

# The Formulation of Optimal Mixtures with Generalized Disjunctive Programming: A Solvent Design Case Study

Suela Jonuzaj, Paul T. Akula, Polyxeni-Margarita Kleniati, and Claire S. Adjiman  
 Dept. of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London,  
 South Kensington Campus, London SW7 2AZ, U.K.

DOI 10.1002/aic.15122

Published online January 28, 2016 in Wiley Online Library (wileyonlinelibrary.com)

*Systematic approaches for the design of mixtures, based on a computer-aided mixture/blend design (CAM<sup>b</sup>D) framework, have the potential to deliver better products and processes. In most existing methodologies the number of mixture ingredients is fixed (usually a binary mixture) and the identity of at least one compound is chosen from a given set of candidate molecules. A novel CAM<sup>b</sup>D methodology is presented for formulating the general mixture design problem where the number, identity and composition of mixture constituents are optimized simultaneously. To this end, generalized disjunctive programming is integrated into the CAM<sup>b</sup>D framework to formulate the discrete choices. This generic methodology is applied to a case study to find an optimal solvent mixture that maximizes the solubility of ibuprofen. The best performance in this case study is obtained with a solvent mixture, showing the benefit of using mixtures instead of pure solvents to attain enhanced behavior.* © 2016 The Authors AICHE Journal published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers AICHE J, 62: 1616–1633, 2016

*Keywords:* computer-aided mixture design, generalized disjunctive programming, optimal mixture design, ibuprofen, solubility

## Introduction

Mixtures play an important role in the process industries and blends of refrigerants,<sup>1–3</sup> polymers,<sup>4,5</sup> and solvents<sup>6</sup> are used in a wide range of applications. Solvent mixtures, for example, are used in separation processes, such as extraction,<sup>7,8</sup> absorption,<sup>9</sup> and crystallisation,<sup>10,11</sup> and in chemical reactions.<sup>12,13</sup> In product design, the desired performance can often only be achieved with a mixture or formulation (e.g., pesticide formulation, crude oils blended into a single product).<sup>14,15</sup> The current regulatory environment is making mixtures increasingly relevant as restrictions are placed on the use of a growing number of compounds. Some common compounds are thus being removed from use as a result of changing regulations (e.g., REACH regulations<sup>16</sup>). Given this context, the formulation of mixtures offers a potential route to enhanced performance, because mixtures can exhibit properties that equate or even surpass those of pure compounds. Some of the benefits of using mixtures have been demonstrated by Granberg and Rasmuson<sup>17</sup> who studied the solubility of paracetamol in a binary mixture of water and acetone. The results of their study are shown in Figure 1, where it

can be seen that, at 30°C, mixtures containing up to 75% water by mass achieve at least as high a solubility as pure acetone, with a 30:70 mixture of water and acetone achieving the highest solubility. A fivefold increase in solubility relative to pure acetone is observed, although paracetamol is poorly water-soluble. This nonlinear behavior, which is commonly observed in solubility experiments (e.g., the work of Pacheco et al.<sup>18</sup>), arises from the nonideal thermodynamics of the ternary mixture.<sup>19,20</sup>

There is thus a crucial need for identifying mixtures that can achieve better performance than pure compounds while being more environmentally benign. Despite the importance of mixtures and their direct impact on the performance and sustainability of products and processes, the design of optimal mixtures, which entails the selection of appropriate components and of their composition, remains challenging. The choice of components is often made based on expensive and time-consuming experiments or database searches,<sup>21</sup> using databases that include thermodynamic property data collections for pure components and mixtures such as the CAPEC<sup>22</sup> database, CHEMSAFE,<sup>23</sup> DECHEMA,<sup>24</sup> DETHERM,<sup>25</sup> DIPPR,<sup>26</sup> HSSDS,<sup>27</sup> NIST,<sup>28</sup> SOLV-DB,<sup>29</sup> SSDS,<sup>30</sup> TAPP.<sup>31</sup> To broaden the search for better mixtures several more systematic approaches for mixture design<sup>10,11,14,32–34</sup> based on computer-aided mixture/blend design (CAM<sup>b</sup>D), have been developed. CAM<sup>b</sup>D has been defined by Gani and coworkers<sup>10,14,35</sup> as the problem of identifying, from a set of given molecules, a mixture with desired properties that optimizes a given performance measure. The general objective of CAM<sup>b</sup>D is to design the optimal number, identity and compositions of the components that participate in a mixture that meets the design criteria. The vast majority of approaches developed can be used in the context of the hybrid computer-aided

