CITRINE INFORMATICS

White paper

UNIQUE CHALLENGES IN AI FOR FORMULATIONS DEVELOPMENT

Unique Challenges in AI for Formulations Development

EXECUTIVE SUMMARY

The development of formulations (processed mixtures of materials and chemicals), such as paints, solvents, and polymer compounds, has unique challenges associated with the vast spectrum of possible ingredient combinations and the complex rules that govern their mixing. Added to this are tight deadlines for development to win contracts and the need to factor in business constraints such as cost of ingredients and regulatory concerns.

Citrine Informatics has worked with many leading formulation developers over the last 6 years and has developed tools to:

- generate and review design spaces with complex constraints and millions of options
- convert chemical objects and recipe structure into machine-readable data
- handle complex process flows

By conquering these challenges, Citrine has unlocked the potential of AI for specialty chemical companies and formulations developers like Showa Denko, and Lanxess. Of particular interest for these companies is the ability to quickly respond to customer requests, identify and mitigate supply risks, and rationalize their portfolios of ingredients and products.

Citrine Informatics continues to develop tools and frameworks specifically for this market, using the knowledge gained from working with cutting-edge chemical companies.

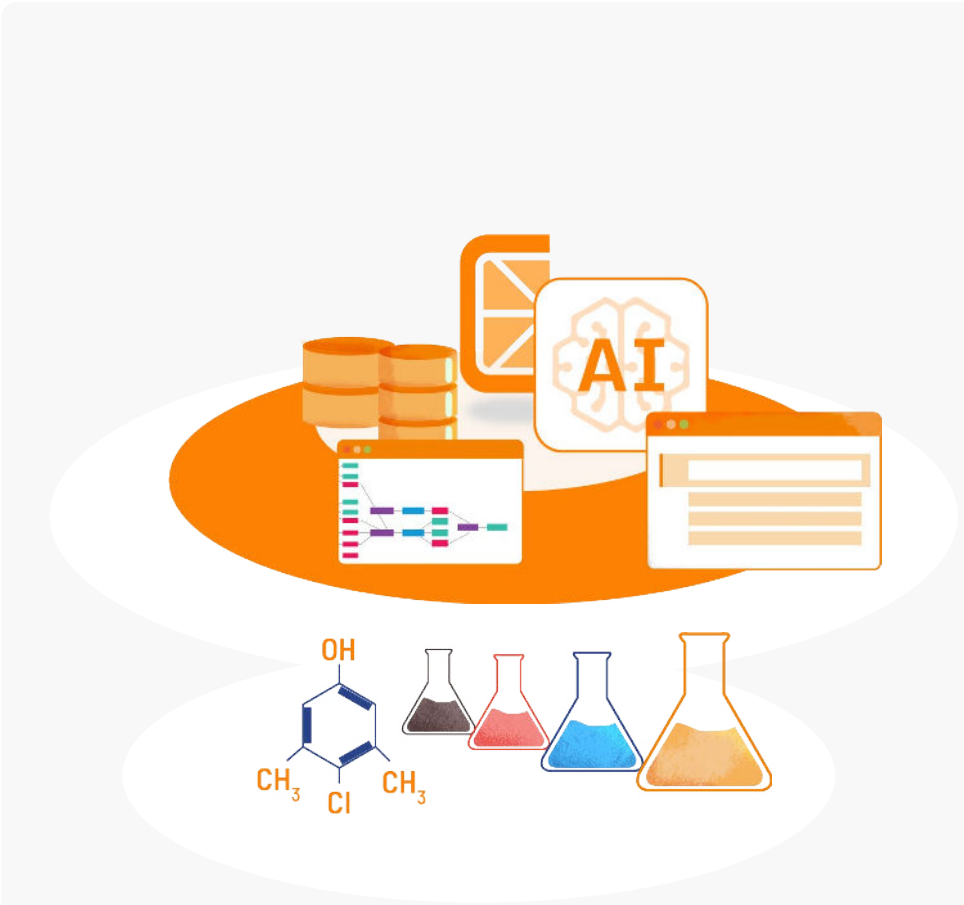
Introduction

INTRODUCTION

Formulations are defined as homogeneous mixtures of chemicals and/or materials. They can be mixtures of powders and solvents like paints or polymers, glasses, and other ingredients to make polymer composites. They can be simple blends of unreactive chemicals, or reactive mixtures where the order of mixing makes a difference, or blends of blends where the final molar composition is dependent on the composition of the original blends. All of this means that there is a huge number of possible ingredients that have complex sets of constraints surrounding their type, number, molar %, and order of mixing.

