible
1,2-Ethanediol	miscible	immiscible	-

••• 25 solvents

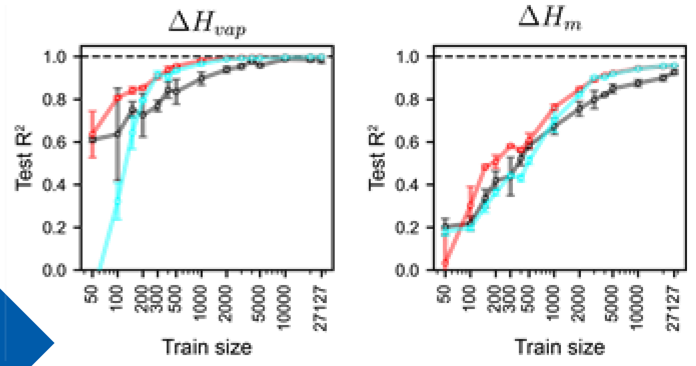
⋮ 81 solvents

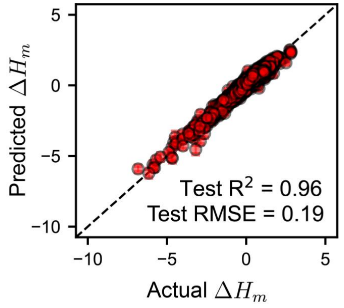
Formulations: Miscible solvents
- ✓ Acetone, Benzene
- ✓ Acetone, 1,2-Ethanediol
- × Benzene, 1,2-Ethanediol

Physics modeling



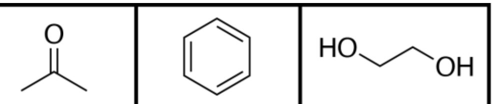
ML modeling





ML needs data: miscibility of organic solvents

CRC handbook of miscible solvents



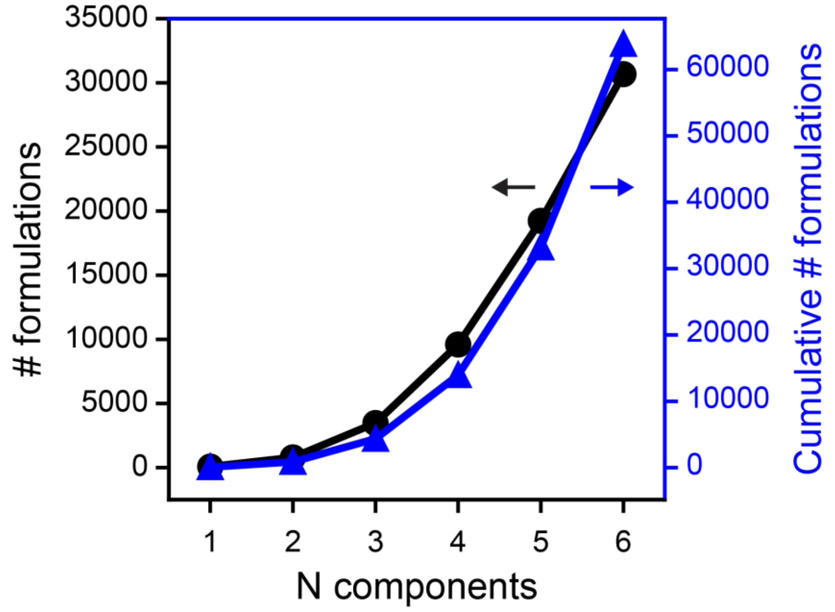
Miscibility	Acetone	Benzene	1,2-Ethanediol
Acetone	-	miscible	miscible
Benzene	miscible	-	immiscible
1,2-Ethanediol	miscible	immiscible	-

••• 25 solvents

⋮ 81 solvents

Formulations: Miscible solvents
- ✓ Acetone, Benzene
- ✓ Acetone, 1,2-Ethanediol
- × Benzene, 1,2-Ethanediol

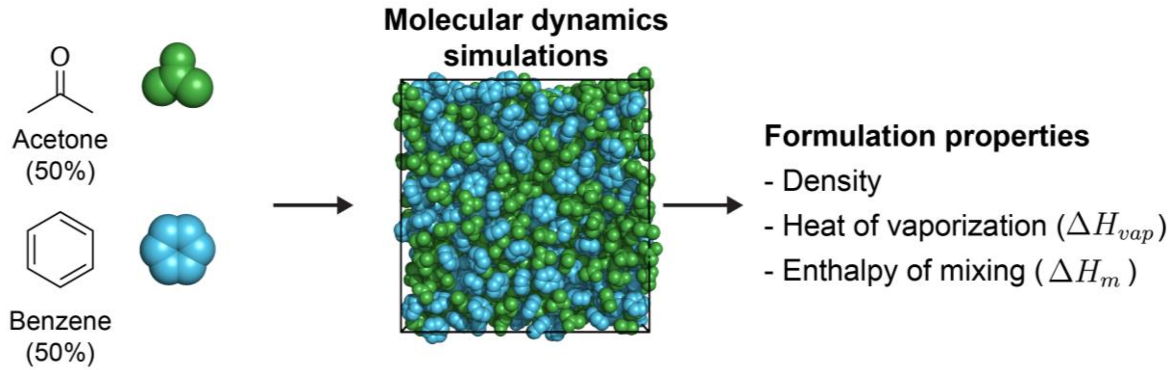
Combinations of 81 solvents



- By enforcing **experimentally determined miscibility rules**, we increase the likelihood of fully miscible, multicomponent mixtures
- For *N* components up to five, **19,238** combinations are possible

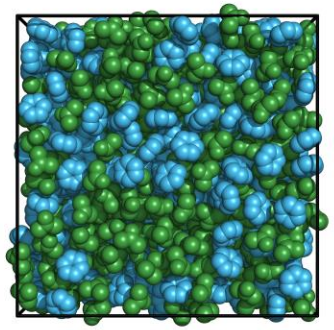
Physics-based simulations to screen mixture properties

MD generates physically relevant properties



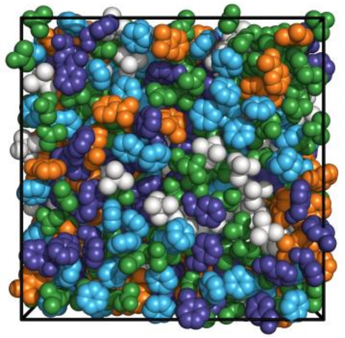
Simulation videos

N = 2



Acetone | Benzene

N = 5

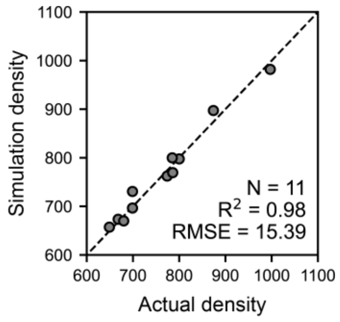


Acetone | Benzene | Benzyl alcohol | N,N-Dimethylaniline | Tetrachloromethane

MD captures experimental trends

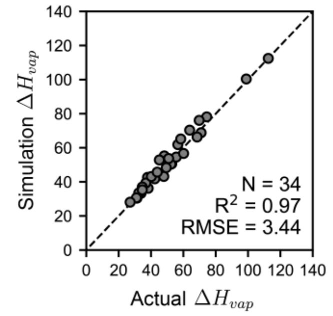
Density

Measures how dense molecules are packed in g/cm^3



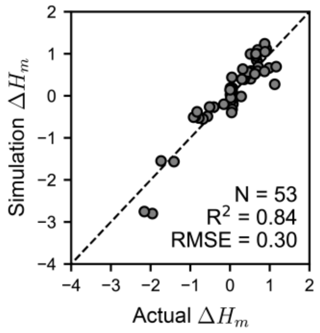
Heat of vaporization

Amount of heat in kJ/mol to convert liquid to vapor



Enthalpy of mixing

Measures energy released or absorbed upon mixing in kcal/mol

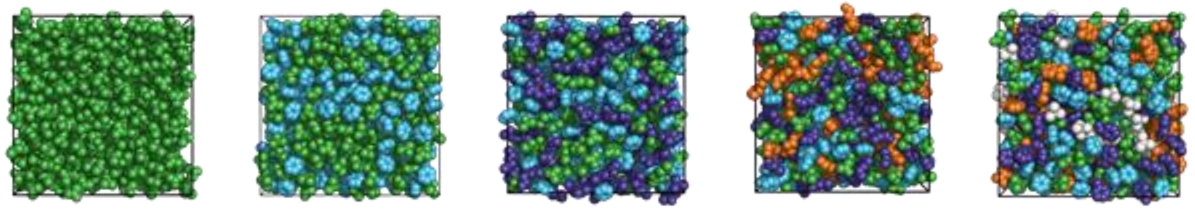


High-throughput MD Simulations to Create Large Dataset

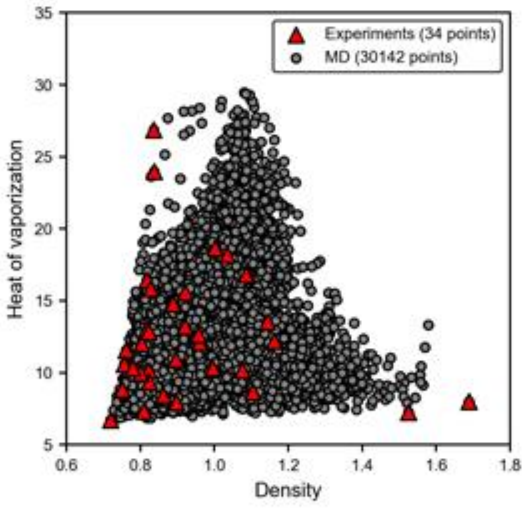
Summary of high-throughput mixture simulations

