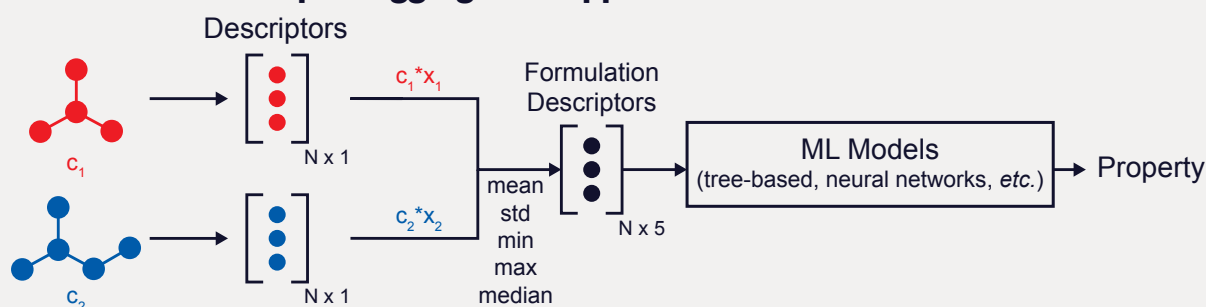


Advanced Machine Learning and Molecular Simulations for Formulation Design

Complex chemical mixtures — or formulations — are used in a wide range of applications, such as gasoline blends in oil & gas, daily care products in consumer goods, and drug delivery in pharmaceuticals. Given the vast number of potential formulations, evolving regulatory requirements, and increasing consumer demand for eco-friendly and sustainable products, we need innovative and cost-effective solutions for designing enhanced formulations. The latest advancements in atomic-scale modeling and machine learning (ML) have enabled computer-aided screening of large numbers of formulation candidates — thus, accelerating the identification of promising formulations and reducing costly experiments.

Schrödinger's **Formulation Machine Learning** tool uses data-driven methods to correlate ingredient structure and composition to formulation properties. This tool uses advanced cheminformatics descriptors and automatic hyperparameter tuning to find the best ML model, and allows external features (e.g., temperature, pressure) from experiments or high-throughput molecular dynamics (MD) calculations to be used as additional input to the ML model. The Formulation ML tool enables R&D teams to quickly train and deploy ML models to rapidly explore the broad design space of formulations by varying the chemical ingredients, compositions, and external features.

Formulation Descriptor Aggregation Approach¹



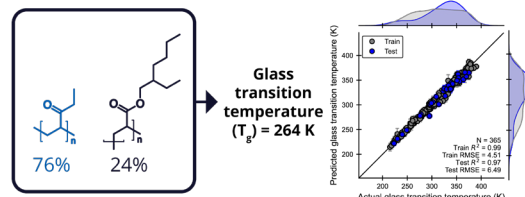
Advantages of Schrödinger Formulation Screening Technology

- **Efficient ML model building and data generation:** Leveraging deep learning technology to build accurate ML models to predict formulation properties, which can be coupled with MD simulations as a way to generate physically meaningful descriptors to improve ML model accuracy
- **Scalable:** ML can be trained and evaluated for mixtures with more than 100 components, extending the capabilities beyond simple mixtures to designing complex mixtures with enhanced properties
- **Automated:** Automatic hyperparameter tuning enables accurate ML model development using expert cheminformatic descriptors with minimal ML expertise required
- **Rapid screening capabilities:** ML can generate predictions in a fraction of a second, which can scale up to screening ~100K formulations in the order of minutes-hours
- **Dedicated support:** Dedicated support team consisting of scientific experts at Schrödinger are available to help users apply computational tools to their applications
- **Multiple platform functionality:** Can be used on laptops, desktops, and high performance clusters

Applications Across Industries

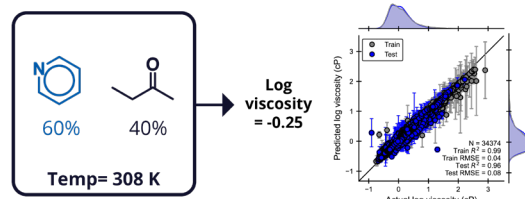
Consumer Products

Random copolymer systems are often found in packaging materials, and glass transition temperature (T_g) is an important parameter that dictates the stability of the polymer as a function of temperature. Formulation ML can accurately predict T_g for 365 examples with a test set coefficient of determination (R^2) of 0.97.²



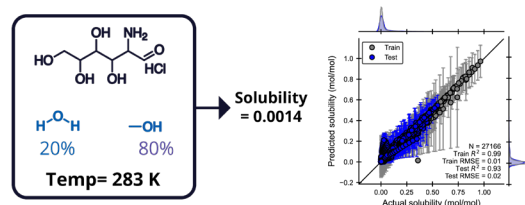
Energy Storage

Liquid electrolytes are often used in batteries to facilitate the movement of electrical charge between an anode and cathode, and viscosity is an important parameter that dictates how easy ions can move through an electrolyte solution. Formulation ML can accurately predict temperature-dependent viscosity given ~34K examples with a test set R^2 of 0.96.³



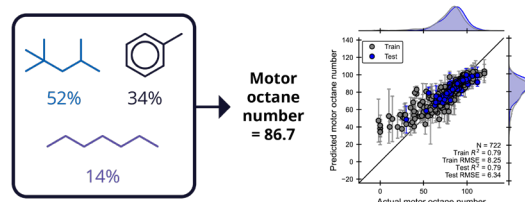
Pharmaceutical Formulation

Solubility of drug molecules in pure and binary mixture solutions is crucial for drug delivery applications for pharmaceutical formulations. Formulation ML can accurately predict temperature-dependent drug solubility for either pure or binary mixture solutions given ~27K examples, which achieves a test set R^2 of 0.93.⁴



Oil and Gas

Mixtures of hydrocarbons are critical in gasoline blends, facilitating efficient combustion for automotive engines, and motor octane number (MON) is an important parameter that measures the fuel behavior under external pressure. Formulation ML can accurately predict MON given ~700 examples with the number of components ranging from pure (single) component systems to 120 components, which achieves a test set R^2 of 0.79.⁵



References

- Leveraging high-throughput molecular simulations and machine learning for the design of chemical mixtures**
Alex, C., et al. *npj Comput Mater* 11, 72, 2025, <https://doi.org/10.1038/s41524-025-01552-2>.
- The glass transition temperature of random copolymers: 1. Experimental data and the Gordon-Taylor equation**
Penzel, E., et al. *Polymer*, 38.2, 1997, 325-337.
- Machine learning for predicting the viscosity of binary liquid mixtures**
Bilodeau, C., et al. *Chemical Engineering Journal*, 464, 2023, 142454.
- Towards the prediction of drug solubility in binary solvent mixtures at various temperatures using machine learning**
Bao, Z., et al. *J Cheminform* 16, 117, 2024, <https://doi.org/10.1186/s13321-024-00911-3>.
- Artificial intelligence-driven design of fuel mixtures**
Kuzhagaliyeva, N., et al. *Communications Chemistry*, 5.1, 2022, 111.

Contact us: ms-sales@schrodinger.com
Learn more: www.schrodinger.com/materials-science

 **Schrödinger**

Copyright © 2025 Schrödinger, Inc.

Speak with
an expert to
get started