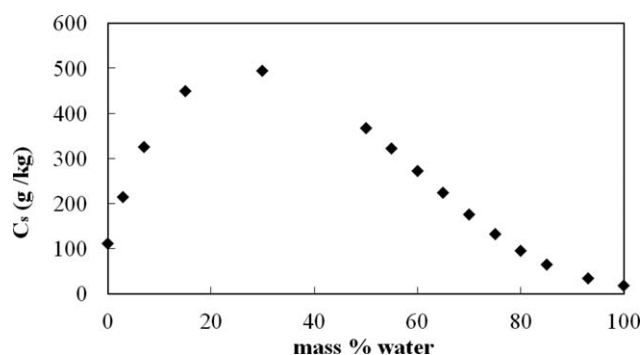
Current address of [Paul T. Akula]: Department of Chemical Engineering, West Virginia University, Morgantown, WV 26506, USA.

Current address of [Polyxeni-Margarita Kleniati]: Process Systems Enterprise Ltd., 26-28 Hammersmith Grove, London W6 7HA, U.K.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Correspondence concerning this article should be addressed to Claire S. Adjiman at c.adjiman@imperial.ac.uk.

© 2016 The Authors AICHE Journal published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers



**Figure 1. Solubility,  $C_s$ , of paracetamol in a mixture of water + acetone, at 30°C, as a function of the mass percent of water in the solvent. Units of solubility: g of paracetamol/kg of solvent.<sup>17</sup>**

molecular design (CAMD) methodology for molecule or mixture design problems proposed by Harper et al.<sup>36</sup> It involves a three-step design process that includes first a pre-design phase where the problem is defined, then a design phase that consists of optimizing a performance index subject to property constraints, and finally a post-design phase where verification and analysis of the results obtained from the design phase take place.

Some studies in the area of CAM<sup>b</sup>D have been focused on developing new methodologies for formulating and solving the mixture design problem. One set of approaches is based on the use of mixed-integer nonlinear programming (MINLP), and this has mostly been applied to the design of binary mixtures. Such a methodology was presented by Duvedi and Achenie<sup>1</sup> who studied the design of environmentally friendly refrigerant mixtures. The authors proposed a mathematical programming problem where the identity of candidate molecules and of the components in the mixture are defined by binary variables, whereas continuous decision variables are used to represent mixture properties and composition. This design methodology was also used by Churi and Achenie<sup>2</sup> to design optimal refrigerant mixtures that have the highest cooling effect in a double-evaporator refrigeration system. In this study, optimal binary mixtures that can give higher efficiencies were identified for a refrigeration cycle with two evaporators operating at two different temperatures. Sioukrou et al.<sup>37</sup> investigated the design of binary solvent mixtures as part of conceptual process design. Their approach focused on the design of a CO<sub>2</sub>-expanded solvent and its impact on process performance. They used enumeration to solve the resulting MINLP due to the small number of discrete choices.

A general and systematic methodology based on an MINLP formulation was also proposed by Vaidyanathan and El-Halwagi<sup>4</sup> for the design of polymer blends, where binary polymer mixtures that match a set of target properties were determined. In the resulting design problem, the identity and the compositions of the components in the mixture were considered as decision variables. Local solutions of the problem were obtained with commercial software packages and an example of small dimensionality was solved globally using an optimization algorithm based on interval analysis.

Much effort has also been devoted to developing strategies that can address the complexity of the MINLP mixture design problem. An interval-analysis based optimization framework was developed by Sinha et al.<sup>38,39</sup> to solve such problems. In their study, an eight-step interval-based domain reduction

algorithm, LIBRA, was developed and used successfully to identify the globally optimal binary mixture in the design of environmentally acceptable blanket wash solvent blend. The solvents that participate in the mixture were selected from a list of promising candidates.

