Computer-Aided Design of Formulated Products

Kontogeorgis, Georgios M.; Jhamb, Spardha; Liang, Xiaodong; Dam-Johansen, Kim

Published in:
Current Opinion in Colloid & Interface Science

Link to article, DOI:
10.1016/j.cocis.2021.101536

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Computer-aided design of formulated products
Georgios M. Kontogeorgis, Spardha Jhamb, Xiaodong Liang and Kim Dam-Johansen

Abstract
Formulated products represent a particular class of complex chemical products, and their design is typically based on experience and extensive experimentation. Although still at an early stage, and despite that their potential is not fully accessed and not fully used by the industry, computer-aided design (CAD) methods and tools offer many possibilities in the design of formulated products. The CAD methodology based on computerized models enables the formulation chemists to speed up the design process, without completely replacing experiments. In this work, we summarize previous studies in the field and present important elements of the CAD framework, emphasizing estimation methods for key target properties, link to specifications, and finally, some case studies will illustrate how the CAD framework can be used in practice for formulated products.

Addresses
CERE and CoaST, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, Soltofts Plads 229, 2800, Kgs. Lyngby, Denmark

Corresponding author: Kontogeorgis, Georgios M. (gk@kt.dtu.dk)

Keywords

Introduction to complex product design
Design of many chemical and biochemical origin products is being characterized by an increased complexity over the last decades, coming from a shift that is observed in parts of the chemical industry: from materials valued for their purity, such as the commodities, to materials sold for their performance behavior, for example, the market-driven consumer products [1]. There are, in fact, limited raw materials, which are processed to obtain the commodity products (basic products). Specialty chemicals are then manufactured from the commodities, and finally, a very large portfolio of higher value products is obtained by processing and/or combining the chemicals of the previous classes. The number of products belonging to each category grows exponentially from around 10 for the raw materials up to more than 30,000 in the last class of higher value-added products [2]. Usually, the last class of chemical products is classified into three categories [3]: devices; complex molecules such as the pharmaceuticals; and (micro)structures including several consumer products, such as cosmetic and food products, where the key is the product function.

The class of structures includes the wide subclass of formulated products. Examples are the pharmaceuticals, paints, creams, detergents, and pesticides, where 5—20 (or more) ingredients are usually present, representing a wide range of chemical compounds, such as polymers, surfactants, solid particles including pigments and fillers, solvents, and aromas. For these (micro)structured products, the common practice in their development is still the experiment-based and trial-and-error approach. However, a systematic integrated procedure, where suggested higher value-added products are designed through a model-based methodology and then validated and/or refined by means of dedicated experiments, represents an efficient alternative, with respect to time and resources, speeding up the product development.

These consumer chemical-based products can be present in various physical forms such as liquid formulations, emulsions, and solid products. Their performance is related both to the presence of active ingredients and additives in the formulation and also to the product’s structural and material properties [4].

Among the complex chemical products, the formulated ones are particularly interesting, and despite the great variety in form and applications (e.g. skincare creams, cosmetics, paints, adhesives, detergents, inkjet printer inks, and so on), they have many characteristics in common. First, they are not ‘simple’ chemicals because they have many components, and each has a purpose. Typically, they include active ingredients (which satisfy the main needs, present in medium concentration), solvents (to ensure the product delivery, present in high concentration), and additives (to satisfy the secondary
needs, present in low concentration). Polymers, surfac-
tants, and other complex compounds are often present.
Nearly all structured products have a microstructure or
nanostructure which is essential for the application.
Colloidal and interfacial phenomena are very important
for such products (as their structures are typically in the
colloidal domain) [5], and making a good structure is
often a difficult job which is connected to both prop-
erties and process conditions. Moreover, such micro-
structured products often have anisotropic or complex
structure and properties, which can change during
application (e.g., wet liquid paint in the can and a dried
film after application).

Another common complexity of the design of micro-
structured products is the often complex and, some-
times, fragmented knowledge of the involved science
and engineering, as illustrated with some examples
in Table 1.

The design of many complex chemical products
including microstructures has been systematized over
the recent decades, and the discipline ‘chemical prod-
uct design’ has been today established with several well-
known textbooks in the field [3,6,7]. The approach
followed in this systematic design approach is first to
define the needs that a product should fulfill (often via
customer interviews). This can be a difficult step, and
the next one which is the transformation of the needs
into quantifiable specifications, which is especially
tough for many consumer products. After sometimes a
brainstorming section, many potential solutions can be
generated, and finally, one will, hopefully, be selected for
further processing and manufacturing.

The importance of product design and associated com-
plexities, including teaching challenges, has been
discussed extensively in the literature [5,8–12]. In
several of these works, for example, the study reported
by Uhlemann et al [5], it has been mentioned that a
comprehensive discussion of product design as one of
the building blocks of chemical engineering education,
research, and practice is still lacking, and this prevents
the discipline from achieving a broader academic and
industrial impact.

Owing to not only its importance in the field but also the
appearance of systematic procedures like the one
outlined previously, many [13–17] have over the years
stated or believed that product design is (or can turn out
to be) an emerging or the third paradigm in chemical
engineering (the first one being the unit operations from
the 1920s and the second being the transport phe-
omena from the 1960s). Specifically, according to Hill
[15], chemical product design and engineering shall be
recognized as the third chemical engineering paradigm.
This is so, because not only additional chemical engi-
neering approaches are needed but also a new mindset,
for solving the product design problems. Within this
paradigm, a systematic consideration of product formu-
lations before experimentation shall be preferred to
minimize the experimentation. However, so far this has
not been widely accepted, and it remains to be seen, but
nevertheless, complex product design is often featured
in many journals, for example, in the ‘Perspectives’
section of American Institute of Chemical Engineers
(AIChe) Journal over many years (2003, 2004, 2006,
and 2010).

One of the difficulties in establishing the complex
product design as a ‘paradigm’ can be related to the
difficulties in developing computational tools of mole-
cular structure–property relationships which can
indeed provide the scientific framework for this new
paradigm [13].

