CLEANING UP SHAMPOO R&D
AI-DRIVEN FORMULATION DESIGN WITH BIOVIA PIPELINE PILOT

Application Brief

INTRODUCTION

Maintaining a healthy head of hair is a daily challenge almost everyone faces, and for many a good hair day is a point of pride. Often, they have to "earn" this pride, managing collections of shampoos, conditioners, hair masks, serums, gels, mousses, pomades, sprays and more. Yet at the core of any of these routines lies shampoo. It cleanses hair of dirt and unnecessary oils, priming hair for treatment. As a result, for some consumers shampoo is more than just soap; it is an experience. In addition to effectively cleaning hair, products should have a rich lather, possess pleasant scents, and leave the hair feeling light, soft and strong. However, no two heads of hair are alike: straight vs. curly, thin vs. coarse, colored vs. uncolored, dry vs. oily, short vs. long. Serving such a diverse customer base provides unique challenges for personal care companies, especially as they continue tailoring products towards individual consumers. Something that moisturizes and tames coarse, curly hair can weigh down light, thin hair, and a shampoo that effectively cleans oily hair can strip dyed hair of its colors. Each shampoo needs to be formulated effectively to serve its target segment effectively, resulting in a range of different end products to achieve the same goal.

These challenges are only compounded when considering the niche properties shampoos can handle: anti-frizz, moisturizing, volume adding, color protection, organic or even sustainable solid shampoos. Each poses unique chemistry challenges for formulation chemists. For example, in the case of anti-frizz shampoo, such as the cyclomethicone shown in Figure 1, various oils and silicone-based additives are left in the hair to lock in moisture. However, shampoos are designed to cleanse the hair of oils and dirt, which frequently possess similar properties to the required additives in anti-frizz shampoos. How can chemists clean out one type of oil while leaving another behind? Often, these shampoos utilize a technique called dilution deposition. When the shampoo is rinsed from the hair – thereby diluting it – the desired oils and other additives self-form into structures called coacervates. These coacervates then precipitate out of the shampoo "solution" and are left in the hair. Achieving this result requires not only ensuring coacervate formation occurs; it must also happen in the right ratios to ensure the right amount of frizz is tamed for a given hair type.

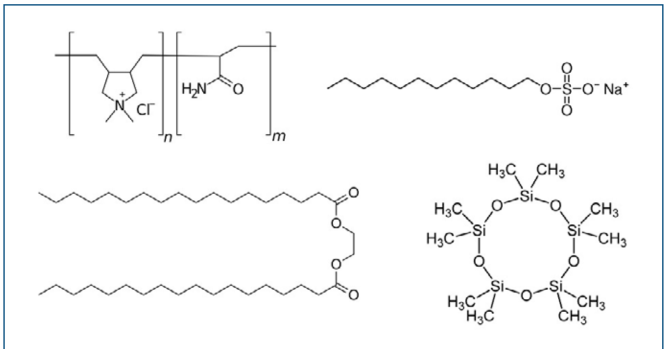


Figure 1: Examples of various hair care ingredients (clockwise from the top left): polyquaternium-7, sodium lauryl sulfate, cyclomethicone (D_5) and glycol distearate.

Each shampoo needs to be formulated effectively to serve its target segment effectively, resulting in a range of different end products to achieve the same goal.

To respond to such stratified end markets, personal care companies need to diversify their portfolios of shampoos to offer shampoo suitable for each unique hair type. Speed is often the key factor, as getting a new product to market after a competitor can result in a significant loss of market share. Yet the sheer number of potential formulations to test presents a significant challenge, as these speed requirements mean only a relatively small region of the planned design space can be tested in the lab. Additionally, design of experiment (DoE) processes for planning new formulations often rely heavily on expert knowledge. As a result, many promising formulations are left untested, limiting the overall potential of each project. This is driving a need to explore other methods to streamline the DoE process and expand the scope of early formulations design while improving R&D agility.

METHOD

To solve these problems, a leading personal care company turned to BIOVIA Pipeline Pilot. Their first project focused on optimizing the formulations of their anti-frizz shampoo. In this case, their marketing team had determined that shampoos which generated higher amounts of lather were considered to be more "premium" and "better cleaning" compared with those that produced less. As a result, their models sought to predict which formulations maximized lather production while maintaining appropriate levels of coacervate retention for given hair types. Pipeline Pilot offered the complete toolset to support this initiative.

Their first step was to gather the data necessary to build out the regression models for predicting lather production. Pipeline Pilot connected three separate internal databases to merge the necessary datasets. They were then able to automate the cleaning of this data with native filters as well as some custom filters developed in Python. For example, the entire dataset listed the ingredients in values such as mL/L or kg/metric ton of shampoo product; with Pipeline Pilot the data science team could automate the conversion of these values into percentages to standardize the comparison of formulations across the board. This helped to ensure that data curation did not happen in a "black box;" the data science team experienced complete control over where they sourced the data and how the data was processed. Additionally, since the processing of the training and testing data was standardized and automated, the team was able to set up the workflow to retrain their models with relevant experimental data as it became available in the future.

In addition to the work done by the data science team on predicting lather generation, formulations chemists sought to expand the types of coacervate-forming additives available to them. Many formulations have previously relied on various silicones such as cyclic and linear dimethicones; however, these silicones have a tendency to build up in hair, minimizing their effect and weighing down hair. As a result, consumers were demanding more sustainable and less maintenance-intensive additives to help fight frizz. With Pipeline Pilot, researchers were able to build out a separate regression model that predicted the surfactant's critical micelle concentration, a key indicator of effective coacervate formation, from chemical properties such as polymer molecular weight, charge density and charge density distribution. From this information researchers could better identify new compounds to test in the lab and incorporate into future formulations.

