The design of optimal mixtures from atom groups using Generalized Disjunctive Programming

Suela Jonuzaj\textsuperscript{a}, Aparana Gupta\textsuperscript{a,b}, Claire S. Adjiman\textsuperscript{a,*}

\textsuperscript{a} Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK
\textsuperscript{b} Department of Chemical Engineering, Indian Institute of Technology Delhi, India

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 29 September 2017
Revised 22 January 2018
Accepted 23 January 2018
Available online 3 February 2018

\textbf{Keywords:}
Mixture design
Generalized Disjunctive Programming
UNIFAC groups
Crystallization
Antisolvent
Liquid–liquid extraction

\textbf{A B S T R A C T}

A comprehensive computer-aided mixture/blend design methodology for formulating a general mixture design problem where the number, identity and composition of mixture constituents are optimized simultaneously is presented in this work. Within this approach, Generalized Disjunctive Programming (GDP) is employed to model the discrete decisions (number and identities of mixture ingredients) in the problems. The identities of the components are determined by designing molecules from UNIFAC groups. The sequential design of pure compounds and blends, and the arbitrary pre-selection of possible mixture ingredients can thus be avoided, making it possible to consider large design spaces with a broad variety of molecules and mixtures. The proposed methodology is first applied to the design of solvents and solvent mixtures for maximising the solubility of ibuprofen, often sought in crystallization processes; next, antisolvents and antisolvent mixtures are generated for minimising the solubility of the drug in drowning out crystallization; and finally, solvent and solvent mixtures are designed for liquid–liquid extraction. The GDP problems are converted into mixed-integer form using the big-M approach. Integer cuts are included in the general models leading to lists of optimal solutions which often contain a combination of pure and mixed solvents.

© 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

The importance of mixture and product design has long been established across the field of chemical engineering (Gani and Ng, 2015). The design of candidate chemicals, be it a single molecule, a mixture or formulation, at minimum cost and time, is essential in improving process and product performance, but is challenging because it requires finding the optimal number, identities and compositions of mixture components and using nonlinear property models. The use of trial-and-error methods to find the most appropriate combination of chemicals and processing technology that will lead to the desired product/process attributes only allows a limited exploration of the design space due to the effort, cost, and time required. Thus, the development of systematic methodologies, tools, and strategies that combine predictive property models within a computer-assisted search for mixture and product design is crucial to sustainability in a highly competitive environment. Computer-Aided Mixture/blend Design (CAM\textsuperscript{D}) (Achenie et al., 2003; Gani, 2004) is a promising approach for identifying suitable mixtures or blends that meet predefined target properties and optimise a given performance measure.

Several methodologies based on a CAM\textsuperscript{D} framework for the design of mixtures/blends have been reported in the literature and have been applied to a wide range of applications, varying from separation processes to product design. Among the previous methodologies, substantial work has been done in the area of solvent and solvent mixture design for separation precesses (Austin et al., 2017, 2016; Buxton et al., 1999; Cignitti et al., 2015; Gopinath et al., 2016; Jonuzaj and Adjiman, 2016, 2017; Jonuzaj et al., 2016; Karunanithi et al., 2005; Sinha et al., 2003; Zhang et al., 2015) and chemical reactions (Adjiman et al., 2008; Austin et al., 2017; Follić et al., 2007; Siougkrou et al., 2014; Struebing et al., 2013, 2017). Other approaches have been applied to the design of refrigerants (Churi and Achenie, 1997; Duvedi and Achenie, 1997; Vaidyaraman and Maranas, 2002) and polymer blends (Solvason et al., 2009; Vaidyanathan and El-Halwagi, 1996; Zhang et al., 2015), while some recent studies have addressed the design of heat transfer fluid mixtures (Lampe et al., 2014; Lee and Mitsos, 2017; Mavrou et al., 2015a, 2015b; Molina-Thierry and Flores-Tlacuahuac, 2015; Papadopoulos et al., 2013), blended liquid products (Yunus et al., 2013, 2014), emulsified chemical products (Mattei et al., 2012, 2014) and the design/selection of mixtures for bio-processes...
(Hernández et al., 2017; Ng et al., 2015). Extended reviews of existing methodologies for mixture/product design and their applications are given by Jonuzaj et al. (2016), Gani (2004) and Gani and Ng (2015) - only a few recent publications are described in this paper.

In recent work, Austin et al. (2017) proposed a mixture design approach based on COSMO-RS (Klamt, 1995) and COSMO-SAC (Lin and Sandler, 2002) to predict mixture thermodynamic properties (i.e., molecular volumes, sigma profiles and sigma moments). The mixture problem was decomposed into molecular design and mole fraction subproblems, where the design variables were projected on a lower dimensional space, that of the sigma moments for each unknown component in the mixture. The design problems were solved using derivative-free optimization algorithms described in earlier work from the same group (Austin et al., 2016) which enabled the reduction of the search space. This design approach was applied to the design of solvents and binary solvent mixtures for liquid–liquid extraction and chemical reactions.

A four-step methodology was proposed by Cignitti et al. (2015) for the design of single compounds, mixtures or blends. Within this computer-aided product design framework the product needs and target properties were first defined and then converted into a CAMD or CAMPD formulation. Next, the CAMD/CAMPD problem was formulated as a MINLP problem which was finally solved by employing a decomposition-based algorithm proposed by Karunanithi et al. (2005). The proposed approach was demonstrated though a solvent design case study for a single-stage liquid extraction process. The same four-step framework was adopted by Zhang et al. (2015) in a CAMD model, where both first- and second-order groups were considered simultaneously in the MILP/MINLP formulations in an attempt to avoid excluding optimal solutions from the feasible region. The proposed CAMD model was applied to solvent, polymer and surfactant design case studies.

A general CAMPD methodology was proposed by Jonuzaj et al. (2016) for the design of mixtures and products, based on the use of Generalized Disjunctive Programming (GDP) (Raman and Grossmann, 1994), for formulating the CAMPD problem. The design approach considered for the first time the simultaneous determination of the optimal number of mixture ingredients, their optimal identities and compositions. The desired components participating in the final mixture were selected from a list of candidate molecules. Since the number of ingredients in the optimal mixture were not defined a priori, optimal solutions may consist of a pure compound or a mixture of compounds. Different relaxation techniques of the GDP mixture problems, including the Big-M (BM) approach (Nemhauser and Wolsey, 1999) and Hull Reformulation (HR) (Lee and Grossmann, 2000), were investigated in later work (Jonuzaj and Adîman, 2017) in order to address the challenges of formulating and solving generalised mixture problems. The design methodology and the two different solution strategies were applied to the design of solvent mixtures for separation processes.

The design and selection of working fluid mixtures for Organic Rankine Cycles (ORC) has received increased attention in recent years. A critical review on designing or selecting fluids and fluid mixtures for ORC is given by Linke et al. (2015). Here, we discuss briefly recent studies that focus on the design/selection of fluid mixtures. Mavrou et al. (2015b) studied the use of different working fluid mixtures in solar ORC systems employing Flat Plate Collectors (FPC) with heat storage. The performance of several binary mixtures was assessed based on a multi-criteria methodology where several operating and design parameters of the solar ORC system were considered. The working fluid mixtures investigated in their work were either selected from predefined lists of common fluids used in ORC, or designed based on the methodology proposed by Papadopoulos et al. (2013). In later work the authors (Mavrou et al., 2015a) conducted a sensitivity analysis for the selection of working fluid mixtures, in which the simultaneous selection of optimal mixtures and ORC system specifications under operating variability was addressed.

Molina-Thierry and Flores-Tlacuahuac (2015) also studied the selection of multicomponent working fluid mixtures for ORC. Within their approach, the optimal identities and compositions of fluids in the mixture, as well as the optimum processing conditions of the Rankine cycle were determined simultaneously. The working fluids that participate in the optimal mixture were selected from a predefined set of commonly used organic compounds. Interestingly, although the number of components in the mixture was not a decision variable and the design problem considered mixtures with three, six and eleven ingredients, the best solutions were obtained for binary mixtures or mixtures with four compounds. The optimal mole fractions of the other compounds in the mixture were found to be zero. This approach is different to the general mixture design methodology proposed by Jonuzaj et al. (2016), where the number of mixture ingredients is a decision variable. In most mixture design problems, setting the decisions variables to zero can create numerical difficulties due to the presence of highly nonlinear phase equilibrium relations (the natural logarithms of mole fractions, which are a frequent feature, are not defined at zero). If the proposed approach was applied to problems where singularities could not be avoided, then only solutions in which all the compounds would be present in the final mixture could be identified and thus, the overall performance would be affected. A mixture of six or eleven components, for example, would yield quite different results compared to an optimal binary mixture.

In a different approach, Lee and Mitsos (2017) proposed a hybrid two-stage methodology for the selection of working fluid blends for ORC that yield maximum recovery of cryogenic energy from a liquefied natural gas (LNG) evaporation process. Within their sequential approach, an MINLP problem, where optimal working fluids are selected from a large list of candidate compounds, is first formulated and solved with a stochastic solver. Next, the composition of a ternary fluid mixture and the operating conditions are optimized as an NLP problem, which is solved to global optimality using a deterministic algorithm. The proposed decomposition-based method gradually reduces the search space, which can ease the solution of complex mixture problems and yield promising solutions.

In the area of chemical product design, Mattei et al. (2012) conducted a study for the design of emulsion-based chemical products, such as pharmaceutical and household products. The proposed systematic methodology consisted of a seven-step hierarchical framework based on computer-aided methods and tools. In the first step, main and secondary consumer needs were identified and converted into target properties. In the following five steps the formulation was built up by adding, one-by-one, different classes of chemicals that satisfy a set of relevant target properties. In particular, suitable active ingredients (AIs) that meet the main needs of the product were selected from databases in step 2, followed by the identification of dispersed and continuous phase solvents in steps 3 and 4, respectively. The solvents were used to dissolve the selected AI and other chemicals present in the formulation, ensuring that the product forms a single liquid phase. Step 5 consists of the selection of suitable emulsifiers (usually surfactants) to ensure the desired formation and stability of the emulsion. Suitable additives were finally added in step 6 to enhance the quality of the formulated product by satisfying the secondary needs. In the seventh and last step, the overall composition of all selected ingredients was determined. The proposed comprehensive approach was further investigated in later work by the authors (Mattei et al., 2014), where the selection and design of surfactants, as well as the determination of some key properties (e.g., cloud point) were highlighted. The design methodology was illustrated through two case
studies for the design of an emulsified UV sunscreen and an emulsified hand-wash.

In the field of bioprocesses, Hernández et al. (2017) presented a methodology for selecting suitable organic wastes and waste blends for the optimal production of biofuels using anaerobic digestion. Within this study, the optimal blend of biomass wastes that yielded the best composition of biogas, which was used as a source of energy and to produce other chemicals (DME, methanol, ethanol), was determined. The design procedure consisted of two main parts: (a) the biogas production from waste via anaerobic digestion and (b) the use of the generated biogas as a fuel after removing CO₂, or to produce biosynrgas via dry reforming. The use of biogas as a source of energy or for producing other chemicals is affected by the different types of feedstocks and their composition. The optimal waste or waste blends were chosen from a variety of feedstocks including cattle slurry and cattle manure, pig slurry and pig manure, urban food waste, urban green waste and sludge, based on the cost of biomasses and the credit obtained from the digestate. Two different models were used to estimate the digestate price and they yielded two optimal feedstock sources for the production of biogas, including sludge and a blend of 65% cattle slurry and 35% urban food waste.

Formulating molecular and mixture problems without restricting the design space has proved challenging because it requires finding an optimal molecule or mixture (defining identities, number and compositions of ingredients) using nonlinear property models, which results in highly nonlinear problems with a large combinatorial space. In order to circumvent the combinatorial explosion in the number of possible designs and facilitate the solution of mixture problems, most studies in the literature have thus been focused on a reduced version of the general CAM²D problem.

In several methodologies, the identities of some or all components that participate in a mixture are selected from a given set of candidate molecules. Within these approaches there is a risk of excluding from the design space molecular structures which exhibit poor performance as pure compounds, but, when combined in a mixture, can lead to better performance. Only a few studies, such as the work of Austin et al. (2017, 2016), Buxton et al. (1999) and Papadopoulos et al. (2013), address the design of all mixture ingredients from functional groups rather than from a list, using predictive methods such as UNIFAC (Fredenslund et al., 1975), COSMO-RS (Klamt and Schüürmann, 1993) or COSMO-SAC (Lin and Sandler, 2002). Within these more general methodologies only problems where the number of mixture ingredients is fixed to two (binary mixtures) has been considered. The restriction to binary mixtures results in a simpler design problem, but restricts the design space.