New formulations development is typically done in response to a customer request involving multiple ingredient and process constraints and property objectives. The difference between winning and losing large contracts hinges on quickly understanding what you already have available and what can be quickly adapted or developed to suit the new business opportunities.

Procuring, storing, and processing thousands of different ingredients is costly, with obvious savings to be had from rationalization. Costs of ingredients vary, and formulations hitting target properties using lower cost ingredients improve margins. As global events impact supply chains, the ability to quickly reformulate using ingredients from a new supplier is essential. It is also necessary to be able to adapt as new legislation on restricted substances come into force.



CHALLENGE 1
GENERATING COMPLEX DESIGN SPACES

EXAMPLE 1

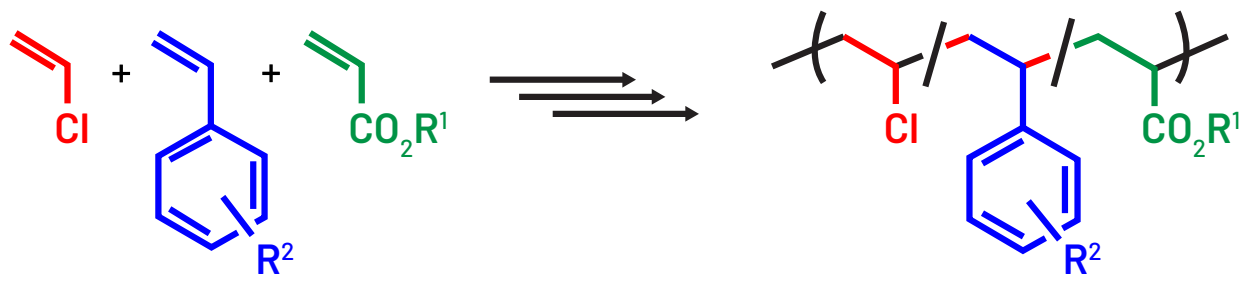
Showa Denko wanted to discover solvent blends that hit specific property targets for their customer. They gave Citrine a list of 380 solvents from their inventory that could be in the blend.

For an equal mix of 4 solvents, there were 855 million possible combinations; however, Showa Denko articulated additional constraints that narrowed this list down to around 100 million.

EXAMPLE 2

A customer wanted to find polymers with improved physical properties. They gave Citrine a list of monomers of interest which could be combined in any permutation. Citrine then programmatically generated a list of thousands of polymers to explore.

The Citrine platform can generate design spaces using the customer's domain knowledge to create a set of rules