A particular difficulty is the establishment of measur-
able specifications for many of the needs of complex

<table>
<thead>
<tr>
<th>Field</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex thermodynamics</td>
<td>Solvent selection, thermopane windows</td>
</tr>
<tr>
<td></td>
<td>‘New’ separations</td>
</tr>
<tr>
<td>Colloids and interfaces</td>
<td>Paint brush, detergents, glues, and lubricants</td>
</tr>
<tr>
<td>Non-Newtonian rheology</td>
<td>Butter, sauce, paint, ink, glue, and whipped cream</td>
</tr>
<tr>
<td>Transport phenomena</td>
<td>Drying of paints, controlled drug release</td>
</tr>
</tbody>
</table>
| Making emulsions — disper-
isons                            | Milling of paint, agglomerated powdered milk, and foaming of cream      |
| Unusual unit operations      | Emulsification, granulation, and so on.                                  |
| Reaction engineering         | Advanced functional coatings, e.g., epoxy curing                       |

<table>
<thead>
<tr>
<th>Comments</th>
</tr>
</thead>
</table>
| Multicomponent (5–50 mole-
  cules), often polymers and
  complex molecules           |
| Supercritical fluid extrac-
  tion, membranes             |
| Wetting, adhesion, (meta)sta-
  bility                      |
| Often flow under shear       |
| Both diffusion and heat trans-
  fer + (often) thermodynamics |
| Particle/powder technology  |
| Extrusion, coating, spray cool-
  ing, ....                   |
| Scale-up is important!       |
| Many characteristics need opti-
  mization, e.g., adhesion, flexi-
  ibility, hardness, gloss, and so on. |
products. In some cases, the needs can be specified directly using physical properties, as also discussed later in the article. For example, as mentioned in the study reported by Cusler and Moggridge [3], the controlled atmosphere packaging for vegetables can be adjusted by the permeability of the packaging, whereas for anti-icing chemicals for planes, the important ones are the freezing point depression and characteristics such as the yield stress so that the chemicals stick to the plane wings and low viscosity to be removed before take-off. The difficulty once this relationship is established is to perform experiments with standard equipment or to have computational methods for the physical properties of interest.

There is, however, a completely different category of products, the so-called consumer products which have certain important characteristics or needs which cannot (always) be measured with conventional instruments or described in scientific terms. These needs are very important and can be related to consumers’ psychology as well. Examples are the color creams or skin lotions (which are emulsions) which contain main different components, for example, emulsifiers, foam stabilizers, surfactants, solubilizers, viscosity conditioners, and numerous other elements. Many of these products are oil-in-water or other types of emulsions. Here, we often encounter ‘less scientific’ concepts, for example, that of ‘creaminess.’ For example, as mentioned in the study reported by Cusler and Moggridge [3], market research shows a need for ‘thinner’ and at the same time ‘creamier’ skin lotions. Experimental data indicate that both thickness and creaminess are functions of viscosity and friction coefficients. Although such equations can provide hints on how much we should change the viscosity and the friction coefficient to obtain the desired effects with respect to ‘thinner’ and ‘creaminess,’ there is no handbook of consumer characteristics—physical property relationships. In most cases, they have to be obtained empirically. More examples are presented in the study reported by Cusler and Moggridge [3], for example, for the “tastiness” of a chocolate or the “crunchiness” of a cereal. Thus, converting needs into specifications and understanding consumer reactions are often difficult tasks including psychological and scientific terms as well as lots of empiricism and experimentation including expert panels. Bagajewicz et al. [18] and Wibono and Ng [19] have discussed extensively the design of creams, pastes, and other consumer products as well as the associated ‘difficult to quantify’ concepts of softness, creaminess, spreadability, greasiness, and effectiveness of the final formulation.

For all the aforementioned reasons, the design of complex chemical products is often based, even today, on tedious experimentation and many years of experience. With the advent, however, of computers and large databases combined with computational techniques and predictive property estimation methods, computer-aided design (CAD) methodologies have been developed. They cannot fully, as yet, replace experiments, but they can contribute significantly to the systematic development of formulated products by searching for various alternatives. These methods are summarized in the next section, followed by a discussion in section 3 of the predictive property estimation methods on which such computational methods heavily rely on and finally in section 4 with four characteristic case studies illustrating some of the characteristics of CAD.

**Computer-aided design framework and previous studies for formulated products**

By now, a wide variety of computer-aided methods and tools have been developed for the design of chemical products represented by the properties of a single molecule or mixtures of molecules. These methods are broadly classified as CAMD (computer-aided molecular or product design) approaches and for an excellent historical overview of these methods including the major contributors, we recommend the 2016 review by Gani et al. [20]. Those authors presented timeline diagrams of CAMD developments since the 1980s for product design and for integrated product-process design.

CAMD methods include the generation of feasible chemical structures, the estimation of the thermophysical properties through property models, and finally the selection of those molecules that match the desired targets [21]. For mixtures, the properties depend on the mixture composition, and the design algorithm needs to identify the molecules and their compositions in solution matching the target properties. Systematic decomposition-based solution approaches are usually used to manage the complexity of these design problems efficiently, by reducing the search space [21]. The main elements of CAD methods are summarized schematically in Figure 1. It should be mentioned that the mathematical problem is often complex and is addressed by a mixed integer linear/nonlinear programming model, as discussed, for example, by Gani et al. in the study reported by Zhang et al. [22]. This is particularly important when highly nonlinear property models are used (such as the Universal Functional-group Activity Coefficients (UNIFAC) estimation method, see next section). Owing to the complexity, a decomposition-based algorithm is often used by first solving subproblems consisting of a set of property constraints only.

CAMD-based methods supplement the experiment-base trial-and-error approach [23]. In the integrated experiment-modeling approach, used when mathematical models are not available for all the target properties, the design problem is decomposed into a hierarchical sequence of subproblems: as one goes from the outer levels to the inner levels, the number of candidates
decreases, and the inner levels use experiments for the final product refinement and/or validation. With this combined approach, high chances for innovation are guaranteed by the model-based structure, and the development time and the consumption of resources are kept low, as the experiments are carefully designed for final refinement. Let us now briefly discuss the main elements of CAMD.

Knowledge base (problem definition)
Even if such data are not systematized in many cases, an appropriate problem definition involves the collection of information, for example, the customer needs, the product quality factors, and the product technical specifications from a wide range of sources (market and customer surveys, patents, literature, and so on). Often several disciplines are involved in this process (engineering, marketing, manufacturing, and economics). The role of the knowledge base at the level of the problem definition is very important, by providing information about all involved materials and their target properties in a specific formulation.

Property models
Knowledge of the product properties, from experiments and models, is crucial in the design and development of chemical formulated products [2], as they are inherently connected to the needs of the product. Many of these properties are thermophysical, in nature, and are helpful in determining many functions of a product, for example, its stability, the evaporation of the solvent mixture on application, or its spread ability, as we discuss later.

Not all properties can be measured, and this is also time-consuming, especially for multicomponent formulations. A hybrid approach is particularly useful, where models can provide estimation for a wide range of properties/systems, and experiments are performed to verify selected promising candidates. A database of experimental data and prediction property models is needed in all cases.

Structured databases (methods and tools)
Many product design problems are so complex that they only partially rely on CAD methods. Very often, the different ingredients of a formulated chemical product are selected from dedicated databases, by means of model-based techniques.

For this reason, several different databases must be available for the categories of ingredients necessary for the solution of the different problems. Examples of databases available in the Integrated Computer-Aided System (ICAS) software developed by Technical University of Denmark (DTU) are presented in Table 2. For many formulated products such as emulsified ones and coatings, at least part of the design is based on databases especially for additives and the ‘secondary’ components, as explained in studies reported by Kontogeorgis et al. and Jhamb et al. [24—26].