In the context of general mixture design, Jonuzaj et al. (2016) addressed the design of multicomponent mixtures by proposing a generalised design approach where the number of components in the mixture is a decision variable. In this approach, however, the mixture ingredients were chosen from a given list of candidate compounds. This is consistent with some existing design practices, especially in the pharmaceutical industry where the list of organic solvents that can currently be used in a drug manufacturing is restricted due to toxicological reasons, as they cannot be completely removed from the final product (Grodowska and Parczewski, 2010). Nevertheless, there is a long-term need to identify new components, as new environmental and safety regulations reduce the lists of allowed compounds (REACH, 2017). The ability to design processing fluids (solvents, heat-transfer fluids) and products from functional groups can help to challenge established practices and identify innovative solutions.

The purpose of the present work is to create a comprehensive approach that bridges the concepts of molecular and mixture design by finding at the same time the optimal number of ingredients, their proportions, which specific molecules should be used, and what atom groups are required. This is achieved by presenting a general CAM²D methodology for the design of molecules and mixtures based on Generalized Disjunctive Programming. Within this approach, all discrete decisions (the number of components, the composition of each component in terms of functional groups) are represented by Boolean variables. The optimal values of these decisions, together with continuous variables such as mixture composition, are made simultaneously in the proposed approach. By designing the pure components and mixture ingredients from functional groups, the selection of compounds from a limited set of choices is avoided, making it possible to consider larger design spaces that consist of a wide variety of molecules and mixtures. This means that we can take full advantage of the non-ideality of mixing promising “good” candidates with “bad” candidate molecules to achieve better performance (e.g., higher solubility of paracetamol can be achieved with a binary mixture of a good solvent like acetone and a bad solvent like water than with pure acetone (Granberg and Rasmuson, 2000)). Furthermore, by not fixing a priori the number of mixture constituents, the explicit evaluation of every choice of the number of components is avoided and the best solution is identified without specifying in advance whether a pure compound or a mixture is sought. The general mixture design problem expressed within a mathematical framework using GDP is converted into mixed integer form using the big-M approach prior to its solution.

The design methodology is demonstrated through three example problems (case studies 1a, 1b and 2) where the design of solvents/antisolvents and solvent/antisolvent mixtures for separation processes, namely crystallization and liquid extraction, is presented. In case study 1a, the design of optimal solvents and solvent mixtures that maximise the solubility of ibuprofen is discussed, whereas in case study 1b, the design of antisolvents and antisolvent mixtures that minimise the solubility of the drug is investigated. In case study 2, optimal solvents and solvent mixtures are determined to separate acetic acid from water in a single-stage liquid extraction process. Integer cuts are introduced in the general mixture formulations and a list of optimal solutions (i.e., list of mixtures with different number, identity and compositions of ingredients) is obtained for each problem.

The paper is organized as follows. In Section 2, the mathematical formulations of the generalized mixture design problem are presented. In Sections 3 and 4, the proposed approach is applied to three example problems, followed by the main conclusions of this work which are summarized in Section 5.

2. The generalized mixture design problem

A general formulation of mixture design problems derived by integrating GDP into a CAM²D framework is described in this section. The GDP formulations where the molecules are designed from a set of atom groups are first presented, followed by the MINLP model obtained by reformulating the GDP problem via the big-M approach.

2.1. Definition of key index sets

The mixture problem is constructed in a systematic way and is presented here in the context of the formulation of a generalized model, where any number of components is designed up to a user-defined maximum value and the molecules are designed based on a list of atom groups subject to structural, pure component and mixture property constraints. The main design variables are the number of the components that participate in the mixture, their identities (which specific groups and how many groups are required) and their compositions.
The following index sets are defined in order to derive the problem formulations: the set \( I = \{1, \ldots, N_I\} \) represents the number of components in the mixture, with \( N_I \) being the total number of components. The final mixture consists of known fixed molecules that must be present in the mixture and of unknown components that are to be designed from atom groups. The number of fixed components in the mixture is defined as \( N' \), whereas the maximum number of components to be designed is \( N_{\text{max}} \). The designed components are described by the set \( N = \{1, \ldots, N_{\text{max}}\} \). The chemical identity of each molecule (fixed and unknown) is defined using functional groups (building blocks such as \( \text{CH}_2 \), \( \text{COOH} \)) that can be used in the calculation of relevant physical properties.

The groups are represented by the set \( G = \{1, \ldots, N_G\} \), where \( N_G \) is the total number of groups. The compounds to be designed can be branched or straight chain molecules (represented by groups \( g \in G \)), aromatic molecules or single group molecules. The aromatic and molecular groups included in the set \( G \) are represented by the sets \( \mathcal{G}_A = \{1, \ldots, N_{\text{G}_A}\} \) and \( \mathcal{G}_I = \{1, \ldots, N_{\text{G}_I}\} \), respectively, with \( \mathcal{G}_A \subset \mathcal{G} \). The design groups are further classified into main, functional, chain-ending and non-chain-ending groups, and are represented by the subsets \( \mathcal{G}_M, \mathcal{G}_F, \mathcal{G}_C, \mathcal{G}_NC \subset \mathcal{G} \), respectively. The type of molecular structures designed from functional groups \( g \) is represented by the set \( \hat{T}_g = \{a, m, b\} \), where \( a = \text{acyclic}, m = \text{monocyclic} \) and \( b = \text{bicyclic} \). When a monocyclic aromatic structure is designed, side chains can appear if branched aromatic groups are included in the designed component. The set of branched aromatic groups considered here is given by \( \mathcal{G}_B = \{1, \ldots, N_{\text{G}_B}\} \). Finally, \( \mathcal{K} = \{1, \ldots, N_K\} \) is a set used to define how many groups of a given type appear in a designed molecule. A set of 49 groups used previously for the design of solvents for chemical reactions (Struebing et al., 2017) is employed here to describe the design of molecules and mixtures with the proposed formulation, and it is included in Table B.1 in Appendix B for completeness. It is noted that the generalized model is valid for different sets of atom groups (smaller/larger lists or different types) which can be used in a wide range of applications.

### 2.2. GDP formulation of the CAMPD problem

#### 2.2.1. General formulation

In previous work Jonuzaj et al. (2016) proposed a general GDP formulation for mixture design where the number, identity and compositions of mixture ingredients are optimised simultaneously, with components that participate in the mixture being selected from a predefined list. Following their derivation, the GDP model of the general mixture problem where molecules are designed from atom groups is given below:

\[
\begin{align*}
\min_{x, Y, \hat{Y}} & \quad f(x) \\
\text{s.t.} & \quad r(x) \leq 0 \\
& \quad \hat{Y}_g, \quad g = 1, \ldots, N_g \\
& \quad h^f_{g}(x, \hat{y}) \leq 0, \quad i = N' + 1, \ldots, N_c \\
& \quad h^C_{g}(x, \hat{y}) \leq 0, \quad g = 1, \ldots, N_g \\
& \quad h^g(x) \leq 0, \quad g = 1, \ldots, N_g \\
& \quad h^r_{g}(x) \leq 0, \quad \forall g, \quad g = 1, \ldots, N_g \\
& \quad h^n_{g}(x) \leq 0, \quad \forall g, \quad g = 1, \ldots, N_g \\
& \quad n = 1, \ldots, N_{\text{max}} \\
& \quad n \leq n_{\text{max}} \\
& \quad \Omega(Y, \hat{y}, \hat{Y}) = \text{True} \\
& \quad g \in \mathcal{G} \\
\end{align*}
\]

where \( f \) is the objective function to be optimised and is a function of continuous variables represented by the \( m \)-dimensional vector \( x \), which denotes operating conditions, pure components and mixture properties (e.g., mole fractions, activity coefficients), or process variables (e.g., flowrates, temperatures). The set of inequalities \( r(x) \leq 0 \) represents general constraints that must hold regardless of the discrete choices. The functions included inside the disjunctions are vectors of linear/nonlinear conditional constraints that depend on the discrete decisions, as represented by Boolean variables. Two different sets of disjunctions are included in this formulation. The first set involves disjunctions for designing each component from a set of functional groups, \( G \), and the identity of a group \( g \) in a designed component \( i \) in the mixture is determined through a Boolean variable \( Y_{i,g} \) (a group \( g \) is present in a component \( i \) if \( Y_{i,g} \) is True and it is not if \( Y_{i,g} \) is False). The presence of groups in a designed component \( i \) is represented through inclusive OR disjunctions (Grossmann and Trespalacios, 2013) \( \bigvee_{g \in \mathcal{G}} Y_{N'+1,g} \), which means that at least one group is present in the first designed component \( N' + 1 \) (i.e., at least one molecule is designed). The conditional constraints included in the first set of disjunctions consists of chemical feasibility constraints (e.g., the octet rule to ensure that the designed molecule has zero valency Odele and Macchietto, 1993), represented by the vector \( h^f_{g}(x, \hat{y}) \leq 0 \); chemical complexity constraints (e.g., constraints to ensure that the number of groups of specific type are within lower and upper limits), given by \( h^C_{g}(x, \hat{y}) \leq 0 \); design constraints (e.g., constraints that ensure the designed molecule/mixture meets specific physical property or health and safety criteria), given by \( h^r_{g}(x) \leq 0 \); and pure component (e.g., constraints to predict critical temperature, pressure and volume, normal boiling and melting points, heat of fusion, etc.) and mixture property constraints (e.g., solubility, miscibility, selectivity and distribution coefficient functions), represented by \( h^g(x) \leq 0 \). The different sets of constraints are described in detail in the following sections. The binary variables \( \hat{y} \) included in chemical feasibility and chemical complexity constraints define the types of molecular structures being designed (acyclic, monocyclic, bicyclic) and indicate the presence of branches in aromatic molecules, as will be described in the next sections. These constraints could also be formulated as embedded (inner) disjunctions (Grossmann and Trespalacios, 2013) determined by Boolean variables \( \hat{Y} \) and included within the disjunctions for the design of mixture ingredients from groups. This disjunctive representation, often seen in superstructures of PSE problems (Caballero et al., 2006; Grossmann and Trespalacios, 2013; Yeomans and Grossmann, 1999), is avoided here by incorporating directly binary variables into the equations to formulate the discrete decisions.

The second set of disjunctions involves conditional constraints that depend on the number of mixture ingredients, \( n \), and that are active when the corresponding Boolean variable \( Y_{n} \) is True. \( \hat{F}_{n}(x) \leq 0 \) is a vector of all the linear/nonlinear constraints that de-
pend on the number of components in the mixture. Common examples of these functions include mixture property and/or process design constraints, such as phase equilibrium relations (e.g., solid-liquid equilibrium (SLE), liquid-liquid equilibrium (LLE)), stability functions (e.g., miscibility equation Gani et al., 1991 to ensure that the designed solvent mixture is in a single liquid phase at given temperature and pressure), mass mole balances of a process unit, etc., which are described in Sections 3 and 4. The mole fraction, $x_i$, of component $i$ in the final mixture is set to be greater than a user-defined value $x_i^0$ if the component is present in the mixture and zero otherwise. This prevents the generation of mixtures with unrealistic compositions, in which some components are present in very small quantities. Exactly one disjunction (i.e., exclusive OR disjunctions) for the number of designed components must be active; this is ensured by the expression $\sum_{n \in N} \tilde{y}_n = 1$, i.e., exactly $n$ ($n = 1, \ldots, N_{max}$) mixture ingredients are designed when $\tilde{y}_n$ is True.

The logic conditions, $\Omega(Y, \tilde{y}, \hat{y}) = True$, are in the form of propositional logic relations involving Boolean and auxiliary variables, and express the relationships between the disjunctive sets. A set of logic relations is included in this model to avoid degeneracy of the solutions by ensuring that no two identical components are designed in the mixture. Additional logic propositions are required to relate the Boolean variables for the number of designed components in the mixture, $\tilde{y}_n$, $n = 1, \ldots, N_{max}$, to the Boolean variables for the identity of each group $g$ in designed molecule $i$, $Y_{i,g}$, $i = N' + 1, \ldots, N$, $g = 1, \ldots, N_g$. Similar relations are derived to relate the Boolean variables $\tilde{y}_n$ to the discrete variables $\hat{y}$ that define the specific structures of the designed molecules. All logic relations can be translated into linear algebraic equations by replacing the Boolean variables $Y_{i,g}$ and $\tilde{y}_n$ with the binary variables $\hat{y}_{i,g}$ and $\tilde{y}_n$, respectively, and applying Boolean algebra rules (Raman and Grossmann, 1991; Williams, 1985). All the logic propositions and their equivalent algebraic equations are summarized in Table A.1 in Appendix A. A more detailed description of the derivation of logic relations for GDP mixture models and how they are converted into linear algebraic inequalities can be found in Jonuzaj et al. (2016).