N components	Compositions	#unique formulations	#examples
1	100	81	81
2	20:80 40:60 50:50 60:40 80:20	716	3,580
3	20:20:60 20:60:20 60:20:20 33:33:33	2,680	10,720
4	25:25:25:25	6,122	6,122
5	20:20:20:20:20	9,639	3639
Total		**19,238**	**30,142**

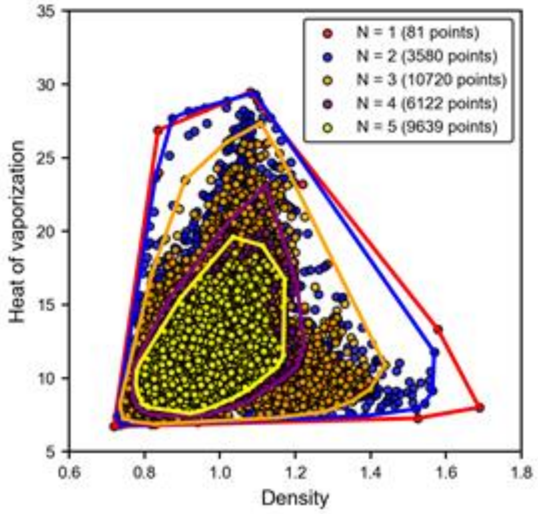
N = 1 N = 2 N = 3 N = 4 N = 5



MD opens opportunities to fine-tune property space



Increasing *N* leads to highly specific properties

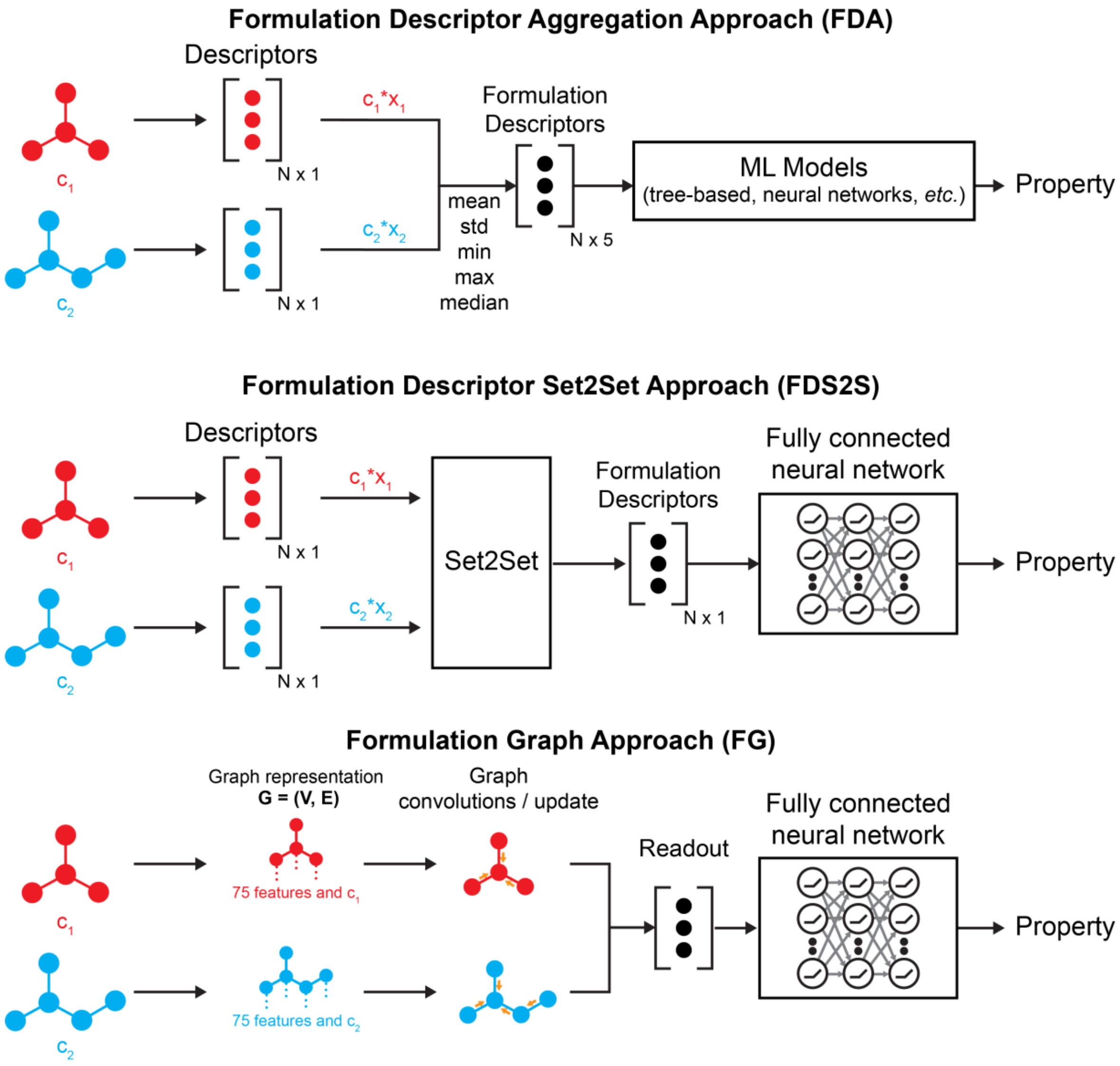


How scalable are our MD workflows?
30,142 simulations (651 ns / day | ~1.48 hours per simulation) = **45,000 GPU-hours** (~5 GPU-years)

Different approaches for machine learning in formulations

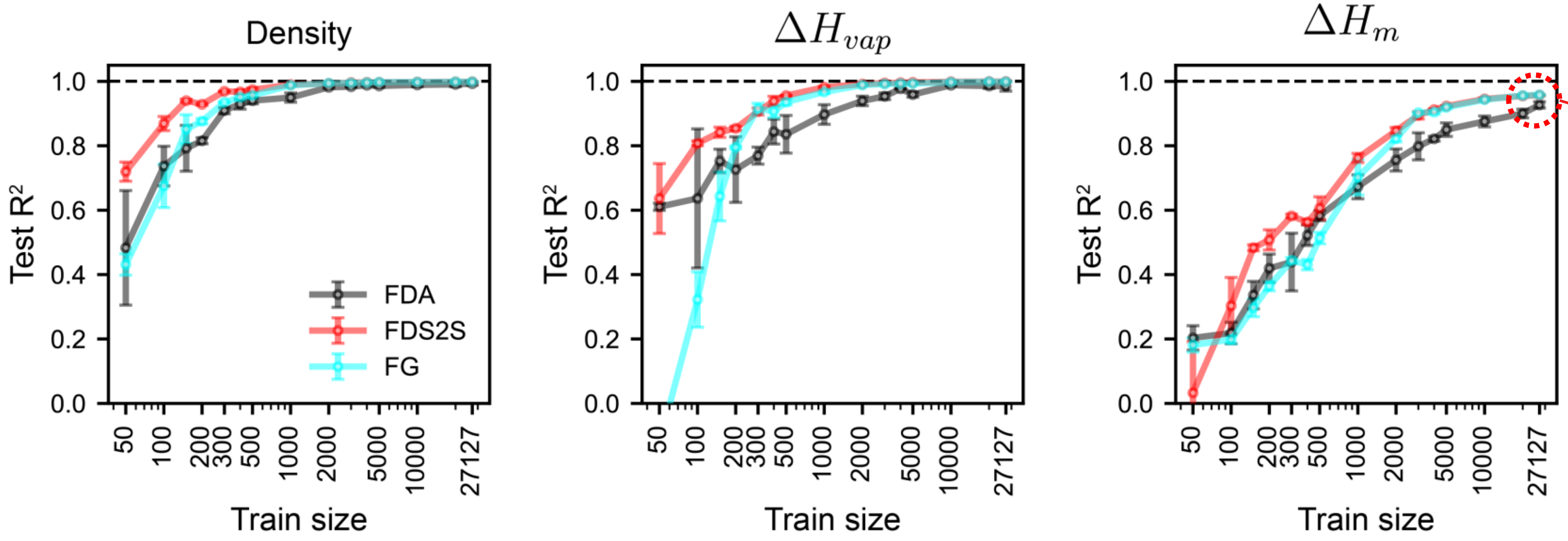
Requirements for Formulations ML:

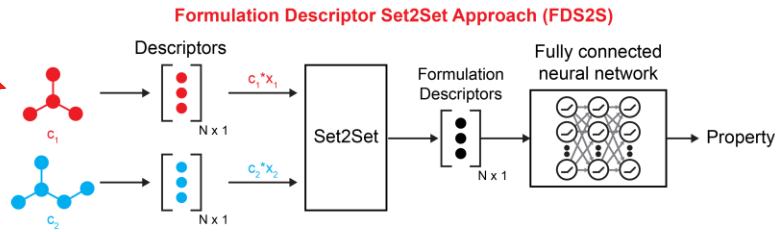
- **Compositionally-aware:** Composition must be embedded into the model
- **Permutation invariance**: Switching ingredient orders do not impact predictions
- **Component independence:** Flexibility in the number of components



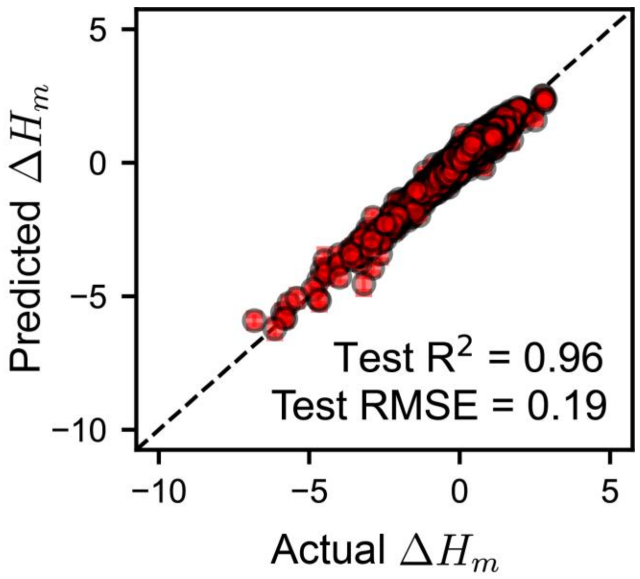
ML accurately predicts formulation properties

Learning curve: Left-out test set performance as a function of train size





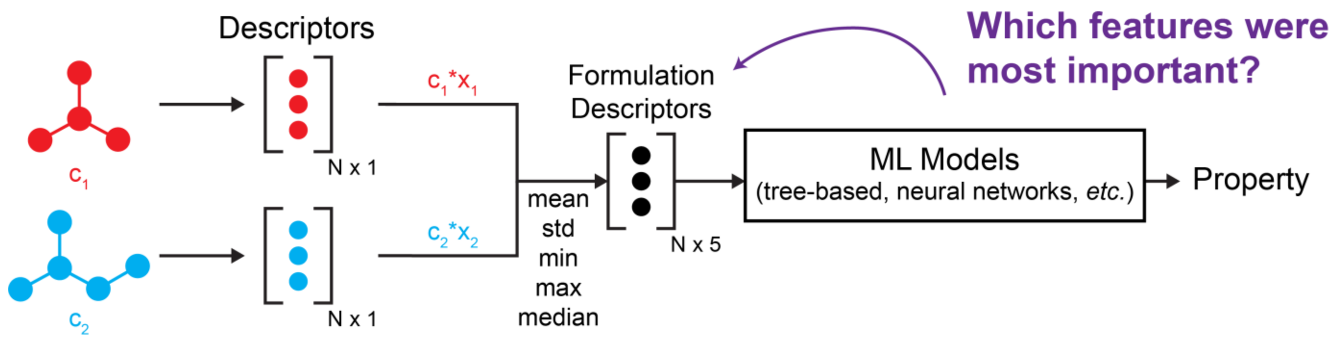
Parity plot of test set



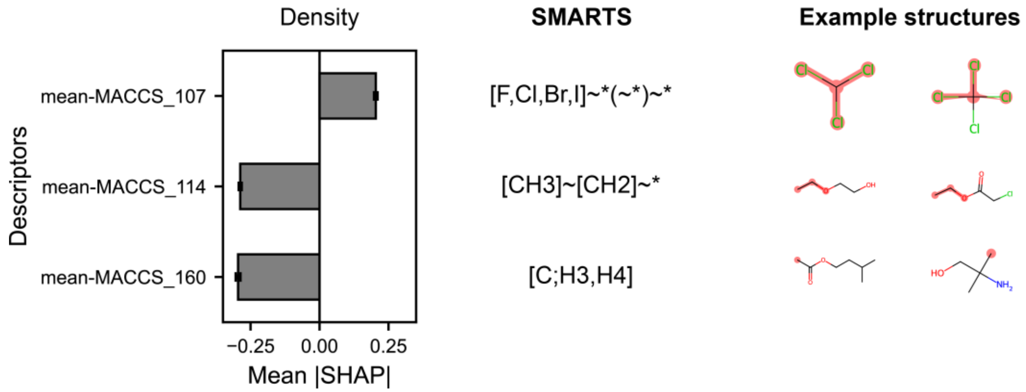
- FDS2S performs the best in predicting formulation properties at all data scales
- With enough data, ML can accurately predict formulation properties (even ΔH_m)

Schrödinger

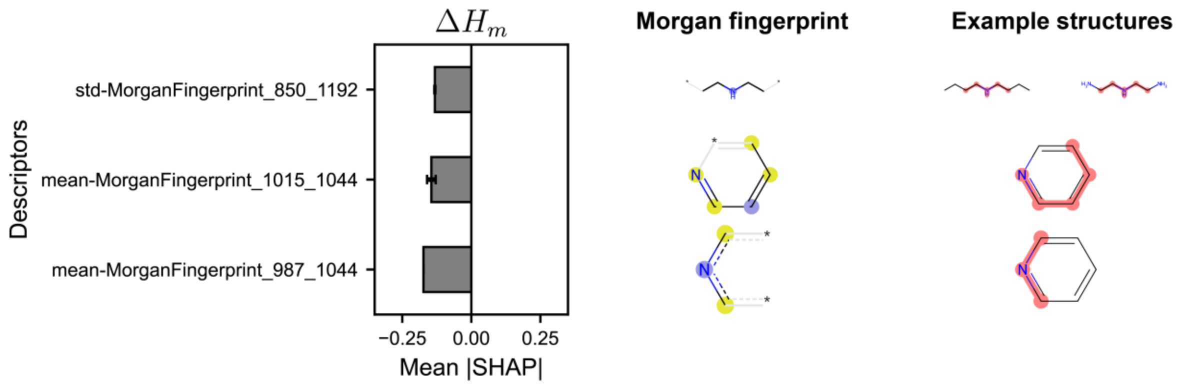
Important features highlighted from ML models



Feature importance: Density



Feature importance: Enthalpy of mixing

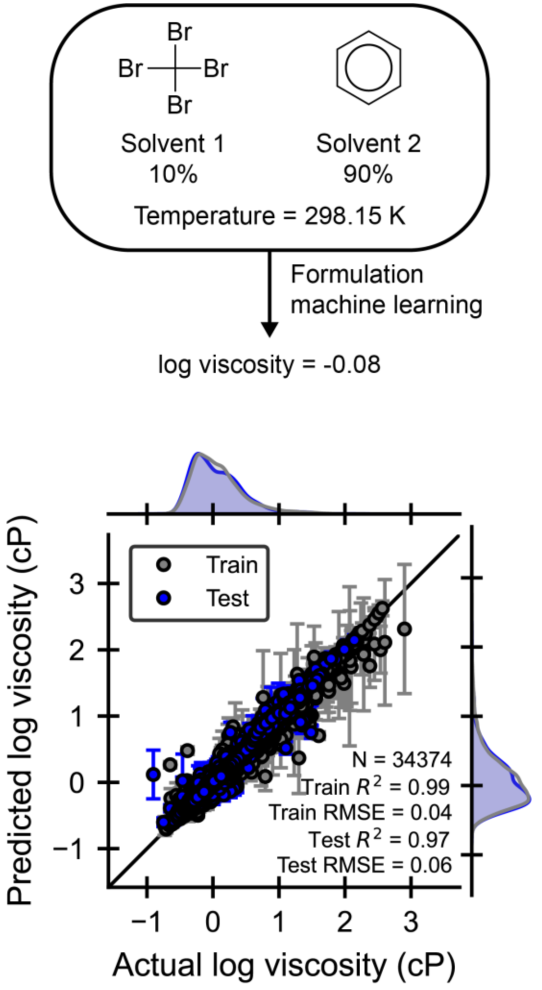


Inclusion of halogens and removal of methyl groups increase density

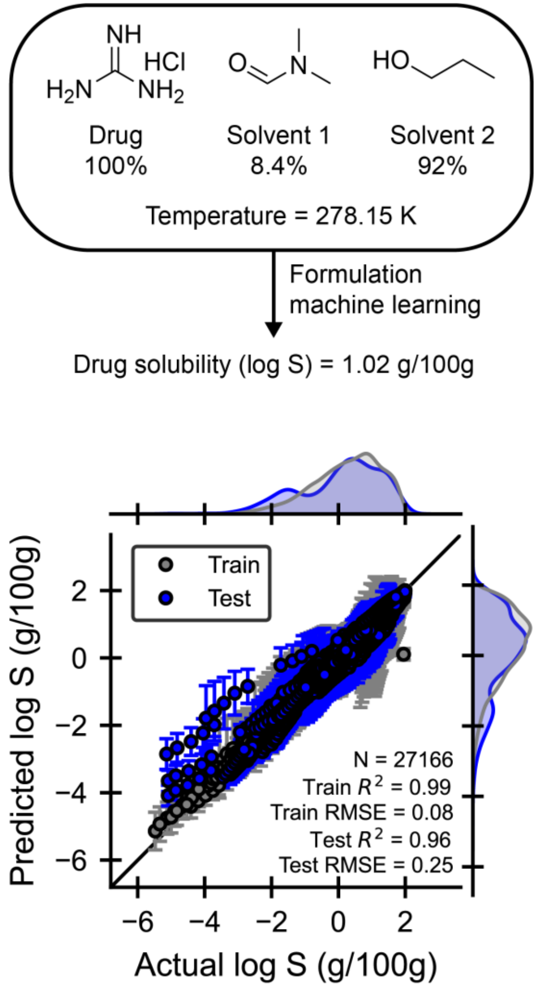
Inclusion of nitrogen groups lowers the enthalpy of mixing