Framework and methodology
The systematic methodology, integrating model-based and experiment-based techniques as described previously, is based on the product of interest. The output will be a validated formulation, containing a list of...
The exact framework depends both on the type of product and computational possibilities. Despite the fact that sequential design approach offers a very efficient workflow, Jonuzaj et al. [28] argued that it is not possible to answer separately how many compounds and their composition shall be in a product, and what the best active ingredients and solvents are to achieve the specific product attributes. Therefore, the authors developed a computer-aided product design method for designing solvent-based acrylic adhesives without the need to be restricted in a sequential design. With the given design example of environmentally benign solvent-based adhesive products, the authors demonstrated that 8–12% of toxicity can be reduced in comparison with the sequential design approach, that is, fixing the active ingredient a priori. The authors emphasized that uncertainty analysis shall be an important aspect to be investigated for more robust designs.

On the other hand, as Arrieta-Escobar et al. [29] emphasize, whereas for relatively simple formulations, CAD methods perform very well, the situation is far more difficult for more complex formulated micro-structured products, often owing to the lack of adequate property models. These authors attempted to integrate heuristic rules formally, in a mathematical manner, into the computer-aided product design framework and demonstrate its application for a hair conditioner case. As a typical cosmetic formulation, many ingredients are usually involved in a hair conditioner, and their heuristic knowledge regarding qualitative functions, impacts on sensorial attributes, compatibilities, and synergies is very rich. For the given case study, 35 ingredients have been considered which belong to emulsifiers, thickeners, and emollients. The authors have explained how the heuristic knowledge in different categories can be incorporated explicitly into the optimization problem. Different from many other works on computer-aided product design, the authors have made experimental validations for their designs. In total nine formulations that fulfill expected properties were designed by their methodology with cost as the primary objective function, and all of them were found similar as per the measured rheological, textural, and microstructural properties. The authors concluded that any of these nine formulations can be used as a starting point for inexperienced formulation designers, and they also emphasize that the proposed methodology shall be applicable to any formulated products.
Another important aspect is, according to Raslan et al. [30,31], the early identification of harmful ingredients which can significantly minimize product development cost and delays, but where, as they state, limited work has been performed in identifying hazardous chemicals used in a formulated product during the early stages of the design process. They recognize that safety and health elements have been included in some product design studies but are often limited to flammability and toxicity properties and less so for severe health hazards such as carcinogenic nature of certain chemicals. They conclude that safety, health, and environment are necessary aspects to consider in designing formulation products, and they developed safety subindices including gas explosiveness, liquid flammability, solid/liquid/gas toxicity, chemical reactivity and solid dust explosiveness, and health subindices owing to exposure route via eyes, inhalation, ingestion, dermal, or their combinations. These indices formed a so-called product safety and health index, which was then integrated into the formulated product design framework proposed by Zhang et al. (2017). The authors further applied the design framework to the case study presented by Conte et al. [32]. It is identified that one of the chosen solvent components, toluene by Conte et al., has high flammability potential and high health hazard via the ingestion exposure route and the additive Sodium Dodecyl Sulphate (SDS) to be hazardous as well after dermal exposure. In parallel, the authors have developed two assessment tiers for safety and health index, which was then integrated into the formulated product design framework proposed by Zhang et al. (2017). The authors further applied the design framework to the case study presented by Conte et al. [32]. It is identified that one of the chosen solvent components, toluene by Conte et al., has high flammability potential and high health hazard via the ingestion exposure route and the additive Sodium Dodecyl Sulphate (SDS) to be hazardous as well after dermal exposure. In parallel, the authors have developed two assessment tiers for safety and health index, which was then integrated into the formulated product design framework proposed by Zhang et al. (2017), and the application of the developed methodology was demonstrated for two formulations of sunscreen products.

Maybe many of the complexities owing to lack of accurate methods could be accommodated, according to Cao et al. [33], by a combination of machine learning (ML) methods and the advancements of digitalization. According to these authors, the trade-off between exploitation and exploration is still a challenge in reducing time and resources needed for product development. Cao et al. [33] considered designing and developing formulations in a fully automated fashion. By combining fast and reliable data collection and targeted experimentation guided by surrogate models, it is possible to make fully robotically automated product developments efficiently via a closed-loop optimization approach. The authors also discussed the challenges regarding the data, hardware, models, and software tools in this approach, although they believe a fully digitized workflow is the general trend that the development of the formulation technology is likely to follow in chemical R&D and manufacture.

**Algorithm descriptions in general terms**

With reference to Figure 2, we present a discussion of the various steps.

The problem definition stage is very important, as it is the first in the hierarchical structure of the methodology, and backward interactions from the other stages may not be always possible. That is, any decision taken at this level influences the decisions taken in the following steps, but not necessarily vice versa, as shown in Figure 2. At this point, a list of target properties and their target values and a list of necessary categories of ingredients are generated for the type of product of interest as the input. The main tool used in this stage is the knowledge base.

During the model-based stage, the results of the problem definition stage are converted into suitable formulated products, for example, emulsions. This happens via a combined use of property models, structured
databases, and dedicated algorithms. Several candidate ingredients and their relative concentrations are determined. The screening process of thousands of candidates is achieved by applying a decomposition strategy so that the solution method is divided into a set of subproblems which are solved individually. During the experiment-based stage, the selected formulated product is verified through experiments. If the results of these experiments do not match with the expected results, appropriate corrections can be taken. All three stages have been applied for some emulsified products, for example, in Ref. [24]. As illustrated in Figure 2, iterations between the model-based stage and the experiment-based stages are typically needed until a candidate formulated product generated by the model-based stage is verified by the experiment-based stage.

The main objective of the proposed methodology is to efficiently screen between many candidates by means of mathematical models and algorithms so that the valuable experimental resources are reserved for the final verification/refinement. This, however, is possible only if a set of product needs, in terms of target properties to be satisfied by and necessary categories of ingredients to be included in the candidate formulated product, is generated.

**Software tool (methods and tools)**

The solution of a chemical product design problem requires management of numerous models, data, and procedures. Computer-aided tools are necessary, and one such tool is the virtual process-product design laboratory (VPPD-Lab), originally developed for the design and analysis of homogeneous formulated products, which has been refined and extended to include formulated products [24].

The VPPD-Lab allows searching for the most promising candidates while also recommending experiments to verify the product formulations using an integrated knowledge base. Computer-aided techniques are used to search through a wide range of alternatives. Similarly, to a process simulator analyzing different chemical processes, the VPPD-Lab is able to design and analyze different chemical products. The VPPD-Lab is supported by tools, such as a property model library, a knowledge base, structured databases, and calculation routines, see the study reported by Kalakul et al. [34] for more information. The software can be provided to universities by contacting the corresponding author.