### 2.2.2. Chemical feasibility and complexity constraints

The design of pure solvents from atom groups was recently studied by Struebing et al. (2017) and was successfully applied to chemical reactions. The chemical feasibility and chemical complexity constraints proposed in their study are modified and employed in this work for general mixture design and they are described in the following paragraphs.

The chemical feasibility constraints ($n'_{i,g}(x, \hat{y}) \le 0$) are derived to ensure that only chemically-meaningful molecules are designed. Such constraints include equations for the design of cyclic, acyclic, aromatic and single group molecules, the octet rule (Odele and Macchietto, 1993) and the modified bonding rule (Buxton et al., 1999). Thus, the following equation allows only the design of acyclic, monomeric or bicyclic molecules:

$$\sum_{i \in I} \hat{y}_{i,a} = 1, \; i \in I$$

where the binary variable $\hat{y}_{i,a}$ is used to define the type of molecular structure being designed, with $\hat{y}_{i,a} = 1$ indicating the design of an acyclic molecule, $\hat{y}_{i,m} = 1$ the design of a monocyclic molecule and $\hat{y}_{i,b} = 1$ the design of a bicyclic molecule. When designing cyclic molecules only aromatic compounds are considered in this work and thus the following constraint is derived based on the fact that exactly six aromatic groups are necessary to create a monocyclic molecule and ten aromatic groups are necessary to create a bicyclic molecule:

$$\sum_{g \in I_{a}} n'_{i,g} - 6\hat{y}_{i,m} - 10\hat{y}_{i,b} = 0, \; i \in I$$

where $n'_{i,g}$ is the number of groups of type $g$ that appear in component $i$. The variable $n'_{i,g}$ should take only integer values and thus, it is converted into a pseudo-integer variable as follows (Floudas, 1995; Struebing et al., 2017):

$$n'_{i,g} = \sum_{k \in K} 2^{k-1} \tilde{y}_{i,g,k}, \; i \in I, \; g \in G$$

where the binary variable $\tilde{y}_{i,g,k}$ relates to group $g$ in molecule $i$ and the maximum number of groups of type $g$ that appear in a designed molecule $i$ is set via a user-specified value, $K$. If the value of $K$ is chosen to be 3 ($k = \{1, 2, 3\}$), for example, then the maximum number of occurrences of group $g$ in a molecule is 7. The binary variable $\tilde{y}_{i,g,k}$ is related to the binary variable $y_{i,g}$ which represents the presence or absence of a group $g$ in a designed component, as shown below:

$$y_{i,g} \le \sum_{k \in K} \tilde{y}_{i,g,k} \; \text{and} \; \tilde{y}_{i,g,k} \le y_{i,g} \; i \in I, \; g \in G, \; k \in K$$

The following octet rule was proposed by Odele and Macchietto (1993) to ensure that the designed molecule does not have any free bonds:

$$\sum_{g \in G} (2 - v_{g})n'_{i,g} - 2m_{i} = 0, \; i \in I$$

where $v_{g}$ is the valency of group $g$ (cf. Table B.3 in Appendix B). $m$ is a vector of continuous variables, the values of which are determined by the design of acyclic and cyclic (monomeric–bicyclic) molecules, using the relation $m_{i} = \hat{y}_{i,a} - \hat{y}_{i,b}, \; i \in I$. Thus, $m_{i}$ takes the value of 1 if an acyclic compound is designed, 0 if a monomeric compound is designed and −1 if a bicyclic compound is designed. It should be noted that Eq. (5) does not prevent a single-molecule group ($g \in G_1$) from appearing alongside a combination of groups forming a chemically-feasible molecule since the valency of molecular groups is zero and thus the octet rule is satisfied. In order to avoid the design of such molecules, we employ the following constraints (Struebing et al., 2017) to ensure that no other groups are included in the design when a single-molecule group is selected:

$$\sum_{g \in G} n'_{i,g} \le 1, \; i \in I$$

$$\sum_{g \in G} n'_{i,g} - \sum_{g \in G_1} n'_{i,g} \le \left( 1 - \sum_{g \in G_1} n'_{i,g} \right) n'_{i,g}^{\text{max}} \; i \in I$$

where $n'_{i,g}^{\text{max}}$ is the maximum number of a group $g$ that can appear in each designed component. The presence of single-molecule groups is also considered in the following modified bonding rule proposed by Struebing et al. (2017), based on the modified bonding rule of Buxton et al. (1999) and bonding rule of Odele and Macchietto (1993):

$$n'_{i,g}(m_{i} - 1) + 2 \left( m_{i} - \sum_{g \in G} n'_{i,g} \right) - \sum_{g \in G} n'_{i,g} \le 0, \; i \in I, \; g \in G$$

Eq. (8) ensures that adjacent groups are connected by a single covalent bond since atom groups are defined so that double or triple bonds appear within atom groups (e.g., CH₂=CH).
and an upper bound on the number of groups that appear in the molecule, as shown below:

$$n_{i,g}^{\min} \leq \sum_{g \in G} n_{i,g} \leq n_{i,g}^{\max}, \ i \in I$$

(9)

where $n_{i,g}^{\min}$ and $n_{i,g}^{\max}$ are the minimum and maximum number of groups, respectively, in each designed component $i$. In addition, the number of groups of each type is constrained by a specific upper bound as shown in the following general form:

$$n_{i,g}^{\prime} \leq \tilde{f}_{i,g}(\hat{y}_{i,g}), \ i \in I; \ g \in G; \ t \in T_{M}$$

(10)

where $\tilde{f}_{i,g}(\hat{y}_{i,g})$ is a linear function that represents the upper limit derived from the possible structures in which each group can participate, i.e., occurrences of a specific group in acyclic, monocyclic and bicyclic molecules. The linear functions $f_{i,g}(\hat{y}_{i,t})$ are presented in Table B.4 in Appendix B for completeness. Further restrictions are imposed on the functional ($G_{F}$) subclass by Eq. (11):

$$\sum_{g \in G_{F}} n_{i,g}^{\prime} \leq \hat{y}_{i,a} + \hat{y}_{i,m}, \ i \in I$$

(11)

where $n_{i,g}^{\prime}$ denotes the upper limit on the number of functional groups of type $g$ in designed molecule $i$. The values of $n_{i,g}^{\prime}$ are presented in Table B.5 in Appendix B. Eq. (11) ensures that acyclic and monocyclic components consists of only a limited number of functional groups, whereas such groups do not appear in bicyclic molecules. The chemical stability of the designed compounds is also improved by allowing at most one carbon–carbon double bond group $g \in G_{CNC}$ in the molecule, as shown below (Struebing et al., 2017):

$$\sum_{g \in G_{CNC}} n_{i,g} \leq 1$$

(12)

The design of branched molecules is also considered in this work. In the case of cyclic molecules, the design of branched components requires the presence of an aromatic group $g \in G_{Br} = \{aC, aCCH, aCHC\}$ that links the aromatic ring with a side chain. The presence of each branched aromatic group in the designed component $i$ is denoted with a binary variable $\hat{y}_{i,g} (i \in I; g \in G_{Br})$, which is used to constrain the number of the branching groups in the designed molecule as follows (Folić et al., 2008):

$$\hat{y}_{i,g} - n_{i,g}^{\prime} \leq 0, \ i \in I; \ g \in G_{Br}$$

(13)

Thus, if a branched monocyclic or bicyclic compound is designed (i.e., $\hat{y}_{i,g} = 1$), the number of branching aromatic groups of type $g$ in molecule $i$, $n_{i,g}^{\prime}$, takes a positive value; otherwise it becomes zero. In the specific case of branched monocyclic components, another binary variable, $\hat{y}_{i,m,aC}$, is introduced to define whether an aC group is present in the designed molecule. $\hat{y}_{i,m,aC}$ takes the value of 1 if a monocyclic compound is designed (i.e., $\hat{y}_{i,m} = 1$) and an aC group is included in the set of groups that make up the molecule (i.e., $\hat{y}_{i,aC} = 1$). The relation between the three binary variables is expressed with the following algebraic form:

$$\hat{y}_{i,m} + \hat{y}_{i,aC} - \hat{y}_{i,m,aC} \leq 1, \ i \in I$$

(14)

or equivalently:

$$2\hat{y}_{i,m,aC} \leq \hat{y}_{i,m} + \hat{y}_{i,aC}. \ i \in I$$

(15)

In order to reduce the complexity of the designed branched molecules, the following constraints are derived to ensure that at most one side chain appears in a monocyclic compound (i.e., at most one aromatic group $g \in G_{Br}$ participates in the molecule), and monocyclic and bicyclic components consist of at most one and two aC groups, respectively:

$$\hat{y}_{i,m,aC} + \hat{y}_{i,m,aCCH} + \hat{y}_{i,m,aCCH_2} \leq 1, \ i \in I$$

(16)

The groups that appear in side chains are classified into chain-ending, $C_{CE}$, and non-chain-ending groups, $G_{NCE}$. A chain-ending group can be connected to a non-chain-ending group or attached directly to the aromatic backbone of the designed cyclic molecule. Both chain-ending and non-chain-ending groups can be connected to the same aromatic group aCCH, leading to the design of a more complex molecule with two branches. The complexity of such compounds is reduced by enforcing that one of the free bonds of aCCH be connected to CH$_3$, as shown in Eq. (18):

$$\hat{y}_{i,aCCH} - n_{i,CH_3} \leq 0, \ i \in I$$

(18)

Further restrictions on the structure of the designed molecules are introduced by forcing chain-ending and non-chain-ending groups to appear at most three times in an acyclic compound and at most once in an aromatic one, as shown in Eqs. (19) and (20), respectively:

$$\sum_{g \in G_{CE}} n_{i,g} \leq 3\hat{y}_{i,a} + \hat{y}_{i,m,aC} + \hat{y}_{i,aCCH} + \hat{y}_{i,aCCH_2}, \ i \in I$$

(19)

$$\sum_{g \in G_{NCE}} n_{i,g} \leq 3\hat{y}_{i,a} + \hat{y}_{i,m,aC} + \hat{y}_{i,aCCH_2}, \ i \in I$$

(20)

Note that $\hat{y}_{i,aCCH}$ does not appear in Eq. (20) as relevant restrictions are imposed via Eq. (18). Chemical complexity constraints are user-specified restrictions on the structure of the design molecules. They can be modified, reduced or increased based on heuristics or preferences of the user.

2.2.3. Design constraints

Design constraints ($h_{\ell}(x) \leq 0$) can be derived based on physical properties or health and safety performance of the candidate molecule/mixture. In solvent mixture design, for example, the designed mixture ingredients must be in liquid phase, which is enforced by constraining the normal boiling temperature ($T_{melt,g}$) and boiling temperature ($T_{boil,g}$) points of each solvent, as shown in Eqs. (21) and (22), respectively. The melting and boiling points of the designed components (left-hand-side of Eqs. (21) and (22)) are calculated based on the first-order group contribution (GC) method proposed by Marrero and Gani (2001):

$$\begin{align*}
\sum_{g \in G} n_{i,g} T_{melt,g} & \leq \exp\left(\frac{T_{max,melt,g}}{147.450}\right), \ i \in I \\
\sum_{g \in G} n_{i,g} T_{boil,g} & \leq \exp\left(\frac{T_{min,boil,g}}{222.543}\right), \ i \in I
\end{align*}$$

(21)

(22)

where $T_{max,melt,g}$ and $T_{min,boil,g}$ are user-defined maximum melting and minimum boiling temperatures, respectively; 147.450 and 222.543 are reference values used in the GC method of Marrero and Gani (2001); $T_{melt,g}$ and $T_{boil,g}$ are the contributions of each group $g$ to the melting and boiling points, respectively, and they are presented in Table B.3 in Appendix B.