CHALLENGE 2
CONVERTING CHEMICAL INFORMATION AND RECIPE STRUCTURE TO MACHINE-READABLE DATA

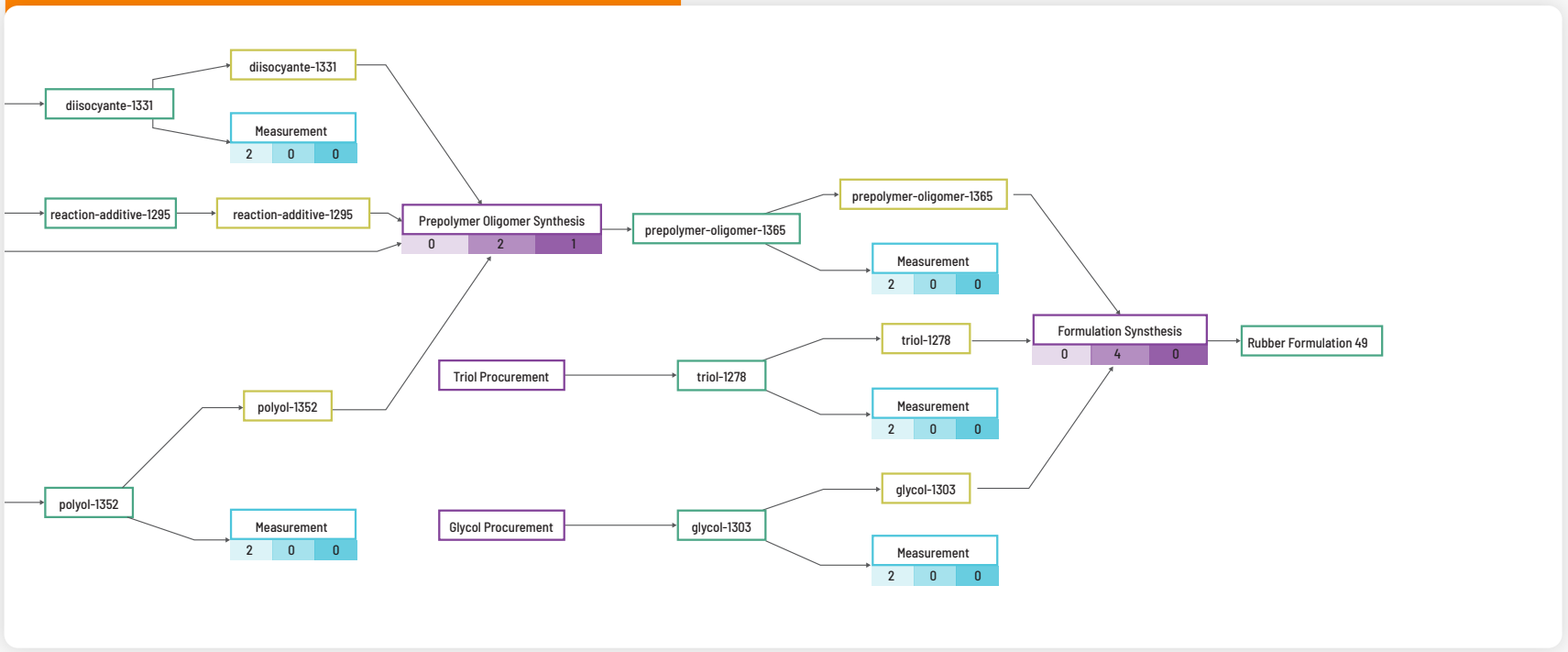
Off-the-shelf AI algorithms read specialized data sets such as chemical formulas and x-ray diffraction images as strings of letters and numbers or strange pictures. The Citrine Platform has a library of descriptors developed specifically for materials and chemicals data that converts information (e.g. SMILES specification) into related machine-readable data (e.g. Atomic Polarizability).

Where customer teams can identify key parameters and processing conditions that are directly related to the target material properties, they are pulled out in a simplified table that can be used in data analysis and modelling. The benefit of this is that data management teams can focus on acquiring detailed data for these key parameters, and product developers can focus their work to understand the underlying mechanisms affecting properties.

CHALLENGE 3
WORKING ON PROCESS FLOW

In product classes where there are complex process flows, the order in which process steps occur and ingredients are introduced is key to resulting properties. The Citrine Platform stores a formulations process history from ingredient procurement through processing, to characterization of the final product using a data model called GEMD (Graphical Expression of Materials Data). This model captures and visualizes process flows which can be extracted and used in AI models.

The Citrine platform captures complex process flows



EXAMPLE

The Citrine team was able to "featurize" the material history data of a semi-crystalline polymer manufacturing process into a simplified structure that could be systematically assessed for its contribution to hard segments in the polymer.