The VPPD-Lab has several possibilities. In the first step, the product type is selected from a list of products available in the database (formulations, gasoline blends, lubricant blends, jet fuel blends, and emulsion-based formulations). Next, product needs are retrieved from the knowledge base and additional information, if needed. In the third step, the product needs are translated to product attributes (properties) and property target values. In the fourth step, we select the ingredients (chemicals) and their amounts such that the property targets are satisfied. The ingredients include the active ingredients, solvents, and additives. In the final step, each feasible product formulation is verified through model-based tests to check for stability, performance enhancement, and so on. The experimental component at the end of a CAD process is discussed in detail in Kontogeorgis et al. [24] and Conte et al. [35].

The above constitutes a rather simplified presentation of the CAD approach for formulated (and other complex) products, and the interested reader is referred to many more extensive reviews available in the literature from Gani et al. and others [20,22,36–42].

In addition to the ICAS/VPPD-Lab, there are several other Computer-Aided Molecular/Product Design (CAMD/CAPD) software and tools available. We mention some characteristic ones:

a) QMaC: A quantum mechanics/machine learning-based computational tool for chemical product design [43]. This computational tool is a computer-aided tool developed at the Dalian University of Technology based on quantum mechanics and ML techniques. The authors claim that they can use it for the better design of organic solvents, inorganic materials, fertilizers, and pesticides, polymers, catalysts, and other chemical products for human needs.

b) OptCAMD: An optimization-based framework and tool for molecular and mixture product design [44]. A versatile tool for chemical product design and evaluation ‘OptCAMD’ wherein an optimization-based mathematical programming model is established and solved to generate feasible molecules and/or mixtures together with optimal product candidates.

Finally, although the focus of this review is on CAMD/CAD methods using conventional computational tools, it is worth mentioning that recent developments in artificial intelligence (AI)-based approaches can play important role in the future use of such CAD methods. When sufficient theoretical correlations of the properties are unavailable, ML-based models coupled with available data could help to develop the needed property models [42]. Data-driven or ML methods have opened new opportunities for the discovery and design of materials. Different packages in Python and toolboxes in MATLAB can be used to build ML-based algorithms. The application of the methods has been shown via three case studies for the design of polymer and porous...
materials, catalytic materials, and energetic materials [45].

**Role of property models in computer-aided design of formulated products**

*In general terms and examples of needs-properties link*

It was apparent from the discussion in the two previous sections that property models play a very important role in CAD. As illustrated in Figure 4, especially the predictive group contribution (GC) methods are particularly useful. Many such methods exist today for many properties, see studies reported by Gani et al. [39,46] and Kontogeorgis et al. [47] for reviews and also references [20,36] for an overview of properties which can be predicted by GC methods. Although these methods can be used to predict the properties, in the inverse scenario, they can be used to generate functional groups and at the end molecules and mixtures when the required property values are known or specified. This is, in essence, the ‘heart’ of the CAD approach.

In the classical approach, when the structure of molecules is known, the properties can be estimated using predictive, for example, GC methods. In the computer-aided product design, the inverse approach is being used. We start from a set of desired properties, and using suitable GC methods, several compounds or mixtures satisfying these properties are generated.

Some examples are shown in Tables 3–6 for a blended design and several formulated products. More details and references about the estimation methods are provided in the references indicated in the tables. The list of properties included in these tables is not complete. Several more properties could be included for specific applications or to satisfy certain conditions. For example, in the case of fuels (Table 3), the environmental impact may include as well as trace metals, nitrogen and sulfur content, and so on.

Reference [48] includes property estimation models for 10 properties which are of direct relevance for the design of blended products.

In addition to identifying the correct correspondence between needs and target properties and the availability of predictive estimation methods, the boundaries of the target values (not shown in Tables 3–6) should also be known. These are sometimes available from product requirements, environmental and safety constraints, and so on, but they are not always precisely known. Hence, a sensitivity analysis should be carried out.

There are many similarities in the target properties for the different categories of products. For example, Mattei et al. [54] present a list of target properties for homogeneous liquid mixtures and emulsion-based products (11 and 10, respectively). Almost all are the same (Gibbs energy of mixing only needed for liquid mixtures), but, of course, the precise number and type of properties needed will depend on the application. A similar conclusion can be drawn from Zhang et al. [20] where they presented a table with target properties for six types of products (refrigerant, detergent, insect repellent, solvent mixture, gasoline blend, and herbal medicine). 5–9 properties are identified per product type, and some are common, but there are also several differences. Over 20 different properties were needed for these products mentioned by Zhang et al. [20].

Indeed, as can be seen from Tables 3–6, a wide range of properties are needed for many complex products including formulations. Some of these properties are thermodynamic (density, miscibility, vapor pressure, …), others are transport (often viscosity), many related to colloids and interfaces (surfactant characteristics such as critical micelle concentration, Krafft and cloud point, the hydrophilic-lipophilic balance for emulsions, and the surface tension), see Table 7, some across disciplines, for example, solubility parameters (especially Hansen ones), environmental and safety properties, for example, flash point and toxicity parameters and many combined properties such as permeability and evaporation rate. Colloid and interfacial science have been emphasized by many as a key discipline in connection to the design of formulated products [5,6].

Many of these properties are needed also for mixtures, and in some cases, linear mixing rules can be, at least approximately, used, but there are several cases where
complex nonlinear mixing rules are required (e.g. for miscibility/Gibbs energy of mixing, evaporation rate, and flash point).

For these reasons, extensive databases with property prediction methods especially GC ones are highly useful for CAPD methods. Figure 5 shows such a list from the ICAS software (DTU), but other collections are also available. As indicated, all these property databases are hardly complete, and new methods are often being developed for missing properties, for example, the study reported by Jhamb et al. [55] presents a method for biodegradability of a wide range of compounds and reference [56] for the Hansen solubility parameters (HSPs) of pigments and in the study reported by Fardi et al. [57] for the HSP of amino acids and fatty acids, in both cases, using both first and second-order group contributions. These are some characteristic examples of recent developments. The combination of the two well-established concepts of HSP and GC is particularly powerful in many product design applications, for example, Councell and Allwood [58] have used it for designing solvents which can remove the toner print in a way that office paper can be reused. Reference [59] presents an extensive list of GC methods for many properties of ionic liquids, which by many are considered promising ‘green’ solvents in many applications.