2.2.4. Property constraints and physical model

Finally, the set of constraints represented by the vector $h_{\ell}(x) \leq 0$ consists of pure component or mixture property constraints, such as the liquid phase activity coefficient of each component in mixture (it participates in phase equilibrium relations, e.g., SLE, LLE), which is evaluated using group contribution methods (e.g., UNIFAC, Fredenslund et al., 1975). Some functions of the group contribution models depend on the identity of the group and thus are included in the disjunctions.
2.3. Reformulation of problem (G-GDP) as an MINLP via Big-M approach

The general GDP mixture problem (G-GDP) is converted into a mixed integer form using the Big-M approach as shown below:

\[
\begin{align*}
\min_{x, y, y_0} & \quad f(x) \\
\text{s.t.} & \quad r(x) \leq 0 \\
& \quad h_{ig}(x, y) \leq M_{h_{ig}} (1 - y_{ig}), \quad g = 1, \ldots, N_g, \quad i = N' + 1, \ldots, N_c \\
& \quad h_{ig}(x, y) \leq M_{h_{ig}} (1 - y_{ig}), \quad g = 1, \ldots, N_g, \quad i = N' + 1, \ldots, N_c \\
& \quad h_{ig}(x) \leq M_{h_{ig}} (1 - y_{ig}), \quad g = 1, \ldots, N_g, \quad i = N' + 1, \ldots, N_c \\
& \quad f_g(x) \leq M_{f_g} (1 - y_n), \quad n = 1, \ldots, N_{\text{max}} \\
& \quad x'_n y_n \leq x'_n y_{n+1}, \quad n = 1, \ldots, N_{\text{max}}, \quad i = N' + 2, \ldots, N_c \\
& \quad M_{h_{ig}} \geq x'_n y_{n+1} \\
& \quad \sum_{i=1}^{N_c} y_{n+1, i, g} \geq 1 \\
& \quad \sum_{n=1}^{N_{\text{max}}} y_n = 1 \\
& \quad Ay + By + Cy \leq d \\
& \quad x \in [x'^{1}, x'^{2}] \subset \mathbb{R}^m \\
& \quad y_{ig} \in [0, 1], \quad i = N' + 1, \ldots, N_c, \quad g \in G \\
& \quad y_n \in [0, 1], \quad n = 1, \ldots, N_{\text{max}} \\
& \quad y \in [0, 1]^N
\end{align*}
\]

(G-BM)

Valid upper bounds are derived for the big-M parameters \(M_{h_{ig}}, M_{h_{ig}}, M_{h_{ig}}, \) and \(M_{f_g}\), such that when the binary variables \(y_{ig}\) and \(y_n\) are zero the constraints are always satisfied. In order to avoid numerical difficulties arising from tight bounds and machine precision, the big-M parameters of nonlinear functions are relaxed bounds rather than exact bounds. Conditional constraints can be converted into mixed integer form either as equalities \((h(x) = 0)\) or as two inequalities with opposite signs \((\pm h(x) \leq 0)\), without loss of generality.

2.4. List of optimal solutions

The main objectives of computer-aided molecular and mixture/blend design is to identify promising molecules, mixtures or blends for various applications. When designing molecules from atom groups, the optimal structure generated from the CAMPOD model for a specific process or product could be a new compound that is difficult to synthesize or an existing but expensive compound. Thus, the second or even third optimal solution with different molecular structures may be required. Working towards generating a list of optimal molecules and/or mixtures, the following integer cut is introduced in the (G-BM) mixture design problem (GAMS Development Corporation, 2017a):

\[
\sum_{i=1}^{N_c} \sum_{g \in G} |n'_{ig} - n_{ig, l}| \geq 1, \quad l = 1, \ldots, N_i
\]  

(23)

where \(n_{ig, l}^{\text{opt}}\) is the optimal number of group of type \(g\) in molecule \(i\) generated in previous solutions \(l = 1, \ldots, N_i\). The advantage of including integer cuts in the general mixture problem is that it helps to obtain a list of diverse solutions which can be different molecular structures, or mixtures with different number and/or identities of components.

3. Case study 1: Solvent mixture design for the crystallization of ibuprofen

Crystallization is a key separation and purification unit in the manufacturing process of many pharmaceuticals (intermediates and/or final API products) as it offers high purity and recovery of chemicals, while maintaining low energy consumption and low environmental impact (Crafts, 2007). The careful design or selection of suitable solvents and solvent mixtures for crystallization is essential to ensure the high performance and economics of the process. Solubility is one of the key properties in crystallization and it is used as a main criterion for selecting molecules (Frank et al., 1999; Jonuzaj and Adjiman, 2017; Jonuzaj et al., 2016) or designing molecules (Austin et al., 2016; Karunanithi et al., 2006) that can be solvents and solvent mixtures for many crystallization processes, such as cooling and drowning out crystallization. Frank et al. (1999) reviewed several methods based on solubility to screen solvents from a database for the crystallization of organic solids. The design and selection of suitable solvents or solvent mixtures based on solubility for the crystallization of ibuprofen (a common anti-inflammatory drug), is a well-studied problem in the literature. Karunanithi et al. (2006) and Austin et al. (2016) have investigated the design of appropriate solvents and binary solvent mixtures for the crystallization of the drug, based on solubility, considering other factors such as flammability, viscosity and toxicity of the solvent and its effect on crystal morphology. Jonuzaj et al. (2016) have addressed the determination of multicomponent mixtures for maximizing the solubility of ibuprofen, with optimal solvents being selected from a predefined set of candidate compounds. In this work, we design optimal solvent and antisolvent (drowning out agent) mixtures with fixed and unknown number of mixture ingredients for maximizing and minimizing the solubility of ibuprofen, respectively, as described in the following subsections. The molecules are designed from atom groups (UNIFAC groups) and the solubility of the drug is considered to be the main criterion for design. It is noted that this case study is employed to illustrate the ability to design “solvent” from atom groups without making any assumptions on whether the “solvent” contains one or more compounds. We do not take advantage of the full possibility of the design methodology, since aspects such as the environmental impact of the solvent or the ease of recovery are not taken into account. These metrics can be integrated into a CAMPOD model, as shown in the work of Karunanithi et al. (2005).

3.1. Case study 1a: Solvent mixture design for maximising the solubility of ibuprofen

In all types of crystallization, achieving high solubility of the solute in a solvent or solvent mixture is an essential feature for the successful performance of the process, as solvent capacity has a strong impact on the throughput that can be achieved given a fixed equipment size. In cooling crystallization, for example, which is a commonly used separation mode in the fine chemicals industry, the solid compound of interest dissolves in a solvent mixture at high temperature and it is then recovered as pure solid from the mixed solution after decreasing the temperature so that the solute becomes much less soluble and crystallizes out of the solution (Karunanithi and Achenie, 2007). The selection/design of suitable solvents affects process efficiency and good solvents that yield high solubility of the solid compound are usually required, so that a sufficient amount of solid is dissolved without increasing the temperature to exceedingly high values. This may not be achieved if a small molecular design space is considered, e.g. if conventional compounds are selected from restricted databases.

In the proposed general approach, the design of molecules from a large set of atom groups (UNIFAC groups) is considered; this can
generate a very large and diverse chemical space. We apply the mixture design methodology to the design of solvents and solvent mixtures that maximise the solubility of ibuprofen, \( x_{\text{ibu}} \) (i.e., the mole fraction of ibuprofen in the mixture), at ambient conditions. A full description of the problem statement was given in previous work (Jonuzaj et al., 2016), where solvent molecules were selected from a small list, and here we present a brief summary of the main features of the case study. Ibuprofen is represented with atom groups as \( 4 \times \text{ACh}, 3 \times \text{CH}_2, \text{aCH}_2, \text{aCH}, \text{CH}, \text{COOH} \) and its solubility in the solvent mixture is calculated as follows (Gmeining et al., 1978; Sandler, 1999):

\[
\ln x_{\text{ibu}} + \ln y_{\text{ibu}} = \frac{\Delta H_{\text{m}}}{R} \left[ \frac{1}{T_m} - \frac{1}{T} \right]
\]

(24)

where \( y_{\text{ibu}} \) is the liquid phase activity coefficient of ibuprofen at composition \( x \), mixture temperature \( T \), and pressure \( P = 1 \) atm; \( T_m \) and \( \Delta H_{\text{m}} \) are the normal melting point and enthalpy of fusion of ibuprofen and their values are taken to be 347.15 K and 25.5 kJ/mol, respectively (Gracin et al., 2002); \( R \) is the gas constant. The following miscibility constraint for every binary pair of solvent molecules is employed in order to ensure that the solvents are mutually miscible in the proportions and at the temperature relevant to the mixture (Gani et al., 1991; Smith et al., 2001):

\[
\left[ \frac{\partial \ln y^{i,j}}{\partial x_{j}^{i,j}} \right]_{T,P} + \frac{1}{x_{j}^{i,j}} \geq 0 
\]

(25)

where \( y_{i,j}^{i,j} \) is the activity coefficient of the designed solvent component \( i \) in a binary solvent mixture of \( i \) and \( j \) and \( x_{j}^{i,j} \) is the mole fraction of component \( i \) in the mixture of \( i \) and \( j \) and can be calculated as \( x_{j}^{i,j} = x_i / \left( x_i + x_j \right) \), with \( x_i \) and \( x_j \) being the mole fractions of components \( i \) and \( j \), respectively, in the multi-component mixture. Although these constraints do not guarantee that the final mixture consists of a single stable liquid phase, it is much less likely to undergo phase separation if the binary pairs are stable. The activity coefficient of ibuprofen and of each designed solvent is estimated using the UNIFAC group contribution method (Fredenslund et al., 1975), and the formulation proposed by Smith et al. (2001), which is presented in a form convenient for programming, is employed in this work. These equations are given in Jonuzaj et al. (2016).

3.1.1. Problem formulations

In order to compare the proposed methodology with the standard (restricted) CAM\(^3\)P-D problem, we first apply a simplified version of the model to the case study, in which the number of solvents in the mixture is fixed. Next, the generalized model is solved, with all the decision variables (number, identity and composition of mixture ingredients) being optimised simultaneously. Hence, two formulations are considered: Formulation 1aR, a restricted problem, where we design mixtures with one \( (N = 1 - \) problem D1), two \( (N = 2 - \) problem D2) or three \( (N = 3 - \) problem D3) solvents; Formulation 1aG, a general problem, where the number of components in the mixture is bounded by a maximum number and we design mixtures with up to three solvents \( (N_{\text{max}} = 3 - \) problem D4). The possible components in the mixture, represented by the set \( I = \{ \text{ibu}, \text{C}_1, \text{C}_2, \text{C}_3 \} \), include ibuprofen (ibu), i.e., the fixed component in the mixture, and the designed solvent components \( \{ \text{C}_1, \text{C}_2, \text{C}_3 \} \). The solvent molecules are designed from a set of 49 atom groups \( (G = \{1, \ldots, 49\}) \), presented in Table B.1 in Appendix B, and each designed molecule \( i \) consists of at most 7 groups \( (n_{\text{max}} = 7) \). The objective here is to maximise the solubility of ibuprofen at 300 K and thus, the objective function is \( f(x) = x_{\text{ibu}} \). In both formulations (1aR, 1aG), the general constraints, \( r(x) \), consist of the solubility function (Eq. (24)) and those equations of the UNIFAC model that are expressed in terms of ibuprofen only (i.e., that do not depend on the identity or number of the designed solvent molecules); thus the general constraints do not depend on the logic decisions. It is noted that in Formulation 1aR, nearly all UNIFAC equations are treated as general constraints. The conditional constraints that depend on the identity of groups in each molecule and/or number of solvents in the mixture are included in the appropriate disjunctions, based on which formulation considered.

In the restricted problem of Formulation 1aR, \( Y_n \) is no longer a decision variable and thus the formulation consists only of disjunctions for assigning the number of groups of type \( g \) to components \( i \) in the mixture, which include the chemical feasibility, chemical complexity and design constraints (Eqs. (1)-(22)) described in Section 2.2; mixture property constraints, such as the stability function (Eq. (25)); and those equations from the UNIFAC model (e.g., the molecular van der Waals volume, \( r_i \), and molecular surface area, \( q_i \), of designed component \( i \)) used to predict the activity coefficient of each solvent component. The conditional constraints can be converted into MINLP using the big-M approach as described in the previous section. Here, however, a simpler MINLP reformulation can be constructed to facilitate the solution of the problem. As shown in Eq. (3), the number, \( n_{ig} \), of groups of type \( g \) in a designed molecule \( i \) is defined by using the binary variable \( y_{ig,k} \) (\( i \in I, g \in G, k \in K \)), whereas the rest of the equations within disjunctions are expressed in terms of the pseudo-integer variables \( n_{ig} \) and the binary variables \( y \). These relations force the number of groups \( n_{ig} \) and all the relevant variables that depend on the identity of the groups to become zero when no atom group of type \( g \) is present in the designed component \( i \) (i.e., \( n_{ig} = 0 \) when \( y_{ig,k} = 0 \)). Thus, the use of large big-M parameter values, which can sometimes lead to poor relaxations of the MINLP model, can be avoided in this simple case. Furthermore, the use of binary variables \( y_{ig,k} \) is sufficient to calculate the number of groups in the molecule and hence variables \( y_{ig,k} \), that represent the presence of a group in a designed component can be eliminated, reducing the size of the MINLP. Finally, logic relations are included in Formulation 1aR in order to eliminate degenerate solutions, so that the design of identical components as two distinct mixture components is avoided (see Table A.1).