Formulation ML predicts broad experimental properties

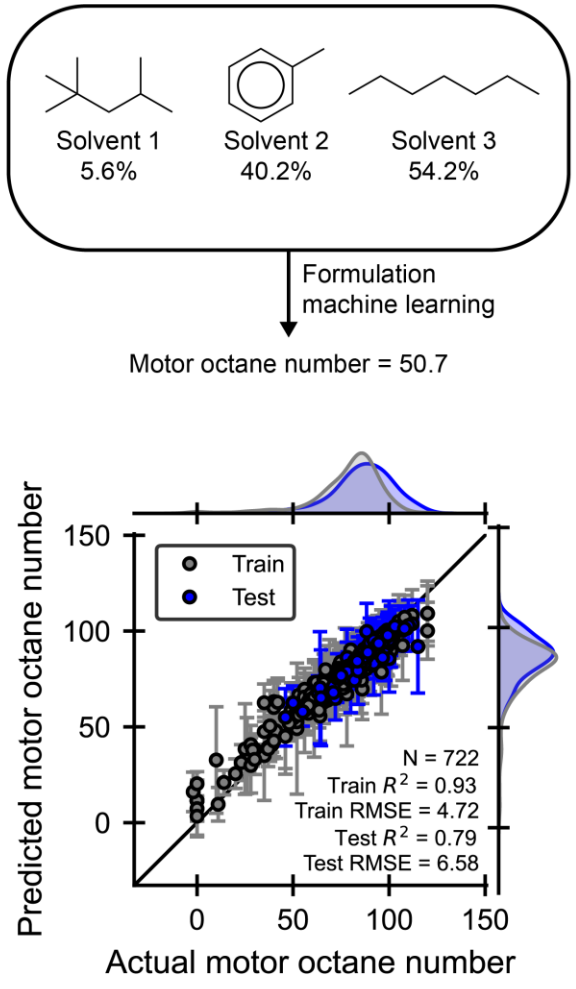
Energy Storage[1]



Pharmaceutical formulations[2]



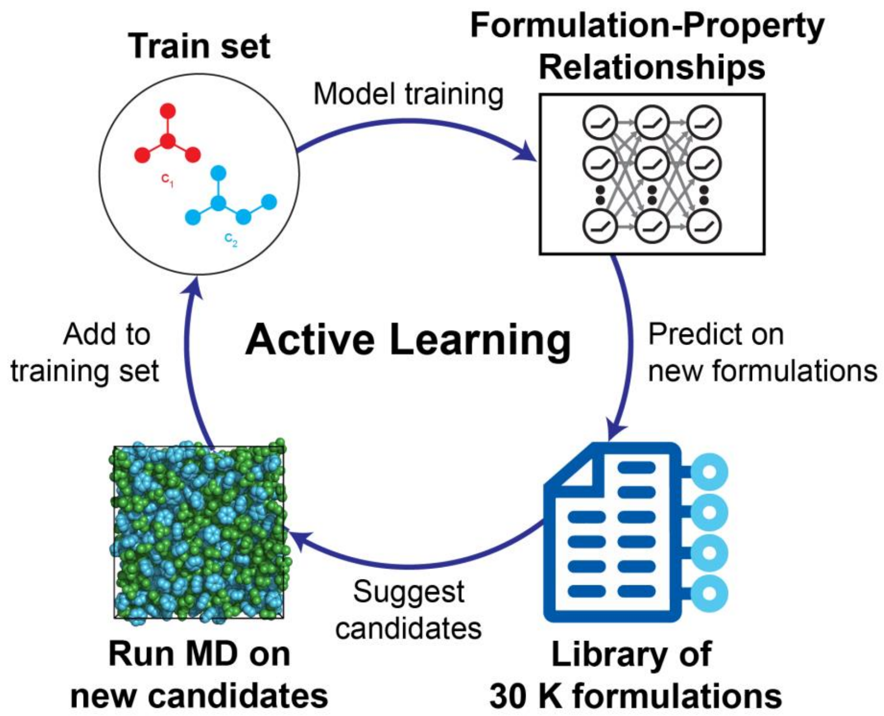
Oil and gas[3]



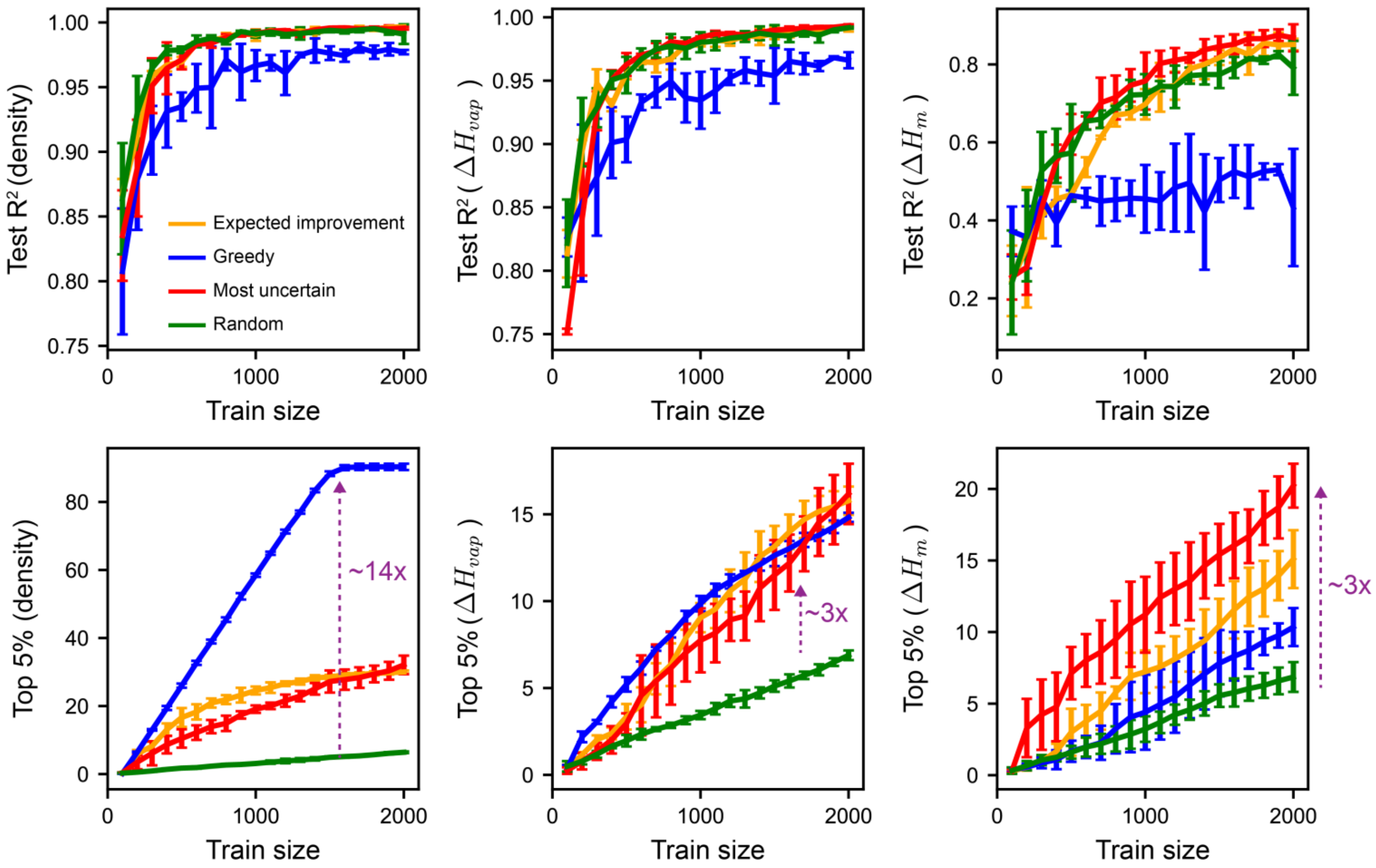
[1]Bilodeau, C., *et al. Chemical Engineering Journal*, 464, **2023**, 142454.
[2]Bao, Z., *et al. Journal of Cheminformatics*,16.1, **2024,** 117.
[3]Kuzhagaliyeva, N., *et al. Communications Chemistry*, 5.1, **2022**, 111.