Important developments can also be mentioned for structure–property relationships which attempt a quantification of the customer needs. For example, Teixeira et al. [60] have used Stevens law and thermodynamic models for describing the smell of perfumes. This is based on a mathematical model of the diffusion of fragrance (perfume) components through the air and the use of the UNIFAC model for activity coefficients for the estimation of vapor-phase concentrations. Although equilibrium is assumed, some parameters need to be estimated related to the minimum detectable concentration, but the overall agreement with the experimental data for many perfumes is very satisfactory.

| Table 3 |
|------------------|------------------|------------------|
| **Need**         | **Target property** | **Estimation method** |
| Ability to be burned | Reid vapor pressure | GC (UNIFAC) |
| Flammability     | Flash point       | GC (UNIFAC) |
| Engine efficiency| Higher heating value | Linear mixing rule |
|                  | Density at 15 C   | Modified Rackett equation |
|                  | Research octane number (RON) | Linear mixing rule |
| Consistency of fuel flow | Dynamic viscosity | Linear mixing rule |
| Liquid blend miscibility | Gibbs energy of mixing for miscibility check | UNIFAC |
| Environmental impacts | Oxygen content | Linear mixing rule |
|                    | -logLC50 (lethal concentration as measure of toxicity) | Linear mixing rule |
|                    | CO2 emission in the combustion engine | |

| Table 4 |
|------------------|------------------|------------------|
| **Need**         | **Target property** | **Estimation method** |
| Drying time (or long lasting for repellents) | Evaporation time | GC + linear mixing rule |
| Toxicity         | Lethal concentration | GC + linear mixing rule |
| Spray ability/flowability | Dynamic viscosity | GC + linear mixing rule |
| Wetting and spread ability | Surface tension | GC + linear mixing rule |
|                  | Molar volume | |
|                  | Viscosity | |
| Solubility       | Gibbs energy of mixing | UNIFAC-LLE |
|                  | Hansen or Hildebrand solubility parameters | |
| Flammability     | Flash point | GC method |
| Water-based formula | Water, water-miscible solvents | GC- UNIFAC |
Despite these successes, color and odor are still very difficult properties to estimate with predictive methods, as also emphasized by Zhang et al. [22] who do not report any generally available methods for these two properties of relevance to product design.

**Complications, some special features, and some new trends**

An evident problem with this methodology, as emphasized by many companies in a recent review [44], is the high complexity of the molecules and mixtures involved in product design as well as the very wide range of properties needed (thermodynamic, interfacial, transport, performance, and so on). The companies in the study reported by Kontogeorgis et al. [47] mentioned that particular gaps are noted for complex and electrolytic mixtures (including ionic liquids) and for large organic molecules with several functional groups. They also mentioned that there is increased need to describe surface effects and complex systems like those involving micelles. We can mention that a rather recent review [54] summarizes many predictive methods for emulsion-related properties, whereas group contribution (GC) methods for the

---

### Table 5

Connection of needs to translated target properties for a handwash in an emulsified form and methods used for their estimation. Adapted from studies reported by Kontogeorgis et al. [24] and Mattei et al. [52]. GC stands for group contribution method, QSPR stands for Quantitative Structure Property Relationship and LD50 stands for the Median Lethal Dose.

<table>
<thead>
<tr>
<th>Consumer needs</th>
<th>Target Properties</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>High foam ability</td>
<td>Surface tension</td>
<td>GC and linear rule</td>
</tr>
<tr>
<td>Nonirritability of the skin</td>
<td>Critical micelle concentration</td>
<td>GC + method</td>
</tr>
<tr>
<td>Wetting of the substrate</td>
<td>Solubility parameters</td>
<td>GC and linear rule</td>
</tr>
<tr>
<td>Spread ability</td>
<td>pH</td>
<td></td>
</tr>
<tr>
<td>Thermal stability</td>
<td>Surface tension</td>
<td>GC-UNIFAC</td>
</tr>
<tr>
<td>Good stability as emulsion</td>
<td>Krafft temperature</td>
<td>GC method</td>
</tr>
<tr>
<td>Antibacterial performances</td>
<td>Antibacterial agent</td>
<td>QSPR method</td>
</tr>
<tr>
<td>Pleasant color</td>
<td>Colorant</td>
<td>Definition</td>
</tr>
<tr>
<td>Pleasant odor</td>
<td>Aroma</td>
<td></td>
</tr>
<tr>
<td>Pleasant skin feeling</td>
<td>Emollient</td>
<td></td>
</tr>
<tr>
<td>Low toxicity</td>
<td>Toxicity parameter, LC50, or LD50</td>
<td>GC method</td>
</tr>
<tr>
<td>High safety</td>
<td>Flash point</td>
<td>GC-UNIFAC</td>
</tr>
</tbody>
</table>

### Table 6

Connection of needs to translated target properties for a tank cleaning detergent and methods used for their estimation. Adapted from the study reported by Mattei et al. [53]. GC stands for group contribution method, QSPR stands for Quantitative Structure Property Relationship, LD50 stands for the Median Lethal Dose and LLE stands for Liquid-Liquid Equilibrium.

<table>
<thead>
<tr>
<th>Need</th>
<th>Target property</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleaning performance</td>
<td>Surface tension</td>
<td>GC or QSPR</td>
</tr>
<tr>
<td>Low-foaming ability</td>
<td>Hansen solubility parameter</td>
<td>GC + linear mixing rule</td>
</tr>
<tr>
<td>Stability</td>
<td>Hydrophilic-lipophilic balance</td>
<td>GC or definition</td>
</tr>
<tr>
<td>Stability</td>
<td>Surface tension</td>
<td>GC or QSPR</td>
</tr>
<tr>
<td>Stability</td>
<td>Critical micelle concentration</td>
<td>GC</td>
</tr>
<tr>
<td>Stability</td>
<td>Cloud point</td>
<td>GC + linear mixing rule</td>
</tr>
<tr>
<td>Stability</td>
<td>Hydrophilic-lipophilic deviation</td>
<td>Definition</td>
</tr>
<tr>
<td>Easy application</td>
<td>Kinematic viscosity</td>
<td>UNIFAC-LLE</td>
</tr>
<tr>
<td>Improved dissolution</td>
<td>Presence of a bleaching agent</td>
<td>GC method</td>
</tr>
<tr>
<td>Defined color</td>
<td>Presence of a pigment</td>
<td></td>
</tr>
<tr>
<td>Low impact on the surface</td>
<td>Presence of a builder</td>
<td></td>
</tr>
</tbody>
</table>
Critical micelle concentration and cloud points of many surfactant families have also recently been developed [52,61], and a thermodynamic-based approach for the assessment of the stability of an emulsified solvent mixture has also been proposed [52].

In addition to these challenges, the companies commented, in the study reported by Kontogeorgis et al. [47], on the, sometimes, weak link between thermophysical and the final product—performance properties. It is emphasized that product performance cannot always be expressed directly by a thermodynamic property, for example, drying time of a paint is related to evaporation rate, hence vapor pressure of the solvent. In addition, the product stability is important, here there is combination of thermodynamics and reaction kinetics. The product performance properties (or lack of methods for them) is considered by many a serious obstacle and general challenge, and the companies do not expect that CAPD methods could fully replace the existing methods in product design in the near future, although an acceleration can be anticipated. A better connection between existing design methods in industrial practical and CAPD approaches will contribute to the faster acceptance of the latter.