CHALLENGE 4
MAKING THE BEST USE OF SIMULATION

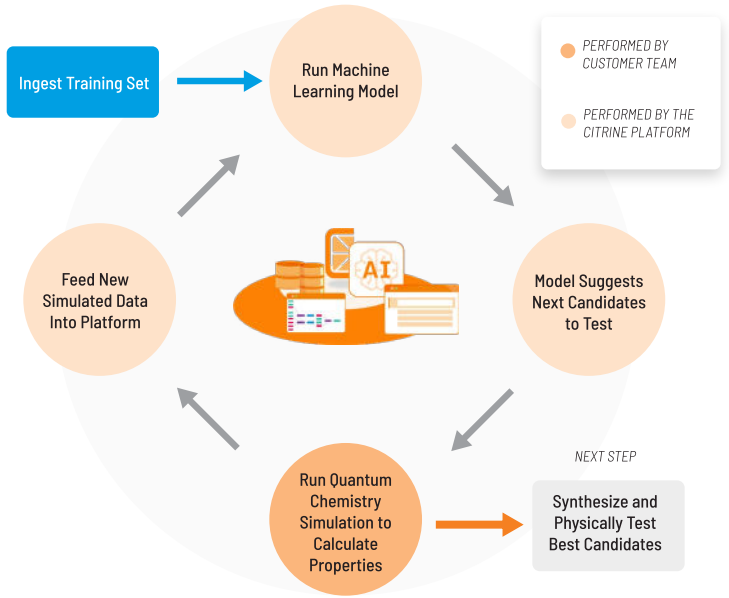
Simulation, including quantum chemistry simulation, can be used to calculate the properties of molecules and formulations. While less resource intensive than trial and error lab experiments, calculations can still take hours and days. To streamline the use of simulation, the Citrine Platform facilitates a Sequential Learning workflow. Sequential Learning is the next generation of design of experiment. It is an iterative process that combines simulation or experimentation with uncertainty calculations to enable efficient, methodical exploration of design spaces.

EXAMPLE 1

Showa Denko's simulation team used the sequential learning process to reduce the number of calculations and find optimum solvent blend, faster.

Sequential learning workflow

CITRINE PLATFORM



EXAMPLE 2

Panasonic also used this functionality when finding new organic semiconductors. They performed physics-based modelling on candidate molecules suggested by Citrine.

"This work demonstrates the utility of using the sequential learning methodology to design experiments for the discovery of novel materials."

Dr. Nobuyuki Matsuzawa

Panasonic

CHALLENGE 5
RESPONDING TO CUSTOMERS QUICKLY

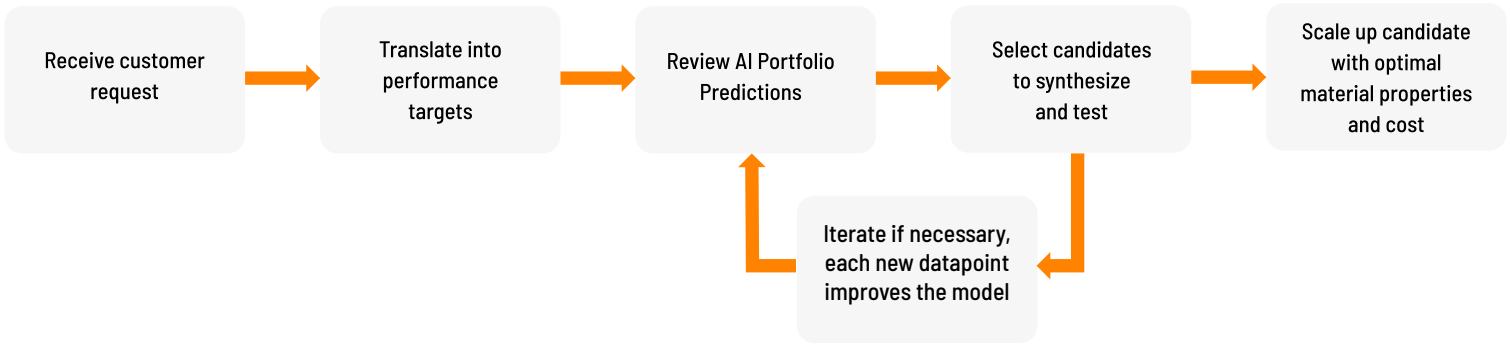
Winning contracts depends on a timely and strategic response to customer requests. For this, data is key. The quickest and easiest response is to deliver a product you already have; however, oftentimes data on existing products can be patchy, with not every property of the product already measured. The Citrine Platform ensures that all your materials and chemicals data is accessible from one place. AI models can then be used to predict the values of missing properties, filling gaps and increasing the likelihood that you find a suitable existing product before conducting costly new product development.