Active Learning: Can ML suggest the best formulations?

Active learning workflow



ML identifies top formulations 2-3x faster than random trial-and-error



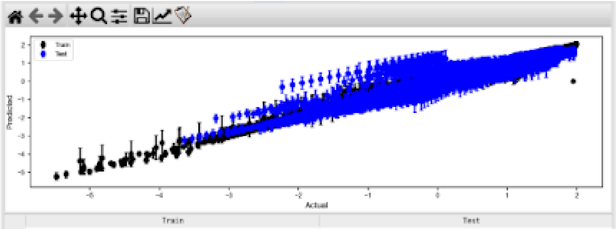
Formulation ML optimization

- Optimize materials formulations for a target property
- Automated model building and validation for a given formulations vs. property dataset

Input {SMILES | composition | property} dataset in CSV format

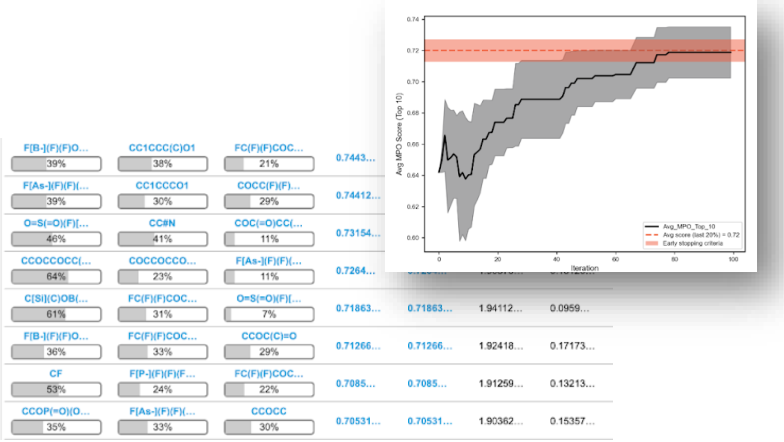
Temperature	Solubility (mol/mol)	DOI	SMILES_0	SMILES_1	SMILES_2	Drug_ExactM	Solvent_1_Ex	Solvent_2_Ex
278.15	0.0202	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.032	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.0471	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.0578	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.0697	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.0829	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.0985	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.1181	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
278.15	0.1464	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
280.65	0.0247	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647
280.65	0.0394	doi.org/10.1(	NC1=CC(N)=CC=C1	N#CC	O	108.068748	41.0265491	18.0105647

Automated model building & validation

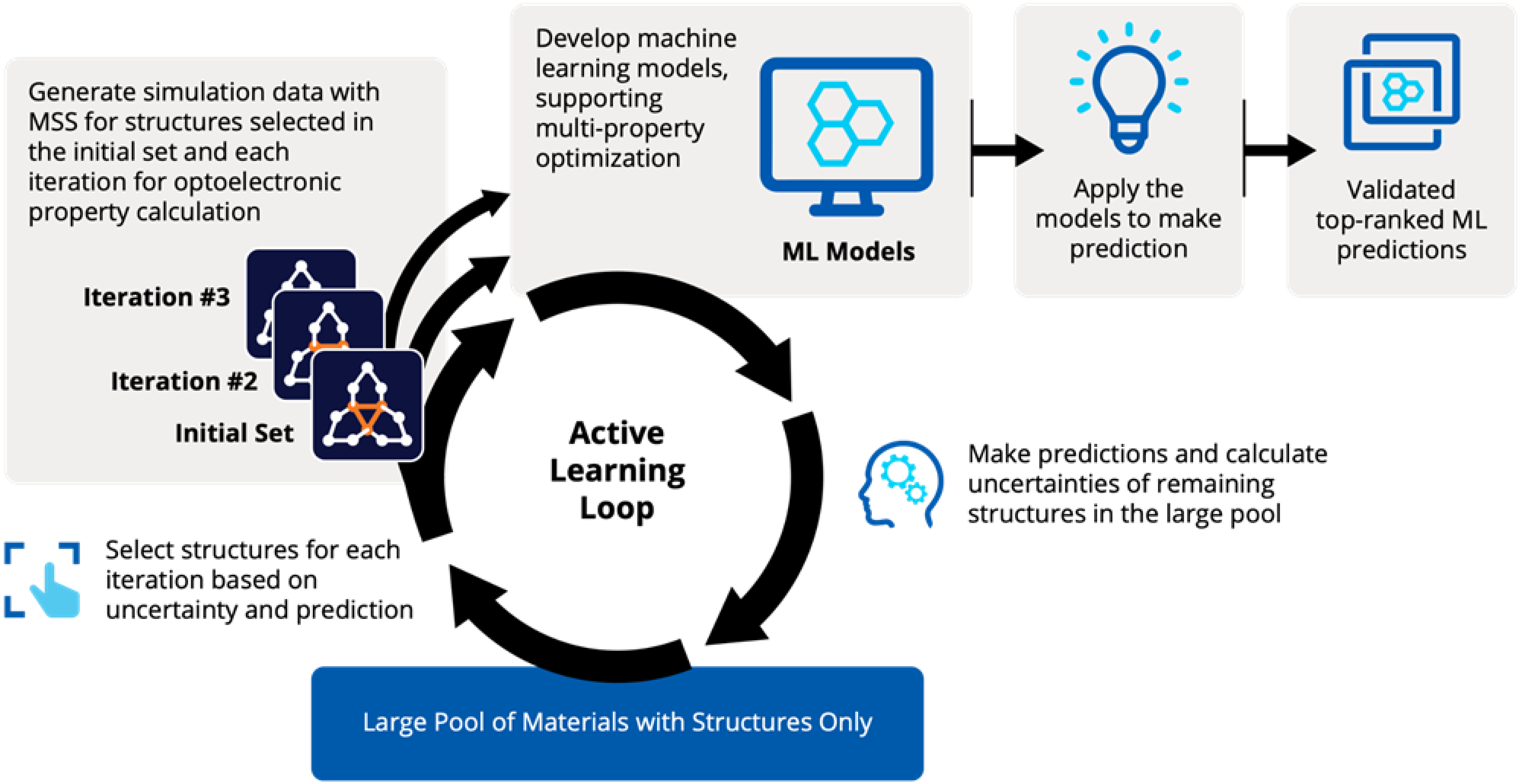


Assess target property space

Check optimized formulations with MPO scores



Supervised Learning: active learning



Case Study 2 – Optimizing shampoo formulations using Formulations ML

Optimizing shampoo formulations

Challenge

Complex physical interactions between ingredients making it difficult to tune formulations to customer-defined property targets

Solution

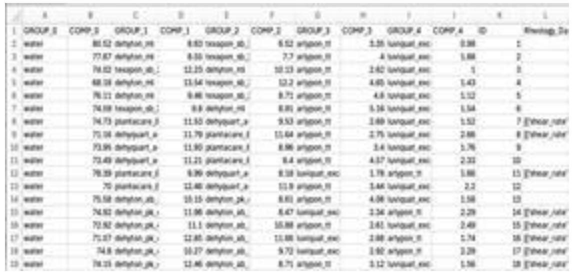
ML models with broad chemical space training and multiple properties can optimize formulations

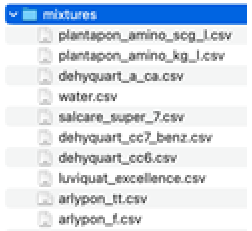
Results

Formulations ML models are able to accurately predict key performance properties & suggest new, optimized formulae

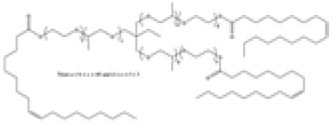
Impact

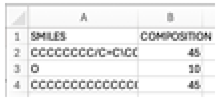
Accurate prediction across design space, encompassing key target properties of shampoo formulations