There are several more special features and complications regarding the predictive estimation methods for properties of relevance to product design. First of all, as mentioned, by far the most popular methods are the GC ones, as they can be used both for property estimation and generation of compounds and mixtures with desired properties (the reserve problem). We have mentioned references summarizing such methods [46,47,54] and recent developments [52,55–57,61]. Sometimes, these methods are enhanced, the so-called GC + versions, with connectivity indices and other ways to extend their applicability for compounds for which no data are available for estimating the missing parameters. A comprehensive review in the study reported by Hukkerikar et al. [62] presents GC and GC + methods for about 20 properties, and a follow-up work from the same group [63] presents GC + methods for 22 environmental-related properties such as global warming potential, ozone depletion potential, LD$_{50}$, LC$_{50}$, bioconcentration factor, and many more. A recent review [46] presents references and discussions for many more GC methods including several for properties of amino acids, acid dissociation constants, ionic liquids, and lipid systems.

Among the various GC methods, UNIFAC is particularly relevant. This is a GC method for activity coefficients and thus useful for stability-miscibility estimations but, indirectly, as indicated previously also for many other properties such as flash point, viscosity, and surface tension. Owing to — among other reasons — the great importance of GC methods in CAD, it has been emphasized both in the most recent [47] and in the previous review [64] carried out from the European Federation of Chemical Engineering that ‘the full phasing out of GC models in favor of modern models is not yet imminent.’ This statement was made in the 2010 review [64], and 10 years later, we could safely conclude that this is still the case for a wide range of practical applications.

Table 7

<table>
<thead>
<tr>
<th>Concept</th>
<th>Need</th>
<th>Product</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hansen solubility parameters, HSPs</td>
<td>Solubilization or dissolution</td>
<td>Paints and coatings</td>
<td>[24–26,32]</td>
</tr>
<tr>
<td></td>
<td>Cleaning</td>
<td>Solvent selection in blends</td>
<td>[45–47,49,50]</td>
</tr>
<tr>
<td></td>
<td>Nonirritability of skin</td>
<td>Tank detergent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hand wash</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lubricants</td>
<td></td>
</tr>
<tr>
<td>Surface tension</td>
<td>Wetting</td>
<td>Paints and coatings</td>
<td>[25–27,32]</td>
</tr>
<tr>
<td></td>
<td>Foaming</td>
<td>Detergent substitution</td>
<td>[49,50]</td>
</tr>
<tr>
<td></td>
<td>Cleaning ability</td>
<td>Tank detergent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spray ability</td>
<td>Hand wash</td>
<td></td>
</tr>
<tr>
<td>Critical micelle concentration, CMC</td>
<td>Foaming</td>
<td>Tank detergent</td>
<td>[49,50]</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>Hand wash</td>
<td></td>
</tr>
<tr>
<td>Cloud point</td>
<td>Thermal stability</td>
<td>Tank detergent</td>
<td>[49,50]</td>
</tr>
<tr>
<td>Krafft temperature</td>
<td>Thermal stability</td>
<td>Hand wash</td>
<td>[49,50]</td>
</tr>
<tr>
<td>Hydrophilic-lipophilic deviation, HLD</td>
<td>Stability</td>
<td>Tank detergent</td>
<td>[49,50]</td>
</tr>
<tr>
<td>Hydrophilic-lipophilic balance, HLB</td>
<td>Solubility</td>
<td>Hand wash</td>
<td>[49,50]</td>
</tr>
<tr>
<td></td>
<td>Cleaning performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some key concepts from colloid and surface science which have been used in computer-aided design of formulated products. Many of these can be predicted from GC methods, see the study reported by Mattei et al. [54] for a review of such estimation methods. These are just some characteristic examples.
On the other hand, as discussed in the study reported by Kontogeorgis et al. [47], GC methods in general, and UNIFAC in particular, have problems for complex functional chemicals like many of the ones used in product design and industries such as biotechnology and agrochemicals. The companies which participated in the survey of Kontogeorgis et al. [47] mention the lack of estimation methods for solid mixtures including solubility prediction and emphasize that the science in this case is clearly less mature than for fluids. One company mentioned characteristically at an internal survey they conducted some years ago showed that for the active ingredients from the current portfolio at the time, only one molecule could be incremented by UNIFAC and that more advanced statistical associating fluid theory (SAFT)-type methods were not helpful either. They concluded that with so many ‘gaps’ (molecules that cannot be described ranging from 20% at best up to 70%) and with the rate of group parameters being generated per year, it is unclear how much predictive
GC methods will become more applicable for dissolved compounds over the next 10—20 years.

Other companies emphasized the complex multicomponent nature of many mixtures of interest (including a wide range of solvents, solids, ionic species, and macromolecules such as proteins and enzymes) and the importance of many phenomena which cannot be easily assessed, for example, polymorphism and bulk behavior-like flowability and also mention that crystallization modeling remains very difficult and requires parameterization, which is difficult to carry out. Despite all these difficulties, UNIFAC is a widely used tool in the industry today, as emphasized by the industry [47,64] both for process and product design, at least at preliminary stages.

In some CAD applications where higher pressures are involved, equations of state (EoS) are to be preferred as the pressure is inherently incorporated in such models. GC-based EoS are more suitable for product design and — among the cubic such EoS — the so-called volume-translated Peng-Robinson is the best choice. VTPR is fast becoming a widely accepted model in the industry, as explained in the study reported by Kontogeorgis et al. [47], and represents a GC-based EoS where UNIFAC is incorporate in a consistent way in the cubic EoS, thus permitting the use of the UNIFAC model at high pressures. This is mostly relevant for process applications.

For more complex mixtures, the SAFT theory is sometimes recommended. SAFT stands for statistical associating fluid theory, its development started in the late ‘80s and is now one of the more popular EoS in applied thermodynamics with numerous variants. For a detailed discussion of the model with many references, see studies reported by Kontogeorgis and Polas [65] and de Hemptinne et al. [66]. SAFT is mentioned in some CAD reviews [25,46], and there are several GC versions of the SAFT model as well. By GC versions, it is meant that GC approaches have been developed for estimating three of the parameters of the model but not all of them (not the hydrogen bonding ones). SAFT is a model which has the potential of being applied to polar and multifunctional molecules, electrolytes, ionic liquids, and polymers and for many properties including transport and interfacial ones and for these reasons has the potential — at some time in the future — of being highly useful for CAD approaches. It has not as yet reached that level, certainly not for practical applications. As discussed in studies reported by Jhamb et al. [25], Liu et al. [43], and Tiexeira et al. [60], but especially in the recent survey [44], industries mention many challenges with the SAFT approach: there are far too many variants, and it is difficult for the industry to choose a priori which one is best, there are several calculation speeds and implementation issues (the model slows down the simulations), calculations are still not very satisfactory or predictive ones for complex situations such as the liquid—liquid equilibrium, and many parameters are missing for many molecules, and parameter estimation is a tedious exercise for such models. Anyhow, SAFT is one of the new trends with potential for CAD in the future. Other trends are mentioned in the survey [47] and in the study reported by Gani [46] which may be of relevance in the CAD methodologies (molecular simulation, artificial intelligence, quantum-chemical approaches such as CONductor like Screening MOdel for Real Solvents (COSMO-RS)), and the interested reader is referred to these references for more details.