The general problem of Formulation 1aG consists of both disjunctions for the identity and number of components, defined by the logic variables \( Y_{ig} \) and \( Y_n \), respectively, as described in Section 2.2. The conditional constraints included in the first set of disjunctions are converted into mixed-integer form as described in Formulation 1aR. The second set of disjunctions consists of all the linear and nonlinear constraints that depend on the number of components in the mixture, such as Eqs. (1)-(22), the mole fraction, \( x_i \), of each designed component in the mixture, the miscibility constraints (25), as well as equations of the UNIFAC model used to predict the activity coefficient of the designed solvent molecules, and they are converted into mixed-integer form using the big-M approach. The big-M parameter values for the nonlinear functions are chosen to be relaxed bounds, so that numerical difficulties arising from tight bounds are avoided. Finally, all the logic relations that appear in Formulation 1aG and their equivalent algebraic form are presented in Table A.1.

3.1.2. Results and discussion

All models reported in this paper were implemented and solved in GAMS (GAMS Development Corporation, 2017b) version 24.8.3, running on a single core of a dual 6 core Intel Xeon E5-1660 machine at 3.30 GHz. SBB (Bussieck and Drud, 2001), a local MINLP solver, based on the implementation of the branch-and-bound method and some standard NLP algorithms supported by GAMS, was used to solve all model instances presented. The results of
Table 1
Optimal solubility of ibuprofen, $x_{ibu}$, obtained when solving the restricted and generalised problems of case study 1a. The lower bound of solvent mole fractions in the mixture is set to $x_i^0 = 0.1$, $i \in \{2, \ldots, N + 1\}.$

| Problem | $x_{ibu}$ | $c_1$ | $c_2$ | $c_3$ | $x_i$ | CPU(s) | N* of nodes |
|---------|-----------|-------|-------|-------|-------|--------|-------------|
| D1 ($N=1$) | 0.5183 | DMSO  | 0.4817 |       | 0.11 | 3      |             |
| D2 ($N=2$) | 0.5077 | DMSO  | 0.3923 | DMF   | 0.1000 | 8.77 | 292       |
| D3 ($N=3$) | 0.4955 | DMSO  | 0.3045 | DMF   | 0.1000 | 5.5-sCF, acCH$_2$, CH$_3$ | 343.84 | 5272 |
| D4 ($N\geq3$) | 0.5183 | DMSO  | 0.4817 |       |       | 36.53 | 42        |

Fig. 1. Designed solvent molecular structures for maximising the solubility of ibuprofen.

Formulation 1aG ($N$ unknown) involves designing an optimal mixture with at most three solvents, and the results validate the optimal solutions obtained in the restricted models, by confirming that the highest solubility is achieved with DMSO and the mole fraction of ibuprofen is 0.5183.

The values presented for computational time in Table 1 correspond to the runs of Formulations 1aR and 1aG where the best solutions were found with SBB. In problems D1 ($N=1$), D2 ($N=2$) and D3 ($N=3$), the CPU time increases rapidly with the number of components, because of the increased size and complexity of the formulations. The general problem, D4 ($N \leq 3$), requires significantly less computational time than problem D3, where a ternary mixture is designed, and it appears to be more effective than enumerating all options in the restricted problem ($N = 1, 2, 3$). The final solutions achieved cannot be guaranteed to be the global solution means that the reported computational times are affected by the starting points. However, similar initial guesses are given to all models (e.g., the same starting points are used in both D3 and D4), so that an initial comparison of the problems can be carried out.

Attempts were made to solve the problems to global optimality using BARON version 17.1.2 (Tawarmalani and Sahinidis, 2005) and only the smallest problem (D1), where the number of designed solvents is fixed to one, is solved globally in 35.61 CPU seconds. The results verify the optimal solution obtained with the SBB MINLP solver, i.e., the maximum solubility is found to have a value of 0.5183 with DMSO as the optimal solvent, as shown in Table 1. Convergence to global optimality was not reached in 60,000 CPU seconds when larger, more challenging problems for the design of binary and ternary mixtures were solved. The solutions reported with SBB for problems D2, D3 and D4 were not identified by the global solver.

The importance of designing mixtures from a large set of atom groups and not restricting the design space is shown in this work. It is observed that better results can be achieved when solvents are designed from groups instead of selecting them from predefined lists. The solvent mixtures generated in this study yield higher solubility values than the optimal solvent mixtures determined in Jonuzaj et al. (2016), where the optimal components (i.e., chloroform, methanol and ethanol) were selected from a set of candidate of 9 compounds. An increase of 48% in the solubility is observed when ibuprofen is dissolved in the ternary mixture from groups compared to its solubility in the optimal ternary mixture of chloroform, methanol and ethanol found in previous work.

3.2. Case study 1b: Antisolvent mixture design for minimising the solubility of ibuprofen

The use of common crystallization techniques, such as cooling or evaporative crystallization, may not be suitable for purifying heat-sensitive chemicals that decompose or react with solvents at higher temperatures, or chemicals that exhibit a weak dependence of solubility on temperature over a practical range of temperatures (Berry et al., 1997). In these cases, drowning out crystallization, another common method for the purification of fine chemicals is used. It is based on the addition of a poor solvent, often referred to as a drowning-out agent or antisolvent, to a solute/solvent mix-
ture, which reduces the solubility of the solute in the mixture and results in crystal formation.

The choice of solvent/antisolvent mixtures in drowning out crystallization is an important decision which affects the success of the process. Water and some organic solvents (e.g., acetonitrile) are often selected as anti-solvents based on knowledge-based methods and database searches. As in most crystallization techniques, solubility is one of the key features of the process and it is considered as one of the main criteria for designing or selecting suitable drowning-out agents (Frank et al., 1999). Only a limited number of modelling approaches have addressed the problem of choosing or designing suitable solvents and antisolvents for drowning out crystallization. Frank et al. (1999) presented a procedure to screen and select candidate solvents that dissolve organic solids, based on the use of existing methods for estimating the solubility of the solids (e.g., UNIFAC, Robbins chart, Hansen solubility parameters, etc.). Karunanithi et al. (2006) and Austin et al. (2016) studied the design of optimal solvent-antisolvent mixtures that result in maximum potential recovery in the drowning out crystallization of ibuprofen. These existing design methodologies have only addressed the design of binary mixtures that consist of a solvent and an antisolvent (i.e., one optimal antisolvent is added to a solution). The design of antisolvent mixtures has not been considered so far.

3.2.2. Problem formulations

Following the strategy discussed in case study 1a, the application of the proposed methodology to the design of antisolvent mixtures is based on two problem formulations: (i) Formulation 1bR, where the number of solvent and antisolvent components is fixed (restricted problem); (ii) Formulation 1bG, where the number of components in the mixture is not fixed but bounded by a maximum Nmax (generalized problem). In Formulation 1bR, mixtures with one antisolvent (N=1 - problem E1) and two antisolvents (N=2 - problem E2) are designed, whereas Formulation 1bG involves the design of mixtures with at most two antisolvents (Nmax ≤ 2 - problem E3). The components in the mixture thus include ibuprofen (ibu) the fixed solvent component (DMSO) and the designed antisolvents (c1, c2). The antisolvent molecules are designed from a set of 49 atom groups, presented in Table B.1 in Appendix B.

Formulation 1bR consists of disjunctions for assigning a group g from the set G to a designed component i in the mixture that ensure that the relevant conditional constraints are active if the corresponding Boolean variable Yi,g is true. Formulation 1bG involves disjunctions for the identity and the number of components in the mixture, which are defined by the Boolean variables Yi,g and Yi, respectively. A detailed description of the GDP restricted and general problems is given in Section 3.1.1 where the problem formulations for maximizing the solubility of ibuprofen are presented. It should be noted that the disjunctive constraints that depend only on the identity of the fixed solvent component are expressed in terms of DMSO and thus, they are treated as general constraints, in the set X(x) ≤ 0.

The GDP formulations are converted into mixed integer form using the big-M approach and the MINLP models of the restricted and generalised problems can be accessed through https://zenodo.org/record/1154203.

3.2.3. Results and discussion

Optimal solutions presented in Table 2 were obtained with SBB (Bussieck and Drud, 2001). Problems E1 and E2 correspond to the models of the restricted problem (Formulation 1bR) and involve the design of one antisolvent (N=1) and a binary mixture with two antisolvents (N=3), respectively. Model E3 represents the general problem of Formulation 1bG which includes the design of a mixture with at most two antisolvents (N≤2).

The results of the restricted problems E1 and E2 show that the lowest solubility of ibuprofen, with a mole fraction value of 0.0025, is achieved with the addition of water (H2O) up to the maximum amount allowed to DMSO. Water is a poor solvent for many organic compounds (including ibuprofen) (Gracin et al., 2002; Gracin and Rasmussen, 2002), and it is commonly used as an antisolvent in drowning out crystallization (Frank et al., 1999). A ternary mixture of DMSO, water and acetic acid (CH3COOH) yields a greater solubility value of 0.0072. The mole fraction of acetic acid is at its lower bound and thus, only a small amount of acetic acid is added to the mixture, which accounts for small differences in the corresponding solubilities of the binary and ternary mixtures. The best solution obtained in restricted problem E1, where the addition of water yields the lowest solubility of ibuprofen, is also found in the generalized problem E3 where the number of antisolvents in the mixture is unknown.

The values of the CPU time presented in Table 2 correspond to the runs where the best solutions were found. It can be observed that the computational performance of the general model E3 is more advantageous than solving problems E1 and E2 sequentially. We note, however, that the problems were solved locally and thus the computational cost of each problem is affected by the initial guesses given to the MINLP solver. An overall comparison of all models could be conducted if the problems were solved globally.
but convergence to global optimality could not be reached with the MINLP algorithms within GAMS.

### 3.3. List of optimal solutions

The design of molecules from a large list of atom groups may sometimes lead to the design of components with undesired properties (e.g., high toxicity values) or new molecular structures that are difficult to synthesize in a laboratory. In order to address these issues, the integer cuts presented in Eq. (23) are introduced in the generalised problem formulation, so that lists of optimal solutions with different molecules and mixtures are generated. The ten best solutions obtained with the general models of both crystallization problem cases (D4 and E3) are ranked in descending order and presented in Tables 3 and 4, respectively.

In case study 1a, the list of optimal solutions includes mainly binary solvent mixtures (apart from the first solution), with DMSO being the main component for dissolving ibuprofen. The small differences in the solubility values among the results obtained are due to the small amount of the second designed solvents added in each mixture (mole fraction values of each solvent are at the lower bound). It can be observed that the second best solution is achieved with the same binary mixture (i.e., a mixture of DMSO and DMF) obtained in model D2 of the restricted problem. The mixture of DMSO, DMF and ethyl pentafluorobenzene found in model D3 yields a solubility value of 0.4955 which is lower than the solubilities obtained in the first ten solutions, and thus no ternary mixtures are included in the generated set. Ethyl pentafluorobenzene, though, is found to provide the third best solution in a binary mixture with DMSO.