In cases where an existing product is not suitable, the Citrine Platform can be used to quickly identify candidate formulation recipes that are likely to meet requirements, or confirm that the customer requirements are unlikely to be achieved with current technologies.

Filling in data gaps

	Young's Modulus	Izod Impact	Fracture Toughness	Tensile Strength	Cost
Existing Product 1	3GPa	330J/m^2	3MPa/m$^{1/2}$	46MPa	5/Kg
Existing Product 2	?	380J/m^2	3MPa/m$^{1/2}$	50MPa	?$/Kg
Potential Product	?	?	?	?	?$/Kg

AI-enabled customer workflow



CHALLENGE 6
RATIONALIZING INGREDIENTS AND PRODUCTS

Formulations companies have hundreds, if not thousands, of ingredients in inventory. Rationalizing the ingredients in stock and bulk buying fewer ingredients reduces costs. The Citrine Platform gives product developers the data that they need to ensure that any additional new ingredients are justified, helping them to understand if substituting one ingredient for another would significantly affect the performance of a formulation. This stops the inventory from growing unnecessarily. They can use the platform to develop reformulations with ingredient substitutions when necessary.

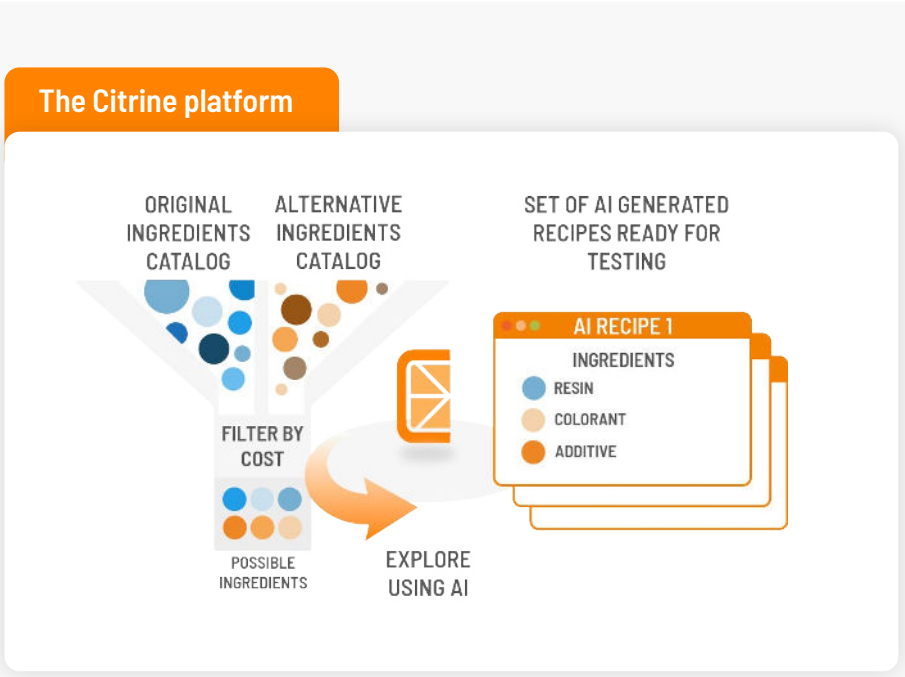
With complete, visual datasets for each product (see Challenge 5), formulators can evaluate whether fewer products can meet customer requirements and rationalize product inventory.