Finally, we should discuss in some detail the application range of the property models as this is inherently connected to the application range of the CAMD method. Selecting the property model implicitly defines the search space of the CAMD approach. We need, therefore, to develop property models that can be used over a wide application range and also have information on the associated uncertainties in property estimations. Maranas [67] has incorporated the uncertainty of property estimation methods within the CAMD problem definition.

Another difficulty with UNIFAC and other GC methods is the nonavailability of model parameters for many molecules. This can eliminate potentially optimum choices. We need predictive property estimation models with few parameters and a large application range. Further development of CAMD methods for applications in structured products and formulations is closely related to the availability of the required property models.

The design of complex molecules involving isomers can be facilitated by Quantitative structure-activity relationship (QSAR) methods which have become popular. Properties estimated through parameters obtained from dynamic modeling and/or molecular modeling are necessary when microscopic and/or mesoscopic scales have been used for molecular structural representations. The need is to develop special quantitative property models based on the data generated from dynamic and/or molecular modeling plus any available experimental data. The property estimation task could be arranged on a hierarchy based on the computational effort and cost related to obtaining a property value.

It is worth discussing QSAR/QSPR methods a bit more. The QSAR methodology makes use of the differences in observed biological activity for a series of compounds that can be quantitatively correlated with differences in their molecular structure. These models have been used for modern drug discovery and design to arrive at molecules that have a combination of optimized
pharmacodynamic and pharmacokinetic properties [68,69]. In the case of QSAR, a chemical property (instead of activity) is modeled by correlating with the topostructural or topochemical features. The QSAR/QSPR models have several types of variants. They could either be GC methods, three-dimensional (3D)-QSAR, or chemical-descriptor-based QSAR [70].

Obviously, the experimental measurement of the property should be at one (high) end and simple, first-order GC methods could be at the other (low) end. The largest number of compounds of different types is handled at the lower end and as one proceeds upward, the number of compounds of different types decreases, but the number of isomers that can be handled increases. In this way, the computationally intensive calculations are saved only for those candidates that have satisfied all other constraints based on the lower-level property models. Note that even in this approach, the uncertainties of prediction accuracy may eliminate some candidates. On the other hand, the method would systematically move toward the solution, provide useful insights, and keep the computational load at a manageable level. Note that if pure component and mixture properties were needed in a CAMD problem, the pure component properties would be estimated first. This would reduce the computational load significantly for the estimation of mixture properties. In addition, this may make the mixture property model more acceptable because some molecules that could not otherwise be handled would be removed owing to a specified property constraint and not because of unavailable model parameters.

Case studies on computer-aided design of formulated products

The application of CAD to formulate chemical and biochemical-based products in different sectors is highlighted through two selected case studies.

Paints and surface coatings

Background

An antifouling coating that is based on the self-polishing copolymer coating technology typically contains acrylic or methacrylic copolymers which are easily hydrolyzable in seawater, biocides, or biocidal pigments, a solvent or mixture of solvents and molecules that regulate the release of biocides into seawater. To formulate such a paint, the molecules that regulate the leaching of the biocides are a critical additive and hence very important to determine the performance of the paint. This performance is measured in terms of the thickness of the leached layer. Rosin is commonly used to regulate the biocide release in commercial self-polishing copolymer coating technology paints. Apart from rosin, there could be other molecules that serve the same function and deliver an improved product performance. To identify such molecules, property models could play a vital role.

Problem definition

An ‘antifouling coating formulation,’ required to be formulated with the technology that is described previously, consists of acrylic copolymers as the binder, a biocidal pigment, and xylene as the solvent. Xylene is chosen as the solvent owing to its versatility and ability to dissolve acrylic copolymers. The goal is to find a molecule that would regulate the leaching of the biocidal pigment. Accordingly, this molecule should be such that it has a good solubility in xylene. Besides, it should be slightly soluble in seawater. Moreover, as per the REACH regulations by the European Chemical Agency, the aquatic toxicity of chemicals discharged into the water bodies should be low, and the primary biodegradation time should be ‘less than a month.’ Let this molecule be named ‘X’.

Problem solution

The solution to this problem can be found using CAMD through the ProCAMD tool in ICAS [71]. As a result of running this algorithm, all feasible candidates that satisfy the specified property and structural constraints are generated. The ‘total HSP’ for xylene (the carrier solvent for the antifouling coating) is 18.1 MPa½. Therefore, for ‘X’ to be soluble in xylene, the ‘total HSP’ of this molecule should be in the range of ±3 MPa½ of this parameter for xylene. It is also known that rosin is a commonly used molecule that serves the function of regulating the leaching behavior in antifouling paints. Therefore, it can be said that the seawater solubility of ‘X’ should be around that of rosin. Here, it is to be noted that the solubility in seawater taken into account at this stage is the same as its solubility in pure water, that is, the effect of salinity and pH on the solubility is not taken into account. Moreover, for a low aquatic toxicity, the \( LC_{50,FM} \) value for fathead minnow must exceed 100 mg L\(^{-1} \) (or -log \( LC_{50,FM} \) (log mol.L\(^{-1}\))<3.3). The ‘primary biodegradation time score’ is a metric developed by the United States Environmental Protection Agency (EPA) to quantify the rate of aerobic biodegradation. It represents the time taken for a chemical to biodegrade such that there is a loss of parent chemical identity under aerobic conditions. The values of this score are in the range of 1–5 as follows: wherein 5 = hours; 4 = days; 3 = weeks; 2 = months; 1 = longer. Because as per the European Chemical Agency guidelines for ready biodegradability, the primary biodegradation time should be less than 28 days, the range for the ‘primary biodegradation time score’ is chosen in between 2 and 5. The qualitative needs, target properties, and constraints on them are summarized in Table 8.
EChA, European Chemical Agency; US, United States.

Considering the building blocks of acyclic hydrocarbons, alcohols, esters, aldehydes, ketones, ethers, and amines, 13,337 candidates are generated and screened using GC property models. Four molecules satisfy the target property constraints. Their properties are listed in Table 9 whereas their molecular structures are shown in Figure 6.