### Table 2

| Problem | x_{DMSO} | x_{H_2O} | c_2 | x_{c_1} | c_3 | x_{c_2} | CPU(s) | N^c of nodes |
|---------|----------|----------|-----|---------|-----|---------|--------|-------------|
| E1 (N = 1) | 0.0025 | 0.2410 | 0.7565 | 0.317 | 67 |
| E2 (N = 2) | 0.0072 | 0.2410 | 0.6518 | 0.1000 | 389 |
| E3 (N ≤ 2) | 0.0025 | 0.2410 | 0.7565 | 12.87 | 84 |

### Table 3

| Ranking | Solvent mixtures | x_{DMSO} | c_1 | x_{c_1} |
|---------|-----------------|----------|-----|---------|
| 1       | Pure solvent    | 0.5183   | c_1: DMSO | 0.4817 |
| 2       | Binary mixture  | 0.5077   | c_1: DMSO | 0.3923 |
| 3       | Binary mixture  | 0.5075   | c_1: DMSO | 0.3925 |
| 4       | Binary mixture  | 0.5067   | c_1: DMSO | 0.3933 |
| 5       | Binary mixture  | 0.5059   | c_1: DMSO | 0.3941 |
| 6       | Binary mixture  | 0.5058   | c_1: DMSO | 0.3942 |
| 7       | Binary mixture  | 0.5050   | c_1: DMSO | 0.3950 |
| 8       | Binary mixture  | 0.5048   | c_1: DMSO | 0.3952 |
| 9       | Binary mixture  | 0.5046   | c_1: DMSO | 0.3954 |
| 10      | Binary mixture  | 0.5038   | c_1: DMSO | 0.3962 |

### Table 4

| Ranking | Antisolvent mixtures | x_{DMSO} | i | c_1 | x_{c_1} |
|---------|---------------------|----------|---|-----|---------|
| 1       | Pure antisolvent    | 0.0025   | c_1: H_2O | 0.7565 |
| 2       | Binary mixture      | 0.0072   | c_1: H_2O | 0.6518 |
| 3       | Binary mixture      | 0.0077   | c_1: CH_3, COOH | 0.1000 |
| 4       | Binary mixture      | 0.0104   | c_1: OH, CH_2NO_2 | 0.1000 |
| 5       | Binary mixture      | 0.2131   | c_1: CH_3, COOH | 0.5459 |
| 6       | Pure antisolvent    | 0.2310   | c_1: OH, CH_2NO_2 | 0.5280 |
| 7       | Pure antisolvent    | 0.2436   | c_1: CH_3, CH_2, COOH | 0.5154 |
| 8       | Pure antisolvent    | 0.2628   | c_1: 2 × CH_3, CH, COOH | 0.4962 |
| 9       | Pure antisolvent    | 0.2877   | c_1: OH, CH_2, CH_2NO_2 | 0.4713 |
| 10      | Pure antisolvent    | 0.3002   | c_1: OH, CH_2Cl | 0.4588 |
In case study 1b, the top ten solutions consist of optimal mixtures with one or two antisolvent components. The first two optimal solutions with the lowest solubility values validate the results obtained in models E1 and E2 of the restricted problem (see Table 2). It is seen that extremely low solubility values can be achieved when water is added to the DMSO + ibuprofen mixture (see first four solutions of Table 4), in contrast with the higher solubility values found for ibuprofen in different antisolvent components (see last six solutions of Table 4). DMF is found to be the second best pure solvent which yields a solubility value of 0.4628. There is a large difference in solubility between the first and second pure solvents and binary mixtures provide alternative high performance options.

4. Case study 2: Separation of acetic acid from water by liquid–liquid extraction

The problem of identifying optimal solvent mixtures from a given set of candidate compounds to separate acetic acid from water in a single stage liquid extraction process provides a more challenging case study due to the participation of all mixture components in the liquid–liquid equilibrium equation.

Liquid–liquid extraction has been proven to be an economical and environmentally friendly method for the purification of acetic acid from an aqueous stream (Alkaya et al., 2009; Austin et al., 2017; Ijzner et al., 2014; Karunanithi et al., 2005). As shown in the single stage extraction unit illustrated in Jonuzaj and Adjiman (2017), a feed mixture of acetic acid and water and a solvent mixture are first mixed and are then separated into extract and raffinate phases in a settler. The design specifications given in Seader et al. (2011) are used in this case study: an aqueous solution with 2.5 mol % acetic acid and a feed flowrate of 708 kmol/h at 298 K and 1 atm are considered. The raffinate phase should contain at most 0.3 mol % of acetic acid.

The objective here is to design an optimal solvent mixture that minimises the solvent-to-feed ratio and also achieves a large fraction of acetic acid and small fraction of water in the extract phase, as shown in the single objective function presented below:

$$\min \frac{F_i}{F} + x_{E,w} - x_{E,a}$$ (26)

where $F_i$ and $F$ are the solvent and feed flowrates, respectively, while $x_{E,a}$ and $x_{E,w}$ are the mole fractions of acetic acid and water, respectively, in the extract phase. The problem involves liquid–liquid equilibrium (LLE) relations between extract and raffinate phases, as described in Eq. (27):

$$\gamma_{E,i} x_{E,i} = \gamma_{R,i} x_{R,i}, \quad i \in I$$ (27)

where $x_{E,i}$ and $x_{R,i}$ are the mole fractions of component $i$ in the extract and raffinate streams, respectively, and $\gamma_{E,i}$ is the liquid phase activity coefficient of component $i$ in stream $p$ at temperature $T$, composition $x_p$ and pressure $P$. The activity coefficient values are calculated using the modified UNIFAC (Dortmund) model (Gmehling et al., 1993; 2002) and the relevant equations are presented in Jonuzaj and Adjiman (2017). The mole balances for each component in the system are described in Eqs. (28)–(31), respectively:

$$(x_{E,a} M = x_{M,a} M)$$ (28)

$$(x_{E,w} M = x_{M,w} M)$$ (29)

$$(x_{E,i} S = x_{M,i} M, \quad i \in \{c_1, c_2, c_3\})$$ (30)

$$(x_{M,i} M = x_{E,i} E + x_{R,i} R, \quad i \in I)$$ (31)

where $M$ is the flowrate of the output stream of the mixer, and $E$ and $R$ are the extract and raffinate flowrates, respectively; $x_{E,i}$ represents the mole fraction of component $i$ in stream $p$.

The molecular structures of acetic acid and water are represented through the groups CH$_3$, COOH and H$_2$O, respectively, whereas the solvent molecules are designed from a set of 25 atom groups ($G = \{1, \ldots, 25\}$) presented in Table B.2 in Appendix B. Each solvent structure $i$ consists of at most 10 groups ($n_{i}^{\text{max}} = 10$). We note that a reduced list of groups is used in this case study, compared to the set of 49 groups employed in the crystallization problem, due to the smaller number of available interaction parameters in the publicly available Dortmund UNIFAC database.

4.1. Problem formulations

Once again, a GDP formulation, where the number of components is fixed (restricted problem, 2R), and a formulation with unknown number of solvents (generalized problem, 2G), are presented in this section. In Formulation 2R, mixtures with one (problem F1), two (problem F2) or three solvents (problem F3) are designed and in Formulation 2G, mixtures with at most three solvents (problem F4) are identified. The components present in the final mixture include acetic acid ($a$) and water ($w$), which are the fixed components in the mixture, and the designed solvent molecules ($c_1, c_2, c_3$).

The general constraints in the GDP models consist of the LLE relations and mole balances that hold for acetic acid and water, as well as the UNIFAC model equations to predict the activity coefficients of the fixed components, as they do not depend on the logic decisions (i.e., identity and number of the designed molecules). All the equations that run over the designed components in the mixture ($c_1, c_2, c_3$), such as chemical feasibility and chemical complexity constraints, design constrains, LLE equations, mole balances and UNIFAC equations for the designed solvents, depend on the identity of the groups in each molecule and on the number of solvents in the mixture. Thus, these conditional constraints are included in appropriate disjunctions for assigning each group $g$ to component $i$ and for the number of designed components in the final mixture, and they are active when the corresponding logic variables $Y_{g,i}$ and $\hat{Y}_i$ are true. The disjunctive constraints are converted into mixed integer form using the big-M approach.

As discussed in case study 1a, the number, $n_{i}^{\text{cg}}$, of each group $g$ in a designed molecule given in Eq. (3) and several constraints derived for the design of solvent molecules (described in Section 2.2) are formulated via the binary variables $\hat{y}$. These relations force the number of groups $n_{i}^{\text{cg}}$ and all related variables to become zero when a group $g$ is not present in the designed component. Therefore, the use of big-M parameters which may lead to poor relaxations is avoided when transforming these equations in the MINLP models.

The logic propositions and the corresponding linear algebraic inequalities presented in Table A.1 are included in the resulting GDP and MINLP formulations. The MINLP models and all the parameter values (e.g., big-M parameters) included in the problems can be found at https://www.zenodo.org/record/1154203.

4.2. Results and discussion

The results of the restricted (Formulation 2R) and generalized (Formulation 2G) problem statements are summarized in Tables 5 and 6. In Formulation 2R (F1–F3), the minimum objective function (with a value of 1.3129) and the highest recovery of acetic acid in the extract phase (with a value of 93.21%) are obtained with a pure solvent, pentafluoroanisole (5-xAC, aCCH$_2$, CH$_3$COO). The best binary and ternary mixtures generated include pentafluoroanisole (5-xAC, aCCH$_2$, CH$_2$COO, CH$_3$) and tetrafluoro-
Table 5
Optimal objective function, optimal solvent mixtures, feed mole fractions, CPU times and number of nodes obtained when solving the restricted and generalised problems of case study 2. The lower bound of solvent feed mole fractions in the solvent mixture is set to $x_{i}^{\text{L}} = 0.1, \ i \in [1 \ldots N + 1].$

| Problem | Objective | Components | $x_{i}(x_{i})$ | CPU(s) | $N^*$ of nodes |
|---------|-----------|------------|----------------|---------|----------------|
| F1 ($N = 1$) | 1.3129 | Acetic acid | CH$_2$, COOH | H$_2$O | 0.0250 | 3.84 | 125 |
|          |          | Water      |                |          | 0.9750 |       |      |
|          |          | $c_1$      | $5 \times aCF, aCCH_2, CH_3COO$ | 1.0000 |       |      |      |
| F2 ($N = 2$) | 1.3141 | Acetic acid | CH$_2$, COOH | H$_2$O | 0.0250 | 7.69 | 121 |
|          |          | Water      |                |          | 0.9750 |       |      |
|          |          | $c_1$      | $5 \times aCF, aCCH_2, CH_3COO$ | 0.9000 |       |      |      |
|          |          | $c_2$      | $5 \times aCF, aCCH_2, CH_3COO, CH_3$ | 0.1000 |       |      |      |
| F3 ($N = 3$) | 1.3219 | Acetic acid | CH$_2$, COOH | H$_2$O | 0.0250 | 158.98 | 2063 |
|          |          | Water      |                |          | 0.9750 |       |      |
|          |          | $c_1$      | $5 \times aCF, aCCH_2, CH_3COO$ | 0.8000 |       |      |      |
|          |          | $c_2$      | $5 \times aCF, aCCH_2, CH_3COO, CH_3$ | 0.1000 |       |      |      |
|          |          | $c_3$      | $4 \times aCF, aCCH_2, aCCH_3COO$ | 0.0000 |       |      |      |
| F4 ($N \leq 3$) | 1.3129 | Acetic acid | CH$_2$, COOH | H$_2$O | 0.0250 | 58.06 | 375 |
|          |          | Water      |                |          | 0.9750 |       |      |
|          |          | $c_1$      | $5 \times aCF, aCCH_2, CH_3COO$ | 1.0000 |       |      |      |

Table 6
Solvent feed ($F_i$), extract ($E$) and raffinate ($R$) flowrates in kmol/h, extract and raffinate mole fractions and acetic acid recovery obtained when solving the restricted and generalised problems of case study 2.

| Problem | $F_i$ | $E$ | $R$ | $x_{i_E}$ | $x_{i_R}$ | $x_{i_E}$ | $x_{i_E}$ | $x_{i_S}$ | $x_{i_E}$ | $%$ Recovery of acetic acid |
|---------|-------|-----|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------------------------|
| F1      | 744.45 | 1051.91 | 400.54 | 0.0157 | 0.2717 | 0.7072 | 93.21 |
| F2      | 748.99 | 1050.66 | 406.32 | 0.0157 | 0.2719 | 0.6411 | 0.0713 | 93.11 |
| F3      | 765.11 | 1051.00 | 422.12 | 0.0157 | 0.2569 | 0.5819 | 0.0030 | 0.0728 | 92.85 |
| F4      | 744.45 | 1051.91 | 400.54 | 0.0157 | 0.2717 | 0.7072 | 93.21 |

Fig. 2. Designed solvent molecular structures for separating acetic acid from water in a liquid extraction process.

4-methylbenzyl acetate ($4 \times aCF, aCCH_3, aCCH_2, CH_3COO$) as the second and third designed solvent molecules, respectively, and they result in worse objective function values than the pure component, due to the increased flowrates ($F_i$) of the solvent mixtures (see Table 6).