CHALLENGE 7
REFORMULATING WITH RESTRICTED INGREDIENTS

Ingredients can be restricted suddenly, by new regulations, supply chain issues, and sustainability goals. In each case, an agile development process is needed to quickly reformulate while maintaining product performance. By using the Citrine Platform to accelerate product development, companies are able to adapt to change.

EXAMPLE 1

A polymer manufacturer wanted to broaden the pool of available ingredients to include ingredients from both their European- and US-based operations. Using the Citrine Platform they filtered this super set of ingredients by cost, before reformulating their products.

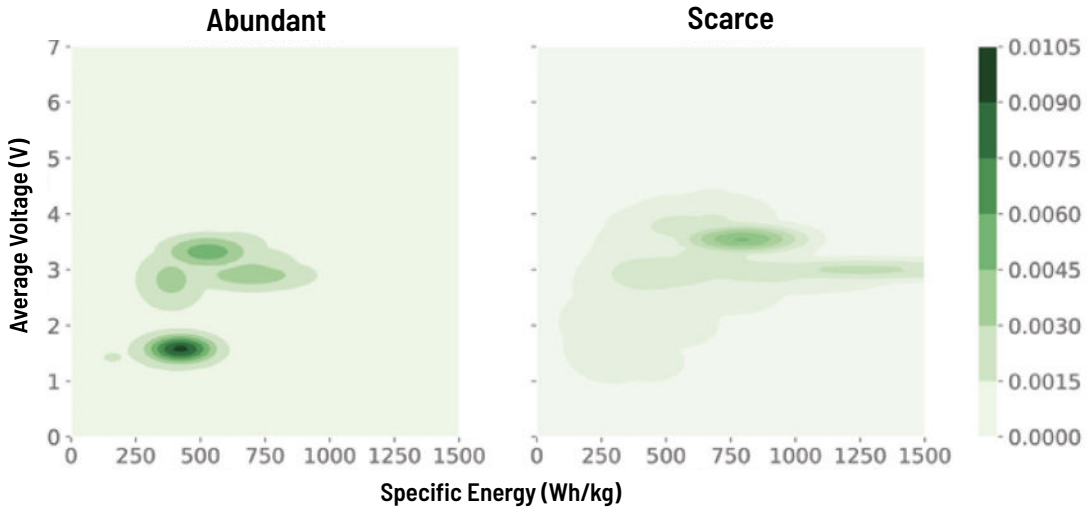


REFORMULATING WITH RESTRICTED INGREDIENTS

EXAMPLE 2

As biodegradability of organic solvents becomes more important, it is useful to understand if using biodegradable molecules affects their performance in this case represented by boiling point and relative polarity. The two charts below plot the aggregated predicted performance of all possible solvents in the design space (using Citrine patented technology), filtered by the molecules being readily biodegradable (left) and not readily biodegradable (right). The Citrine Platform explicitly displays the uncertainty of each prediction it makes. The property predictions displayed on these charts are therefore not exact values but overlapping probability gradients. These visualizations can be used to inform strategic research direction decisions. In this case, product developers can quickly see that restricting the ingredients to biodegradable molecules has little effect on the properties of interest. It is therefore possible to reformulate in a way that is more sustainable, without compromising on performance.

Visualization indicating relative likelihood of achieving properties



SUMMARY

The Citrine Platform has been built to enable companies to get the most out of Materials Informatics. Its design is based on the experience of Citrine as it has engaged with world-leading chemicals and materials companies, and universities and research organizations over the past 7 years. By not just concentrating on AI algorithms and data management, but paying due attention to the user experience, Citrine has made a platform that can be used autonomously across a global enterprise. The talents of both data scientists and scientific researchers can be harnessed to accelerate materials and chemicals development.

Citrine s mature approach is already reaping benefits for innovative players in the materials and chemicals industry. Contact Citrine to see a demonstration of the platform.

Citrine Informatics Inc.
2629 Broadway St
Redwood City, CA 94063

citrine.io
info@citrine.io