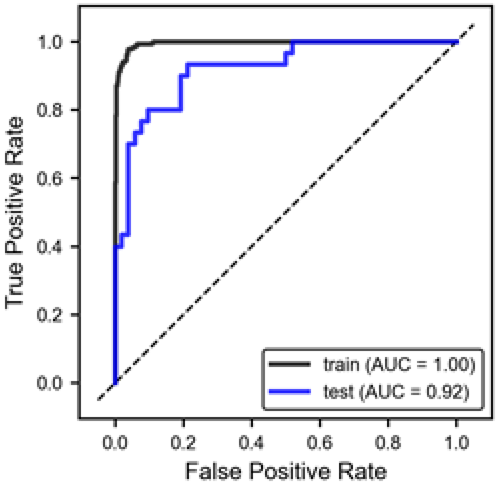


Arlypon TT

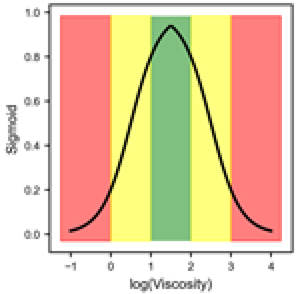


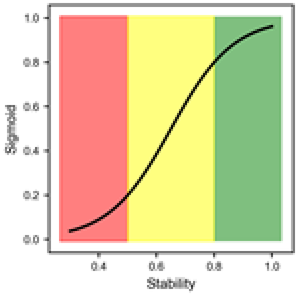


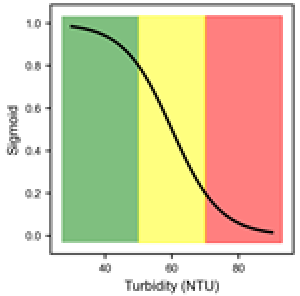
ML modeling



Formulation optimization







Schrödinger

BASF Shampoo formulations dataset

Data from the following source*


scientific data

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > scientific data > data descriptors > article

Data Descriptor | Open access | Published: 03 July 2024

Accelerating Formulation Design via Machine Learning: Generating a High-throughput Shampoo Formulations Dataset

Aniket Chitre, Robert C. M. Querimit, Simon D. Rihm, Dogancan Karan, Benchuan Zhu, Ke Wang, Long Wang, Kedar Hippalgaonkar & Alexei A. Lapkin

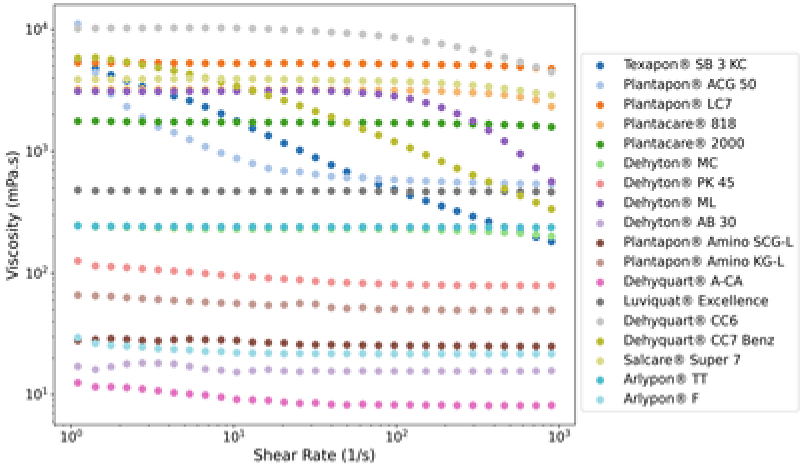
Scientific Data 11, Article number: 728 (2024) | Cite this article

3404 Accesses | 1 Altmetric | Metrics

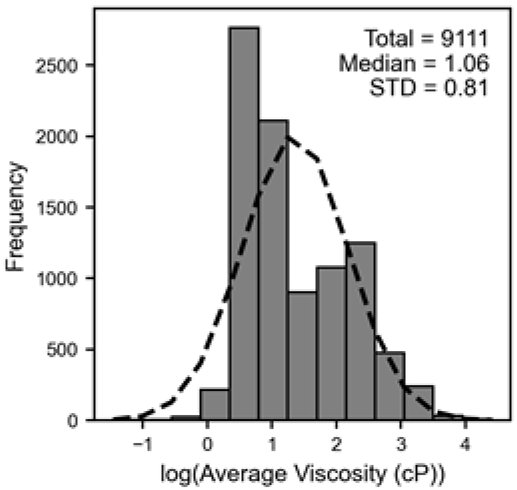
Dataset overview from paper:

- **812** unique formulations – mixtures of one **surfactant**, one **conditioning polymer**, and one **thickener**
  - Turbidity (regression) for stable formulations
  - Stability (classification)
- **9,633** shear-rate dependent viscosities

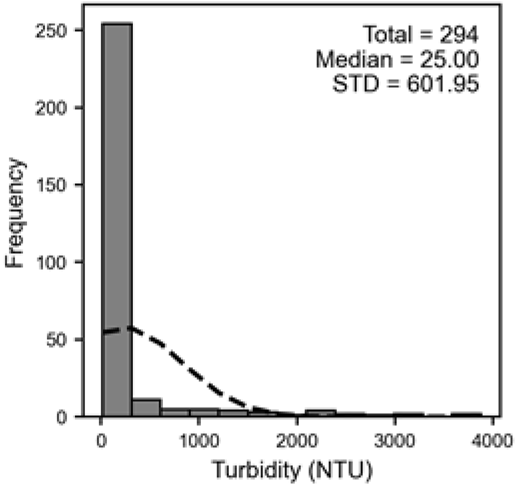
18 Ingredient Mixtures



Log-viscosity distribution



Turbidity distribution



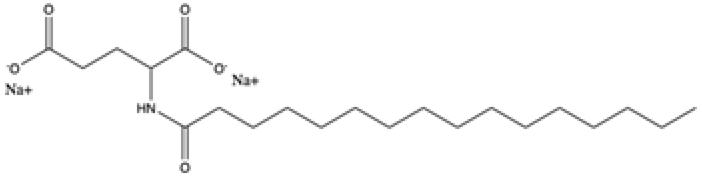
*Chitre, A., et al. Accelerating Formulation Design via Machine Learning: Generating a High-throughput Shampoo Formulations Dataset. Sci Data 11, 728 (2024). https://doi.org/10.1038/s41597-024-03573-w

Data conversion for complex ingredient mixtures

Formulation Definitions

	A	B	C	D	E	F	G	H	I	J	K	L
1	GROUP_0	COMP_0	GROUP_1	COMP_1	GROUP_2	COMP_2	GROUP_3	COMP_3	GROUP_4	COMP_4	ID	Rheology_Da
2	water	80.52	dehyton_ml	8.63	texapon_sb_:	6.52	arlypon_tt	3.35	luviquat_exc	0.98	1	
3	water	77.87	dehyton_ml	8.55	texapon_sb_:	7.7	arlypon_tt	4	luviquat_exc	1.88	2	
4	water	74.02	texapon_sb_:	12.23	dehyton_ml	10.13	arlypon_tt	2.62	luviquat_exc	1	3	
5	water	68.18	dehyton_ml	13.54	texapon_sb_:	12.2	arlypon_tt	4.65	luviquat_exc	1.43	4	
6	water	76.11	dehyton_ml	9.46	texapon_sb_:	8.71	arlypon_tt	4.6	luviquat_exc	1.12	5	
7	water	74.59	texapon_sb_:	9.8	dehyton_ml	8.91	arlypon_tt	5.16	luviquat_exc	1.54	6	
8	water	74.73	plantacare_8	11.53	dehyquart_a-	9.53	arlypon_tt	2.69	luviquat_exc	1.52	7	[{'shear_rate'
9	water	71.16	dehyquart_a-	11.79	plantacare_8	11.64	arlypon_tt	2.75	luviquat_exc	2.66	8	[{'shear_rate'
10	water	73.95	dehyquart_a-	11.93	plantacare_8	8.96	arlypon_tt	3.4	luviquat_exc	1.76	9	
11	water	73.49	dehyquart_a-	11.21	plantacare_8	8.4	arlypon_tt	4.57	luviquat_exc	2.33	10	
12	water	78.39	plantacare_8	9.99	dehyquart_a-	8.18	luviquat_exc	1.78	arlypon_tt	1.66	11	[{'shear_rate'
13	water	70	plantacare_8	12.46	dehyquart_a-	11.9	arlypon_tt	3.44	luviquat_exc	2.2	12	
14	water	75.58	dehyton_ab_:	10.15	dehyton_pk_4	8.61	arlypon_tt	4.08	luviquat_exc	1.58	13	
15	water	74.92	dehyton_pk_4	11.98	dehyton_ab_:	8.47	luviquat_exc	2.34	arlypon_tt	2.29	14	[{'shear_rate'
16	water	72.92	dehyton_pk_4	11.1	dehyton_ab_:	10.88	arlypon_tt	2.61	luviquat_exc	2.49	15	[{'shear_rate'
17	water	71.07	dehyton_pk_4	12.85	dehyton_ab_:	11.66	luviquat_exc	2.68	arlypon_tt	1.74	16	[{'shear_rate'
18	water	74.8	dehyton_pk_4	10.27	dehyton_ab_:	9.72	luviquat_exc	2.92	arlypon_tt	2.29	17	[{'shear_rate'
19	water	74.15	dehyton_pk_4	12.46	dehyton_ab_:	8.71	arlypon_tt	3.12	luviquat_exc	1.56	18	[{'shear_rate'

Plantapon Amino SCG-L



	A	B
1	SMILES	COMPOSITION
2	[O-]C(CCC(NC(C	27.5
3	O	72.5
4		

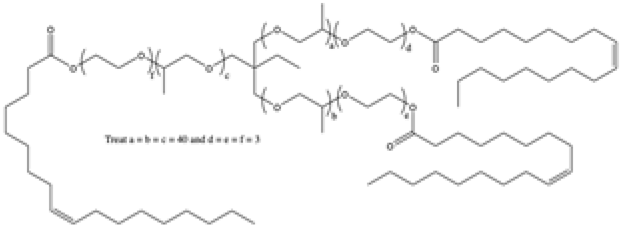
Ingredient Mixtures (to SMILES, COMPOSITION)