As shown in Fig. 2, similar molecular structures are designed in all cases as optimal solvent components in the mixture, with aCF being the most frequent group in all molecules. Therefore, the percentage of second and third solvent molecules do not affect significantly the separation process and the results obtained with the optimal pure component and the optimal solvent mixtures are very similar. We note that pentafluorobenzoyl acetate and pentafluorobenzyl propionate are not commonly used as solvents, and currently very expensive (Synquest Laboratories, 2010;
Table 7
List of optimal solutions obtained for the general mixture problem (F4) of case study 2 when including integer cuts. \( i \) represents the solvent molecules and \( x_i \) is the mole fraction of each solvent in the mixture (\( i = c_1, c_2, c_3 \)).

| Ranking | Mixtures | Objective | \( i \) |
|---------|----------|-----------|--------|
| 1       | Pure solvent | 1.3129 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 1.0000 |
| [2]     | Binary mixture | 1.3141 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
| [3]     | Ternary mixture | 1.3219 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.8000 |
|         |           |          | \( c_2: 5 \times aCF, aCH\_2, CH\_2COO \) |
|         |           |          | \( c_2: 5 \times aCF, aCH\_2, CH\_2COO \) |
| 4       | Binary mixture | 1.3234 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) |
| 5       | Binary mixture | 1.3254 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_2: 4 \times aCF, ac, aCH, CH\_2COO \) 0.0100 |
| 6       | Binary mixture | 1.3283 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_2: 5 \times aCF, aCH\_1, CH\_2COO \) 0.0100 |
| 7       | Binary mixture | 1.3284 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_2: 5 \times aCF, aCH\_1, CH\_2COO \) 0.0100 |
| 8       | Binary mixture | 1.3325 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_1: 3 \times aCF, 2 \times ac, ac, CH\_2COO \) 0.0100 |
| 9       | Binary mixture | 1.3338 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.9900 |
|         |           |          | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.0100 |
| 10      | Ternary mixture | 1.3358 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.8000 |
|         |           |          | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.0100 |
|         |           |          | \( c_1: 4 \times aCF, ac, aCH, CH\_2COO \) 0.0100 |
| 11      | Ternary mixture | 1.3686 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 0.8000 |
|         |           |          | \( c_1: 4 \times aCF, ac, aCH, CH\_2COO \) 0.0100 |
| 12      | Pure solvent | 1.4481 | \( c_1: 5 \times aCF, aCH\_1, CH\_2COO \) 1.0000 |
| 13      | Ternary mixture | 2.0927 | \( c_1: 2 \times CH\_1, CH\_2COO \) 0.7597 |
|         |           |          | \( c_1: 2 \times CH\_1, CH\_2COO \) 0.1403 |
|         |           |          | \( c_1: 2 \times CH\_1, CH\_2COO \) 0.1000 |

VWR Avantor, 2010), in part due to their limited uses/production. Tetrafluoro-4-methylbenzyl acetate is not found in the available solvent databases (DETERM, 2008; NIST, 1996) so that no experimental property data could be found. Hence, the generated molecules should be tested with respect to additional criteria, such as environmental, safety, or cost constraints, not considered in the design specifications.

The general problem (F4), which was solved for at most three designed solvents, yields the same optimal solution obtained in restricted problem F1. ANTIGONE and BARON were employed to solve the problems globally, but convergence to global optimality was not reached in 100,000 CPU seconds.

The computational times presented in Table 5 for Formulation 2R (F1-F3) increase with the number of components in the mixture. The solution of Formulation 2G (F4) requires less CPU time than solving Formulation 2R with three solvents (F3). Similar initial guesses were given in each problem, so that the solution times are not affected significantly by the choice of starting points.

It can be observed that the identity of the solvent molecules has a significant impact on the performance of the process. The molecules and mixtures designed in this work yielded better solutions than the solvents selected from a given list of 8 compounds presented in previous work (Jonuzaj and Adjian, 2017). In particular, the solvent mixture flowrate (\( F_2 \)) is decreased by 48% when pentafluorobenzyl acetate is used to separate acetic acid from water, which may lead to reduced capital and operating costs.

4.3. List of optimal solutions

Integer cuts, presented in Eq. (23), are included in the generalised problem (F4) of case study 2 and a list of twelve optimal solutions is presented in Tables 7 and 8. The list contains pure solvents, as well as binary and ternary solvent mixtures. Only ten of the solutions were generated by adding the integer cuts to F4; the second and third solutions (in brackets) of the list were obtained by solving problems F2 and F3 (see Table 5), but were not found in the first ten runs of the generalised problem when introducing integer cuts. This indicates that on at least two occasions, the SBB solver converged to a local solution.

In most solutions in the list, solvent mixtures containing aromatic compounds that consist of one fluorobenzene ring and one or two side chains, are designed. If fluorobenzene rings are excluded from the design of monocyclic compounds, then a ternary mixture with acyclic solvents, butanone (2 × CH\_3, CH\_2CO), methyl butanone (2 × CH\_3, CH\_2CO), and pentanone (2 × CH\_3, CH\_2CO, CH\_3CO) in position 13, yields an optimal solution with higher objective function and flowrate values, but better acetic acid recovery than any of the top 12 solvents. Hence, a trade-off between reduced costs and improved separation and recovery of acetic acid can be achieved with different solvent designs.

5. Conclusions

This study provides a general and comprehensive mathematical formulation for the design of optimal molecules and mixtures based on a computer aided mixture/design (CAM\(^{\text{RD}}\)) framework. Within this systematic approach the number of mixture components, the identities of the components and their compositions are determined simultaneously. The general formulation makes it possible to consider large and diverse design spaces by designing the optimal molecular structures of all mixture components simultaneously from a large set of atom groups (building blocks), and it focuses on identifying the best pure or blended components, without specifying a priori whether one seeks a single compound or a mixture. Disjunctions are employed to model the effect of discrete decisions (i.e., the number of mixture ingredients, the identities and numbers of groups in a molecular structure and of molecules in the final mixture), on the problem equations, leading to a GDP formulation.

The proposed approach has been applied successfully to different case studies in the context of separation processes, where optimal solvent, antisolvents, solvent and antisolvent mixtures are
Table 8
Solvent feed (F), extract (E) and raffinate (R) flowrates in kmol/h, extract and raffinate mole fractions and acetic acid recovery obtained from every run when including integer cuts in the general mixture problem (F4) of case study 2.

| Ranking | FS | ER | X_k,a | X_k,b | X_k,w | X_k,f1 | X_k,f2 | X_k,R1 | X_k,R2 | % Recovery of acetic acid |
|---------|----|----|-------|-------|-------|--------|--------|--------|--------|---------------------------|
| 1       | 744.45 | 1051.91 | 0.0157 | 0.2771 | 0.7072 |        |        |        |        | 93.21                     |
| 2       | 748.99 | 1050.66 | 0.0157 | 0.2719 | 0.6411 | 0.0713 |        |        |        | 93.11                     |
| 3       | 765.11 | 1051.00 | 0.0156 | 0.2569 | 0.5819 | 0.0728 | 0.0728 |        |        | 92.85                     |
| 4       | 758.34 | 1057.48 | 0.0156 | 0.2678 | 0.6449 | 0.0717 |        |        |        | 93.07                     |
| 5       | 754.32 | 1063.05 | 0.0155 | 0.2755 | 0.6381 | 0.0709 |        |        |        | 93.23                     |
| 6       | 752.81 | 1068.48 | 0.0155 | 0.2805 | 0.6336 | 0.0704 |        |        |        | 93.35                     |
| 7       | 760.78 | 1062.86 | 0.0155 | 0.2693 | 0.6437 | 0.0714 |        |        |        | 93.12                     |
| 8       | 765.66 | 1065.53 | 0.0155 | 0.2665 | 0.6462 | 0.0718 |        |        |        | 93.08                     |
| 9       | 761.55 | 1070.26 | 0.0154 | 0.2736 | 0.6399 | 0.0711 |        |        |        | 93.23                     |
| 10      | 768.24 | 1068.37 | 0.0154 | 0.2661 | 0.5748 | 0.0719 | 0.0718 |        |        | 93.09                     |
| 11      | 792.34 | 1098.81 | 0.0150 | 0.2644 | 0.5765 | 0.0720 | 0.0721 |        |        | 93.19                     |
| 12      | 839.70 | 1181.89 | 0.0140 | 0.2762 | 0.7098 |        |        |        |        | 93.80                     |
| 13      | 1220.80 | 1737.72 | 0.0097 | 0.2368 | 0.5718 | 0.1060 | 0.0756 |        |        | 95.07                     |

Further verification of the designed molecular structures should, however, be conducted, especially when new molecules are designed. The GDP formalism provides a useful framework for formulating and solving mixture problems, making it possible to consider large design spaces and to avoid numerical difficulties arising from the absence of groups in a molecular structure and of components in the final mixture (i.e., with a mole fraction of zero).

In the context of separation solvent and process design, future work will focus on extending the problem formulations to provide a more realistic assessment of performance (e.g., include environmental and safety metrics Karunanithi et al., 2005; Karunanithi et al., 2006). Future investigations will also be directed at tackling product design applications, especially formulated products (Mattei et al., 2012; Yunus et al., 2014) which exhibit an increased number of mixture ingredients and where the general mixture design formulation may be especially beneficial: the explicit evaluation of every choice of number of components can be avoided in this way, making it possible to consider larger design spaces. In addition, the development of suitable algorithms to achieve global solutions should be considered, in order to carry out a comprehensive comparison of the proposed formulations.

Acknowledgements

The authors gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EPSRC) of the UK via a Leadership Fellowship (EP/J003840/1), a platform grant (EP/J014958/1) and a Doctoral Prize to S. Jonuzaj.

Data statement: Data underlying this article can be accessed on Zenodo at https://www.zenodo.org/record/1154203, and used under the Creative Commons Attribution license.
Appendix A. Logic relations of the GDP mixture problem

| Logic expressions | Linear inequalities |
|-------------------|---------------------|
| Eliminate degenerate solutions | $\sum_{\beta \in C} |y_{\beta} - y_{\beta'}| \geq 1$ |
| with $D_{\alpha} = (Y_{\alpha} - Y_{\alpha'})$ | $i = N' + 1, \ldots, N_i - 1$, $\beta' = i + 1, \ldots, N_i$ |

Relations between $Y_{\alpha}$ and $Y_{\alpha'}$

- $\neg Y_{\alpha} \lor (\sum_{\beta \in C} Y_{\beta} + 1)$
- $\neg Y_{\alpha} \lor (\sum_{\beta \in C} Y_{\beta} + 1)$, $i = N' + 2, \ldots, N_i$, $g \in G$
- $\neg Y_{\alpha} \lor (\sum_{\beta \in C} Y_{\beta} + 1)$, $i = N' + 3, \ldots, N_i$, $g \in G$

Relations between $Y_{\alpha}$ and $\hat{y}$

Let $\hat{Y}_{\alpha} = (\hat{y}_{\alpha} = 1)$

and $\hat{Y}_{\alpha'} = (\hat{y}_{\alpha'} = 1)$

- $\neg \hat{Y}_{\alpha} \lor (\sum_{\beta \in C} \hat{y}_{\beta} + 1)$, $i = N' + 2, \ldots, N_i$, $g \in G$; $t \in T_M$
- $\neg \hat{Y}_{\alpha} \lor (\sum_{\beta \in C} \hat{y}_{\beta} + 1)$, $i = N' + 3, \ldots, N_i$, $g \in G$; $t \in T_M$
- $\neg \hat{Y}_{\alpha} \lor (\sum_{\beta \in C} \hat{y}_{\beta} + 1)$, $i = N' + 4, \ldots, N_i$, $g \in G$; $t \in T_M$

Appendix B. Data used in the design of molecules from atom (UNIFAC) groups

The data included in this section for the design of molecules from atom groups are taken from the work of Struebing et al. (2017).

| $G_1$ | $G_{CH}$ | $G_{CHO}$ | $G_{CO}$ | $G_{CH}_{D}$ | $G_{CH}_{N}$ | $G_{CH}_{S}$ |
|-------|----------|-----------|----------|-------------|-------------|-------------|
| OH    | CH=CH    | CH=CH     | aCH      | CH$_3$      | CH$_3$      | aC          |
| CH$_2$ | CH$_3$   | CH$_3$    | ac       | CH$_2$      | CH$_2$      | aCH         |
| CH$_3$ | CH$_2$CO | CHO       | aC       | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$CO | CH$_2$CO  | aC       | CH$_2$      | CH$_2$      | aC          |
| CH$_3$ | CH$_2$CO | CH$_2$OH  | CH$_2$   | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$SH  | COOH     | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$Cl  | ac       | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$Cl  | acF      | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$NO$_2$ | acF    | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$NO$_2$ | acF    | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$SH  | ac       | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$   | Br       | CH$_2$      | CH$_2$      | aC          |
| CH$_2$ | CH$_2$  | CH$_2$S   | Br       | CH$_2$      | CH$_2$      | aC          |
### Table B.2
Subsets of set G for the solvent mixture design of case study 2. The subsets $G_a$, $G_{aCCH}$, $G_{CCH}$, $G_{CH}$, $G_1$, and $G_B$ represent functional groups, non-chain-ending groups, chain-ending groups, aromatic groups, main groups, single-molecule group and branched aromatic groups, respectively.