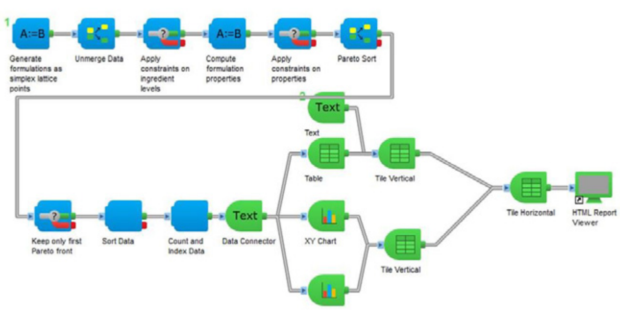


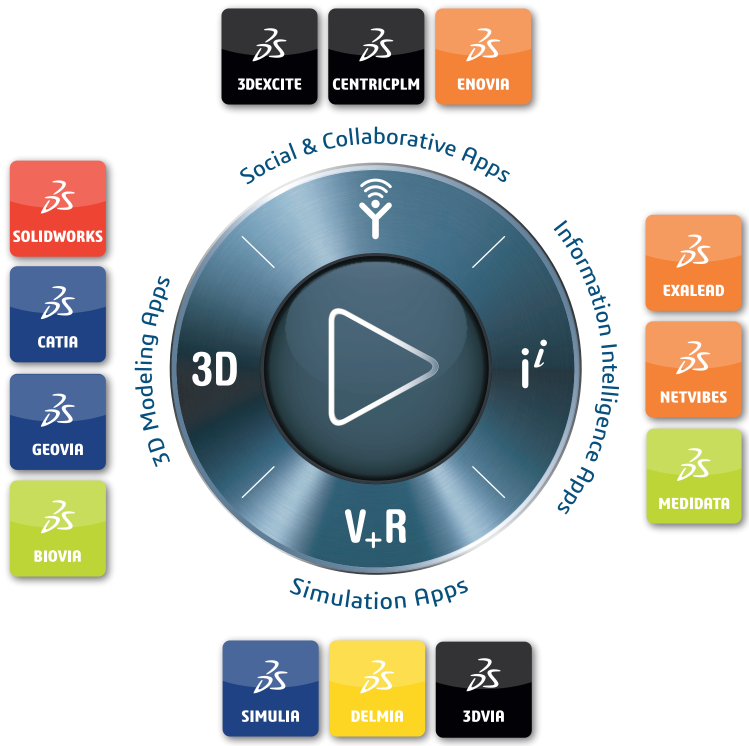
Figure 2: Workflow showing the optimization protocol and the generation of the corresponding DoE report.

With the data prepared, the team turned to examining different methods for their regression model. Pipeline Pilot offers a collection of prebuilt machine learning architectures, and in this case the team used a Random Forest, Genetic Functional Algorithm and a Deep Neural Network to predict the lather generation for a given formulation. They were then able to directly compare their performance to identify which model best suited their purposes: this time the Genetic Functional Algorithm provided the right mix of speed and accuracy. Armed with this model and the one that predicted coacervate formation for a given additive, the researchers could automate the creation and the virtual testing of thousands of potential shampoo formulations, automating their DoE. Finally, Pipeline Pilot allowed researchers to screen these recipes to identify the Pareto optimal formulations with highest lather generation within a given range for coacervate retention, as seen in Figure 2. This helped formulations chemists to target their physical experimentation on the formulations that seemed the most beneficial for each demographic.

RESULTS

Utilizing Pipeline Pilot, this personal care provider was able to efficiently streamline their DoE to help researchers identify the formulations with the highest chance to succeed. Their models effectively predicted lather formation across different ranges of coacervate retention, allowing the formulations team to easily classify optimized recipes for given customer demographics. Additionally, they identified a previously untested blend of amine-functionalized silicones that significantly increased coacervate retention without significant buildup. All of these results were subsequently validated in the laboratory. Since these models were built upon existing data – and retrained periodically as new data became available – researchers could feel confident in the predictions they generated. Looking forward, the protocols developed in Pipeline Pilot could be easily shared and repurposed to optimize formulations for other properties and identify tradeoffs, helping data science teams produce viable solutions for the business significantly faster. Researchers are currently exploring ways to optimize these models for "green only" ingredients to introduce a new line of sustainable shampoo and conditioner products.

Together, these benefits helped to scale this better targeted DoE approach and accelerate the optimization of product lines for specific end user types while also helping to diversify and expand the product portfolio more efficiently. In the end, this personal care provider improved its DoE and formulations design process to enhance the overall performance of its anti-frizz shampoos. They were also able to offer their broad range of customers a new and improved shampoo experience in a shorter more competitive time.



Our **3D**EXPERIENCE® platform powers our brand applications, serving 11 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE Company, is a catalyst for human progress. We provide business and people with collaborative virtual environments to imagine sustainable innovations. By creating 'virtual experience twins' of the real world with our **3D**EXPERIENCE platform and applications, our customers push the boundaries of innovation, learning and production.

Dassault Systèmes' 20,000 employees are bringing value to more than 270,000 customers of all sizes, in all industries, in more than 140 countries. For more information, visit **www.3ds.com**.

Europe/Middle East/Africa
Dassault Systèmes
10, rue Marcel Dassault
CS 40501
78946 Vélizy-Villacoublay Cedex
France

Asia-Pacific
Dassault Systèmes K.K.
ThinkPark Tower
2-1-1 Osaki, Shinagawa-ku,
Tokyo 141-6020
Japan

Americas
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

DASSAULT SYSTEMES | The **3D**EXPERIENCE® Company