	A	B	C	D	E	F	G	H	I	J
1	trade_name	INCI	SMILES_0	COMP_0	SMILES_1	COMP_1	SMILES_2	COMP_2	SMILES_3	COMP_3
2	plantacare 8	Coco-Glucos	CCCCCCCC	52	O	48				
3	dehyton ab 3	Coco-Betain	CCCCCCCC	30	O	63	[Na+].[Cl-]	7		
4	texapon sb 3	Disodium Lau	CCCCCCCC	36	O	63	C/C=C/C=C/	0.5	OC(=O)CC(C	0.4
5	dehyton ml	Sodium Laur	CCCCCCCC	27	O	73				
6	dehyton mc	Sodium Coco	CCCCCCCC	30	O	70				
7	dehyquart a-	Potassium C	CCCCCCCC	25	O	75				
8	plantapon lc	Laureth-7 Cit	OC(CC(O)(C	92.5	O	0.5	OC(=O)CC(C	7		
9	arlypon tt		CCCCCCCC	45	O	10	CCCCCCCC	45		
10	luviquat exce	Polyquaterni	[At]C(N1CCC	2	O	60	[At]C(N1C=[	38		
11	plantapon an	Potassium C	CCCCCCCC	32.5	O	67.5				
12	dehyquart cc	Polyquaterni	[At]CC1C[N+	100						
13	salcare supe	Polyquaterni	[At]CC1C[N+	25	[At]CC[At](C	75				
14	plantapon ac	Disodium Co	CCCCCCCC	45	O	44	CC(O)CO	6		
15	arlypon f	Laureth-2	CCCCCCCC	90.5	O	9.5				
16	plantapon an	Sodium Coco	[O-]C(CCC(N	27.5	O	72.5				
17	dehyton pk 4!	Cocamidopr	CCCCCCCC	45	O	55				
18	plantacare 2	Decyl-Gluco	CCCCCCCC	47	O	53				

→

- mixtures
  - plantapon_amino_scg_l.csv
  - plantapon_amino_kg_l.csv
  - dehyquart_a_ca.csv
  - water.csv
  - salcare_super_7.csv
  - dehyquart_cc7_benz.csv
  - dehyquart_cc6.csv
  - luviquat_excellence.csv
  - arlypon_tt.csv
  - arlypon_f.csv

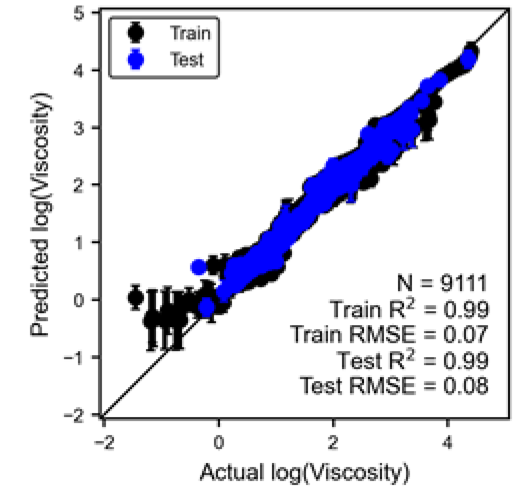
Arlypon TT

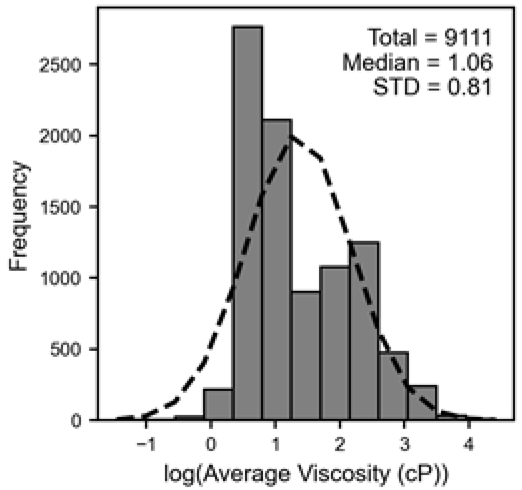


	A	B
1	SMILES	COMPOSITION
2	CCCCCCCC/C=C\CC	45
3	O	10
4	CCCCCCCCCCCCC(	45

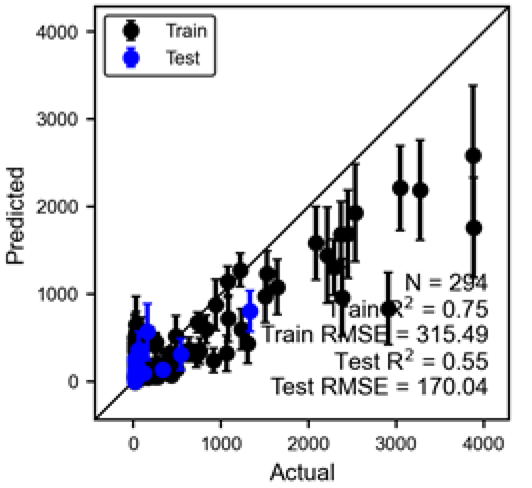
Evaluating multiple property models

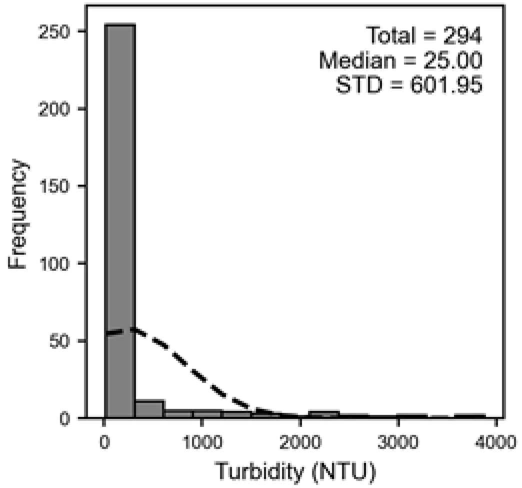
Viscosity (Regression)



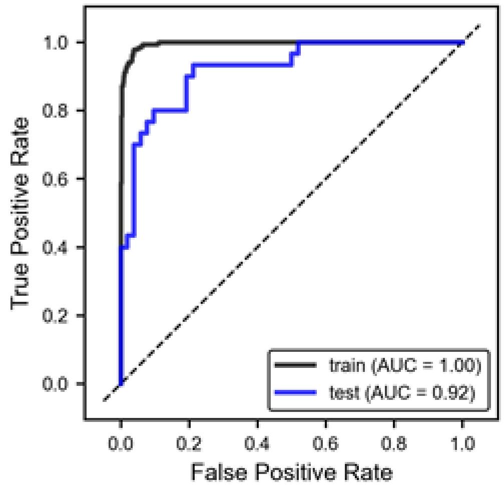


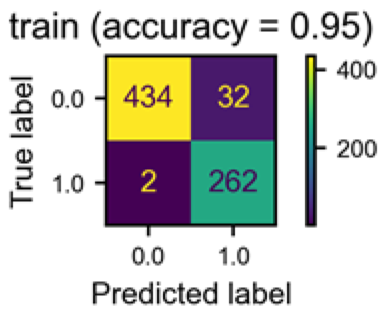
Turbidity (Regression)

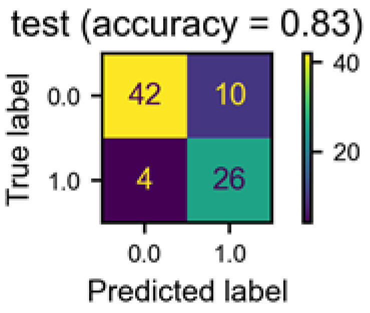




Stability (Classification)