| Group | $G_a$ | $G_{aCCH}$ | $G_{CCH}$ | $G_{CH}$ | $G_1$ | $G_B$ |
|-------|-------|------------|-----------|---------|------|------|
| OH    | CH=CH | CH$_2$=CH | aCH       | CH$_3$  | CHCl$_2$ | aC    |
| CH$_3$CO | CH$_2$CO | CH$_2$CO | aC        | CH$_2$  | H$_2$O | aCH   |
| CH$_2$CO | CH$_3$COO | CHO | aCCH$_2$ | CH     | C      | aCH$_2$ |
| CHO   | CH$_2$COO | COOH | aCCH     | CH$_2$=CH | CH=CH | C=C   |
| CH$_3$COO | COOH | aC      | CH$_2$=CH | CH=CH | C=C   |
| COOH  | aC     | CH$_2$=C | CH=C     | C=C   |
| CHCl$_3$ | C=C | C=C       | C=C   |

### Table B.3
Valency $v_g$, normal melting point $T_{mel,g}$ and normal boiling point $T_{boil,g}$ contributions of each group.

| Group $g \in G$ | $v_g$ | $T_{mel,g}$ (K) | $T_{boil,g}$ (K) |
|-----------------|-------|-----------------|-----------------|
| CH$_3$         | 1     | 0.6953          | 0.8491          |
| CH$_2$         | 2     | 0.2515          | 0.7141          |
| CH             | 3     | $-0.3730$       | 0.2925          |
| C              | 4     | 0.0256          | $-0.0671$       |
| CH$_2$=CH      | 1     | 1.1728          | 1.5596          |
| CH=CH          | 2     | 0.9460          | 1.5597          |
| CH$_2$=C       | 2     | 0.7662          | 1.3621          |
| CH=C           | 3     | 0.1732          | 1.2971          |
| C=C            | 4     | 0.3928          | 1.2739          |
| aCH            | 2     | 0.5860          | 0.8365          |
| aC             | 3     | 1.8955          | 1.7324          |
| aCCH$_2$       | 2     | 1.0068          | 1.5653          |
| aCH$_3$        | 3     | 0.1065          | 1.4925          |
| aCH$_2$        | 4     | $-0.5197$       | 0.8665          |
| OH             | 1     | 2.7888          | 2.5670          |
| aCOH           | 2     | 5.1473          | 3.2205          |
| CH$_2$CO       | 1     | 2.9588          | 3.1178          |
| CH$_3$CO       | 2     | 2.5232          | 2.6761          |
| CHO            | 1     | 3.0186          | 2.5388          |
| CH$_3$COO      | 1     | 2.1657          | 3.1228          |
| CH$_2$COO      | 2     | 1.6329          | 2.9850          |
| CH$_3$O        | 1     | 1.3643          | 1.7703          |
| CH$_2$O        | 2     | 0.8733          | 1.3368          |
| CH-O           | 3     | 0.2461          | 0.8924          |
| CH$_2$NH$_2$   | 1     | 3.2742          | 2.7987          |
| CH$_3$NH       | 1     | 2.4034          | 2.2514          |
| CH$_2$NH$_2$   | 2     | 1.7466          | 1.8750          |
| CH$_3$N        | 2     | 0.9607          | 1.3841          |
| CH$_3$N        | 3     | 0.0442          | 1.1222          |
| aCNO$_3$       | 2     | 3.9889          | 3.8298          |
| CH$_3$CN       | 0     | 4.7365          | 4.2923          |
| CH$_3$CN       | 1     | 2.5760          | 4.5871          |
| COOH           | 1     | 7.4042          | 5.1008          |
| CH$_2$Cl       | 1     | 1.9253          | 2.6364          |
| CHCl$_2$       | 2     | 1.0224          | 2.0246          |
| CHCl$_3$       | 0     | 2.5196          | 3.3420          |
| aC$\equiv$Cl   | 2     | 4.1473          | 4.4919          |
| CH$_2$NO$_3$   | 0     | 1.7134          | 2.0669          |
| CH$_2$NO$_2$   | 1     | 5.2594          | 5.3768          |
| CH$_2$NO$_2$   | 2     | 3.2131          | 4.5311          |
| CHONO$_2$      | 2     | 0.7812          | 3.8089          |
| CH$_3$SH       | 1     | 2.2992          | 3.1974          |
| I               | 1     | 1.9444          | 3.1778          |
| Br             | 1     | 1.7641          | 2.4231          |
| aCF            | 2     | 0.9782          | 0.7945          |
| CH$_3$S        | 2     | 1.0063          | 2.6524          |
| H$_2$O         | 0     | 6.3757          | 5.3475          |
| DMF            | 0     | 7.1981          | 7.9779          |
| DMF            | 0     | 4.2305          | 6.7863          |
Table B.4
Upper bounds provided by linear functions \( \hat{f} \) that define if each group \( g \) is allowed to appear in a designed molecule \( i \). Groups with an * have been deactivated in the studied problems due to lack of sufficient data for accurate predictions.

| Group \( g \) | \( \hat{f}_g(\hat{y}_i), \ i \in T, \ g \in G, \ t \in T_M \) |
|--------------|--------------------------------------------------|
| \( \text{CH}_3 \) | \( m_{\text{CH}_3} \cdot \hat{y}_i + 2 \cdot \hat{y}_i \cdot \text{acCH} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_2 \) | 3 \( \hat{y}_i \) |
| \( \text{CH} \) | 3 \( \hat{y}_i \) |
| \( \text{C} \) | \( \hat{y}_i \) |
| \( \text{CH}_2=\text{CH}^{*} \) | 0 (\( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \)) |
| \( \text{CH}–\text{CH}^{*} \) | 0 \( \hat{y}_i \) |
| \( \text{CH}–\text{C}^{*} \) | 0 \( \hat{y}_i \) |
| \( \text{C}–\text{C}^{*} \) | \( \hat{y}_i \) |
| \( \text{aCH} \) | 6 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + 8 \hat{y}_i \) |
| \( \text{aC} \) | \( \hat{y}_i \cdot \text{ac} + 2 \hat{y}_i \) |
| \( \text{aCCH}_3 \) | 6 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + 8 \hat{y}_i \) |
| \( \text{aCCH}_2 \) | \( \hat{y}_i \cdot \text{acCH} \) |
| \( \text{aCCH} \) | \( \hat{y}_i \cdot \text{ac} \) |
| \( \text{aCOH} \) | 6 \( \hat{y}_i + 8 \hat{y}_i \) |
| \( \text{CH}_3\text{CO} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_3\text{CO} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CHO} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_3\text{COO} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_2\text{COO} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_2\text{O} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_2 \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_3\text{NH}_2^{*} \) | 0 (2 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \)) |
| \( \text{CH}_3\text{NH}^{*} \) | 0 \( \hat{y}_i \) |
| \( \text{CH}_3\text{NH}^{*} \) | 0 \( \hat{y}_i \) |
| \( \text{aCNH}_2^{*} \) | 0 (6 \( \hat{y}_i + 8 \hat{y}_i \)) |
| \( \text{CH}_3\text{CN} \) | \( \hat{y}_i \) |
| \( \text{CH}_2\text{CN} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CDOH} \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_3\text{Cl} \) | 2 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CHCl} \) | \( \hat{y}_i \) |
| \( \text{CHCl}_3 \) | \( \hat{y}_i \) |
| \( \text{H}_2\text{O} \) | \( \hat{y}_i \) |
| \( \text{DMSO} \) | \( \hat{y}_i \) |
| \( \text{DMF} \) | \( \hat{y}_i \) |
| \( \text{aCCl} \) | 6 \( \hat{y}_i + 8 \hat{y}_i \) |
| \( \text{CH}_3\text{NO}_2 \) | \( \hat{y}_i \) |
| \( \text{CH}_3\text{NO}_2 \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CHNO}_2 \) | \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{CH}_3\text{SH}^{*} \) | 0 (\( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \)) |
| \( \text{I} \) | 2 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \) |
| \( \text{Br}^{*} \) | 0 (2 \( \hat{y}_i + \hat{y}_i \cdot \text{ac} + \hat{y}_i \cdot \text{acCH} \)) |
| \( \text{aCF} \) | 6 \( \hat{y}_i + 8 \hat{y}_i \) |
| \( \text{CH}_3\text{F}^{*} \) | 0 \( \hat{y}_i \) |
Table B.5
Upper bounds on the number of groups that limit the simultaneous occurrence of different functional groups. Groups with an * have been deactivated in the studied problems.

| group $g \in G_F$ | $n_{sl}^i \ g \in G_F \ i \in I$ |
|-------------------|---------------------------------|
| OH                | 3.0                             |
| CH$_3$CO          | 1.0                             |
| CH$_2$CO          | 1.0                             |
| CHO               | 1.0                             |
| CH$_3$COO         | 1.0                             |
| CH$_2$COO         | 1.0                             |
| CH$_2$O           | 1.0                             |
| CH$_3$O           | 1.0                             |
| CH-O              | 1.0                             |
| CH$_3$NH$_2^*$    | 2.0                             |
| CH$_3$NH$_3^+$    | 1.0                             |
| CH$_3$N$_3^+$     | 1.0                             |
| CH$_3$N$_3$       | 1.0                             |
| CH$_3$CN          | 1.0                             |
| CH$_3$CN          | 1.0                             |
| COOH              | 1.0                             |
| CH$_2$Cl          | 2.0                             |
| CHCl              | 1.0                             |
| CHCl$_2$          | 2.0                             |
| CHCl$_3$          | 1.0                             |
| CH$_2$NO$_2$      | 1.0                             |
| CH$_3$NO$_2$      | 2.0                             |
| CHNO$_2$          | 1.0                             |
| CH$_2$SH$_2^*$    | 1.0                             |
| I                 | 2.0                             |
| Br$_-$            | 2.0                             |
| CH$_3$S$_2^+$     | 1.0                             |
References

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.

Alkaya, E., Kaptan, S., Özkam, L., Uldudag-Demirer, S., Demirer, G.N., 2009. Recovery of acids from anaerobic acidification broth by liquid-liquid extraction. Chemo-

Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer Aided Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands.

Adjiman, C.S., Clarke, A., Donn, C., 2008. Solvents for ring-closing metathesis reactions. Chem. Commun. 2008–2008.
Smith, J.M., Van Ness, H.C., Abbott, M.M., 2001. Introduction to Chemical Engineering Thermodynamics, 6th ed. McGraw-Hill, Boston.
Solvason, C.C., Chemmangattuvalappil, N.G., Eljack, F.T., Eden, M.R., 2009. Efficient visual mixture design of experiments using property clustering techniques. Ind. Eng. Chem. Res. 48, 2245–2256.
Struebing, H., Ganase, Z., Karamertzakis, P.G., Siougkrou, E., Haycock, P., Piccione, P.M., Armstrong, A., Galindo, A., Adjiman, C.S., 2013. Computer-aided molecular design of solvents for accelerated reaction kinetics. Nat. Chem. 5, 952–957.
Struebing, H., Obermeier, S., Siougkrou, E., Adjiman, C.S., 2017. A qm-camd approach to solvent design for optimal reaction rates. Chem. Eng. Sci. 159, 69–83.
Synquest Laboratories, 2010. 2,3,4,5,6-pentafluorobenzyl acetate. http://www.synquestlabs.com/product/id/130425.html. Accessed 2017.
Tawarmalani, M., Sahinidis, N.V., 2005. A polyhedral branch-and-cut approach to global optimization. Math. Program. 103, 223–249.
Vaidyanathan, R., El-Halwagi, M., 1996. Computer-aided synthesis of polymers and blends with target properties. Ind. Eng. Chem. Res. 35, 627–634.
Vaidyaraman, S., Maranas, C.D., 2002. Synthesis of mixed refrigerant cascade cycles. Chem. Eng. Commun. 189, 1057–1078.
VWR Avantor, 2010. Pentafluorobenzyl propionate. https://www.us.vwr.com/store/product/9003447/pentafluorobenzyl-propionate-97?sessionid=ZgLs7DiGZStifFoV3XJx2WNWvstore8a. Accessed 2017.
Williams, H.P., 1985. Model Building in Mathematical Programming, 2nd ed. Wiley & Sons, California.
Yeomans, H., Grossmann, L.E., 1999. A systematic modeling framework of superstructure optimization in process synthesis. Comput. Chem. Eng. 23, 709–731.
Yunus, N.A., Gernaey, K.V., Woodley, J.M., Gani, R., 2013. Design of sustainable blended products using an integrated methodology. Comput. Aided Chem. Eng. 32, 835–840.
Yunus, N.A., Gernaey, K.V., Woodley, J.M., Gani, R., 2014. A systematic methodology for design of tailor-made blended products. Comput. Chem. Eng. 66, 201–213.
Zhang, L., Cignitti, S., Gani, R., 2015. Generic mathematical programming formulation and solution for computer-aided molecular design. Comput. Chem. Eng. 78, 79–84.