Several authors have developed decomposition-based approaches in which the search space is gradually narrowed. A decomposition-based computer-aided molecular/mixture design methodology was proposed by Karunanithi et al.<sup>10</sup> Within their framework, the original mixture design problem is decomposed into the five following nonlinear subproblems: (1) structural constraints, (2) pure component property constraints, (3) mixture property constraints, (4) miscibility constraints, and (5) process model constraints and objective function (MINLP subproblem). In each subproblem, pure components and/or mixtures that do not satisfy a subset of the constraints of the original problem are eliminated. This leads to a smaller final MINLP problem, in which a large portion of the search region has been eliminated. This decomposition-based methodology was later applied by the same authors to the design of crystallization solvents.<sup>11</sup> Their study involved the design of optimal binary solvent mixtures that maximize the potential recovery of a drug, subject to several property constraints, such as crystal morphology, solubility, viscosity, toxicity, normal boiling point and melting point.

Buxton et al.<sup>40</sup> proposed a systematic decomposition-based procedure to select optimal solvent blends for nonreactive, multicomponent gas absorption processes. Their approach was based on an extension of the work of Pistikopoulos and Stefanis,<sup>9</sup> who considered the design of pure solvents for environmental impact minimization. Within their proposed framework, process operations that make use of solvents are identified first and the solvent candidates are then determined subject to specific property and environmental constraints. Finally, the performance of the solvents is verified on a plant-wide basis and the optimal solvent candidate is selected. The extension of this formulation to the design of binary solvent mixtures<sup>40</sup> requires the inclusion of additional constraints on the physical properties and the operating conditions.

In the recent work of Papadopoulos et al.,<sup>33</sup> a CAM<sup>b</sup>D approach was developed as a Multi-Objective Optimization problem (MOO) for obtaining the optimal binary fluid mixture for organic Rankine cycles. In this approach, the two compounds and their optimal composition in the mixture are designed simultaneously using a two-stage methodology. In the first part, the molecular structure of a pure compound that matches a set of properties and yields the best performance measure is designed; this compound is then selected to be the first component in the binary mixture. The second stage consists of designing a number of feasible molecules for the second component in the mixture and defining the optimal mixture composition. The proposed mixture design methodology is followed by a nonlinear sensitivity analysis to evaluate the effect of the uncertainties arising from the model and in particular from the use of group contribution methods. A useful feature of this approach is that the first component is guaranteed to be a good fluid, so gives good baseline performance, and the second component is guaranteed to provide a performance enhancement, regardless of its performance as a pure compound. In principle, this means that, if this framework was applied to the design of a solvent mixture that maximizes the solubility of paracetamol, a binary mixture with the characteristics of the acetone and water mixture (of Figure 1) could be

identified. Acetone, which is considered to be a good solvent, could be identified in the first part of the methodology and water, which improves the overall performance, could be determined in the second part.

Only a few methods have been proposed to address the design of mixtures with more than two components. This issue was studied systematically by Klein et al.,<sup>41</sup> who proposed a successive regression and linear programming algorithm for the solution of the nonlinear programming formulation of the problem. In this work, the objective was to determine the minimum cost solvent mixture, subject to linear constraints on the solubility parameters and nonlinear constraints on the density and boiling-point temperature. The candidate solvents were selected from a predefined set of molecules.

Recently, Ng et al.<sup>42</sup> presented a two-stage methodology for mixture design in an integrated biorefinery. The first stage consists of the mixture design framework, where optimal mixtures are formulated based on standard CAM<sup>b</sup>D techniques. First, the component that performs the main functionality of the mixture is identified from given products or designed with respect to physico-chemical properties and structural constraints. Next, based on the target properties (product needs), the number of components to participate in the final mixture is defined, and suitable additive components that meet these properties are then designed. In the final part of this stage, the miscibility of the mixture components is investigated. In the second stage, the optimal biomass conversion pathways that produce the optimal mixtures determined in the first stage are identified using a superstructure optimization approach. The design methodology was applied to case study for the design of biofuels from palm-based biomass.