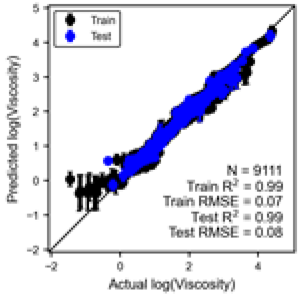




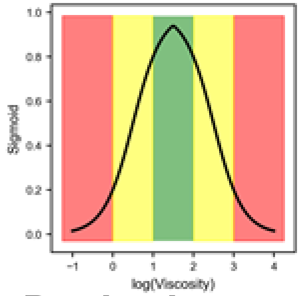
Schrödinger

Formulation optimization for shampoo

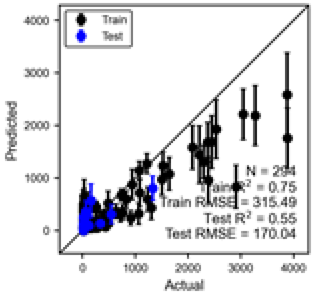
Viscosity



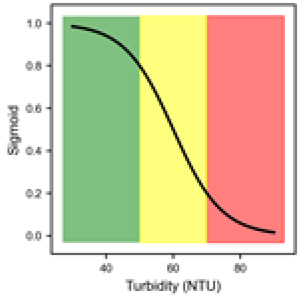
Desired target:
Middle is better



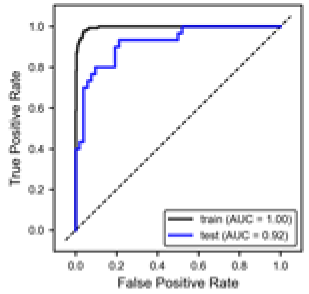
Turbidity



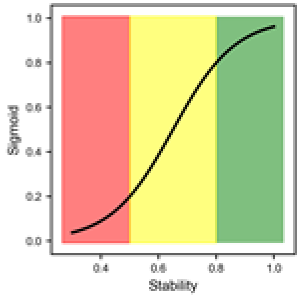
Desired target:
Lower is Better



Stability



Desired target:
Higher is Better



Goal:
Optimize composition for complex mixtures

Surfactant
8–13 w/w%

Conditioning Polymer
1–3 w/w%

Thickener
1–5 w/w%

Ingredient Choices

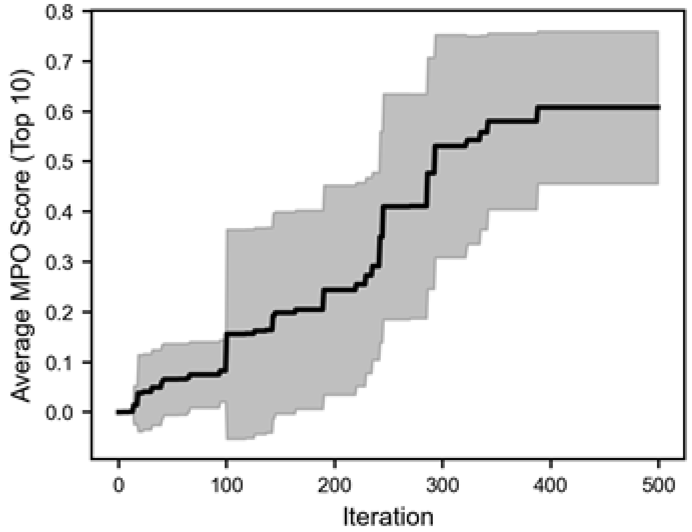
	A	B	C	D	E	F	G
1	SMILES	MIN_COMP	MAX_COMP	LABEL	GROUP	REQUIRED	STEP_SIZE
2	water	0.7	0.9	18	water	TRUE	0.01
3	arlypon_f	0.01	0.05	0	thickener	FALSE	0.01
4	arlypon_tt	0.01	0.05	1	thickener	FALSE	0.01
5	dehyquart_a-ca	0.08	0.13	2	surfactant	FALSE	0.01
6	dehyton_ab_30	0.08	0.13	5	surfactant	FALSE	0.01
7	dehyton_mc	0.08	0.13	6	surfactant	FALSE	0.01
8	dehyton_ml	0.08	0.13	7	surfactant	FALSE	0.01
9	dehyton_pk_45	0.08	0.13	8	surfactant	FALSE	0.01
10	plantacare_2000	0.08	0.13	10	surfactant	FALSE	0.01
11	plantacare_818	0.08	0.13	11	surfactant	FALSE	0.01
12	plantapon_acg_5C	0.08	0.13	12	surfactant	FALSE	0.01
13	plantapon_amino_	0.08	0.13	13	surfactant	FALSE	0.01
14	plantapon_amino_	0.08	0.13	14	surfactant	FALSE	0.01
15	plantapon_lc_7	0.08	0.13	15	surfactant	FALSE	0.01
16	texapon_sb_3_kc	0.08	0.13	17	surfactant	FALSE	0.01
17	dehyquart_cc6	0.01	0.03	3	conditioning	FALSE	0.01
18	dehyquart_cc7_be	0.01	0.03	4	conditioning	FALSE	0.01
19	luviquat_excellen	0.01	0.03	9	conditioning	FALSE	0.01
20	salcare_super_7	0.01	0.03	16	conditioning	FALSE	0.01

Choose one of each ingredient type, remainder of the formulation is water to sum up to 100%

Optimization results and best candidate(s)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	iteration	MPO	log(Viscosity)_predict	Turbidity (NTU)_predict	Stability_predict	INGREDIENT_0	INGREDIENT_1	INGREDIENT_2	INGREDIENT_3	comp_0	comp_1	comp_2	comp_3
2	286	0.798	1.273	55.152	0.732	water	arlypon_f	dehyton_mc	dehyquart_cc7_benz	0.89	0.01	0.09	0.01
3	100	0.758	0.719	27.304	0.929	water	arlypon_tt	plantacare_818	dehyquart_cc7_benz	0.87	0.01	0.11	0.01
4	293	0.737	1.153	63.553	0.895	water	arlypon_f	dehyton_pk_45	salcare_super_7	0.86	0.01	0.12	0.01
5	245	0.734	0.675	34.031	0.923	water	arlypon_f	dehyton_ab_30	salcare_super_7	0.86	0.01	0.12	0.01
6	242	0.692	1.025	59.823	0.689	water	arlypon_tt	plantapon_acg_50	dehyquart_cc7_benz	0.8	0.03	0.15	0.02
7	388	0.592	0.554	27.743	0.638	water	arlypon_f	plantacare_818	luviquat_excellence	0.88	0.01	0.1	0.01
8	342	0.531	0.795	44.228	0.455	water	arlypon_tt	plantacare_2000	luviquat_excellence	0.82	0.02	0.15	0.01
9	190	0.476	1.213	39.343	0.301	water	arlypon_tt	plantacare_2000	luviquat_excellence	0.85	0.02	0.12	0.01
10	335	0.442	0.623	55.201	0.442	water	arlypon_f	plantapon_amino_	dehyquart_cc7_benz	0.9	0.02	0.07	0.01
11	143	0.322	1.424	93.728	0.635	water	arlypon_f	dehyton_mc	dehyquart_cc7_benz	0.89	0.02	0.08	0.01
12	395	0.320	1.451	52.708	0.090	water	arlypon_f	dehyton_pk_45	luviquat_excellence	0.88	0.01	0.08	0.03
13	323	0.315	1.795	52.660	0.103	water	arlypon_tt	dehyton_pk_45	dehyquart_cc7_benz	0.89	0.02	0.08	0.01

MPO Scores



1% thickener

ARLYPON® F

Function

Nonionic Surfactant

Thickener

9% surfactant

Dehyton® MC

Amphoacetate: Amphoteric Surfactant

Predicted Properties:

Viscosity 1.273 log(cP)

Turbidity 55 NTU

Stability 73%

1% conditioning polymer

Dehyquart® CC7 BZ

Aqueous Solution of a cationic Diallyl Dimethyl Ammonium Chloride/Acrylamide Copolymer

Success stories in real world applications

Fast adoption of digital chemistry brings big business impact for consumer packaged goods R&D

CHALLENGE

The consumer packaged goods market faces many challenges, including demands for sustainability, constant requirement changes and short time-to-market timelines.

SOLUTION

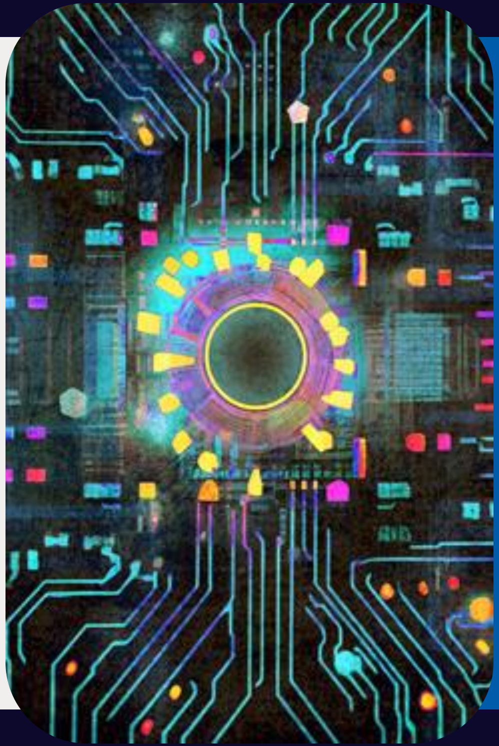
The Reckitt team incorporated digital simulation in several R&D areas (e.g. detergents, drug formulations, packaging materials). Running digital testing before experiment reduced mistakes and improved outcomes.

Result: R&D timelines expedited by **10 times**

Large-scale de novo design of hole-conducting materials for organic electronics

CHALLENGE

Molecules with high mobility are highly desirable for organic electronics. However, it is extremely costly and time-consuming to synthesize and assess every candidate molecule



SOLUTION

Scientists from Panasonic and Schrödinger employed DFT, machine learning and cloud computing to screen over 14 million molecules, predicting hole mobility of the selected top candidates

Result: Over 50 molecules with better performance were identified and the structural effects were discovered

Discussion and questions

Schrödinger

Thank you!