### Cosmetics and personal care

#### Background

A standard ‘nail polish and enamel remover’ comprises an AI that is responsible for dissolution of the nail polish, a solvent or diluent serving as the carrier medium, and an additive that provides a soothing aroma to the formulation. The AI should be such that it can completely solubilize the nail polish and remove it from the applied surface, that is, the keratin on the nails.

A typical nail polish contains a film former, such as nitrocellulose or cellulose acetate butyrate, to make the product hard and shiny when it dries. To make the film tough and resilient, a resin or secondary film former such as tosylamide/formaldehyde resin or tosylamide/epoxy resin is used.

#### Problem definition

To formulate a ‘nail polish and enamel remover,’ the most important step is the selection of the AI. Therefore, considering that a nail polish is made up of the nitrocellulose and epoxy resins, the AI in the remover should be such that it dissolves these two resins. Moreover, the formulation should be easily flowable and nontoxic.

#### Problem solution

To formulate the nail polish remover, the needs, target properties, and the constraints on the AI are listed in Table 10.

AI, active ingredient.

A database of 3078 ‘polar nonassociating compounds’ available in the database manager of ICAS was chosen. One-hundred and ninety five compounds from this database satisfy the constraint on the Hildebrand solubility parameter mentioned in Table 10.

It is known that the relative energy difference (RED) for a polymer-AI combination is the ratio of the ‘Hansen distance between the polymer and the AI’ and ‘the radius of solubility of the polymer.’ This is calculated using the equation as follows:

\[
\text{RED}_{\text{pol} - \text{AI}} = \frac{R_{\text{d,pol} - \text{AI}}}{R_{\text{d,pol}}} = \sqrt{4(\delta_{\text{d,pol}} - \delta_{\text{d,Al}})^2 + (\delta_{\text{p,pol}} - \delta_{\text{p,Al}})^2 + (\delta_{\text{h,pol}} - \delta_{\text{h,Al}})^2} < 1
\]
where, $\delta_{d,pol}$, $\delta_{p,pol}$ and $\delta_{h,pol}$ are the dispersive, polar, and hydrogen-bonding components of the HSPs.

The list of 195 compounds was tested for the condition $\text{RED}_{pol} - \text{AI}$ for both nitrocellulose and epoxy. The compounds that fulfilled the condition for both the polymers, that is, the ones that lie in the overlapping region of the solubility spheres of the two polymeric resins, qualify for being possible AIs of the formulation. A diagrammatic representation of these AIs in the form

![Diagram of compounds](image)

Table 9

<table>
<thead>
<tr>
<th>Result No.</th>
<th>log ($W_i$) at 298 K (log (mg/l))</th>
<th>$\delta_{AV}$ at 298 K (MPa$^{1/2}$)</th>
<th>$T_m$ (K)</th>
<th>Primary biodegradation probability score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$-2.6$</td>
<td>20.53</td>
<td>323.76</td>
<td>3.93</td>
</tr>
<tr>
<td>2</td>
<td>$-1.62$</td>
<td>20.93</td>
<td>336.13</td>
<td>3.69</td>
</tr>
<tr>
<td>3</td>
<td>$-3.12$</td>
<td>20.26</td>
<td>327.7</td>
<td>3.91</td>
</tr>
<tr>
<td>4</td>
<td>$-2.42$</td>
<td>20.94</td>
<td>334.88</td>
<td>3.95</td>
</tr>
</tbody>
</table>

Table 10

<table>
<thead>
<tr>
<th>Needs</th>
<th>Target properties</th>
<th>Target property constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissolve the nail polish</td>
<td>Hildebrand solubility parameter, $\delta$</td>
<td>$18 &lt; \delta$ (MPa$^{1/2}$) &lt; 22, $\text{RED}_{pol} - \text{AI}$ (&lt; 1$</td>
</tr>
<tr>
<td></td>
<td>‘Hansen distance between the polymer and the AI’ ($\text{RED}_{pol} - \text{AI}$)</td>
<td></td>
</tr>
<tr>
<td>Easily flowable</td>
<td>Viscosity ($\mu$)</td>
<td>$0.6 &lt; \mu$ (cP) $&lt; 0.9$</td>
</tr>
<tr>
<td>Non-toxic</td>
<td>Aquatic toxicity ($\text{LC}_{50,FM}$)</td>
<td>$-\log \text{LC}_{50,FM} &lt; 3.3$</td>
</tr>
</tbody>
</table>
of a 3D Hansen solubility sphere plot is shown in Figure 7.

Furthermore, the viscosity and the Lethal Concentration to kill 50% of Fathead Minnow population (LC50,FM) values for the compounds found previously are calculated using GC methods available in the ProPred tool of ICAS. Their properties are also listed in Table 11.

<table>
<thead>
<tr>
<th>Active Ingredient</th>
<th>Properties</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>δd (MPa1/2)</td>
</tr>
<tr>
<td>Acetone</td>
<td></td>
<td>15.5</td>
</tr>
<tr>
<td>Methyl ethyl ketone</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Ethyl acetate</td>
<td></td>
<td>15.8</td>
</tr>
<tr>
<td>Cyclohexanone</td>
<td></td>
<td>17.8</td>
</tr>
<tr>
<td>Acetophenone</td>
<td></td>
<td>19.6</td>
</tr>
<tr>
<td>Propylene 1,2-carbonate</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

Acetophenone has a high toxicity, whereas cyclohexanone and propylene 1,2-carbonate have a high viscosity. Therefore, acetone, methyl ethyl ketone, or ethyl acetate could be suitable AIs for the nail polish remover. Moreover, if a solvent that could significantly lower the viscosity of the full formulation to meet those specified in the target property constraints, then propylene 1,2-carbonate could also be a potential AI for the formulation.

**Conclusions**

The development of systematic computational frameworks for the design of complex formulated products can contribute toward the faster development of products through the screening of many alternatives. The needs of the product and their translation to measurable specifications are always the first steps for a successful product design project. CAD methodologies require a combination of tools (extensive databases, algorithms, and predictive property estimation methods) but when available, they can be applied to a wide range of design problems for blended products, formulations, and so on. Sustainability and environmental principles can also be incorporated by implementing estimation methods for the corresponding properties. The accuracy of the final design and the overall applicability of the method and associated tools will depend on the application range of the property estimation models used and the general methodology and the satisfactory solution of the mathematical problem. The choices obtained will in essentially all cases need to be checked against experimental data or simply identify product choices or formulations which must be validated experimentally. It can be demonstrated that the screening of alternatives through a CAD can save time and resources, and the optimal product candidates can be identified. Through experimental design, the weak points of CAD can be identified, and suggestions for improvement can be provided.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgement

Financial support from the Hempel Foundation to CoaST (The Hempel Foundation Coatings Science and Technology Centre) is acknowledged.

References