A systematic four-step methodology applicable to more than two components was proposed recently by Yunus et al.<sup>34</sup> for the design of blended liquid products. The first step consists of the definition of the problem, where the product needs are identified and translated into physico-chemical target properties, and target values for these properties are determined. In the next step, a set of property models is retrieved from the model library to allow the prediction of pure component and mixture target properties. The third step involves the design of multi-component mixtures based on a decomposition methodology, where pure components that satisfy the property constraints are first identified and then a stability analysis is performed to define possible mixtures. The third step is concluded by optimizing the performance objective subject to the linear and nonlinear property models. The mixture design methodology is applied to binary and ternary mixtures. In the fourth and final step, rigorous models are used to verify the mixture property values, resulting in a set of optimal blends that satisfy all property targets. The proposed methodology was applied to two case studies: (a) designing gasoline blends that can be used in car engines in hot climates and (b) designing environmentally friendly base-oil mixtures that have good lubrication properties, from organic chemicals and mineral oils.

In spite of these advances, there remains significant scope for further research in this area. Most existing methodologies are applicable to binary mixtures only,<sup>1,2,4,10,11,33,38–40</sup> with the exception of the work of Klein et al.<sup>41</sup> and Yunus et al.<sup>34</sup> Furthermore, in many methods<sup>1,2,10,34,38,41</sup> the space of possible components is restricted. In some cases, all components that participate in a mixture are selected from a given set of molecules, while in others, one of the mixture ingredients is defined *a priori*, and the other compounds (usually one

compound) are designed or selected from a set. In these methodologies, the number and the identity of one component (or of all components) are usually chosen in advance, and then the identities of the remaining molecules in a mixture, their compositions, and, where relevant, process topology and operating conditions, are defined. In this sequential approach, where the desired number of mixture ingredients is specified and a single compound is selected from a given list, there is a risk of excluding from the design space molecular structures which, when combined in a mixture, can lead to better performance. Only a small number of studies are reported in the literature in which the simultaneous design of all compounds is considered.<sup>33,40</sup> However, these studies refer to problems where the number of mixture ingredients is fixed to two. In summary, a reduced version of the general CAM<sup>b</sup>D problem has been addressed to date: the number of mixture ingredients is fixed (usually 2) and the identity of a compound (or in a few cases, of all compounds) that can participate in a mixture is chosen from a given set of candidate molecules.

A hurdle in the further development and widespread use of tools for CAM<sup>b</sup>D is the complexity of the mathematical programs that need to be formulated and solved. When modeling a mixture design problem directly as a MINLP problem, several numerical difficulties can arise, related to the nonlinearity (nonconvexity) of the property models and the large design space. Solving the optimization problems can be quite challenging, as these are combinatorial due to the presence of binary variables, and highly nonlinear due to the expressions that relate composition, structure and physical properties.<sup>12</sup>

In view of these challenges, the main purpose of this article is to create a comprehensive and systematic mathematical programming approach for the formulation and solution of the general CAM<sup>b</sup>D problem. To address the difficulties arising from the complexity of expressing the problem within a mathematical framework, we adopt a logic-based methodology in which generalized disjunctive programming (GDP)<sup>43</sup> is used to formulate the discrete choices inherent in mixture design problems. In this framework, we first show how to formulate a design problem in which the number of components ( $N$ ) in the mixture is fixed and these components are selected from a predefined set of molecules. In working toward the generalized CAM<sup>b</sup>D problem, we focus on making  $N$  a design variable where at most  $N_{\max}$  compounds are chosen from a given list. This design methodology is an initial approach to a more general concept where the simultaneous design of the number, identity and compositions of mixture ingredients will be considered. Our proposed logic-based approach fits within the broader framework proposed by Harper et al.<sup>36</sup> and can be used in the second (design) step of such an approach.

This article is organized as follows: first, the background theory is given, including a brief introduction to the CAM<sup>b</sup>D framework, followed by a short description of GDP. The general methodology used in this work and a detailed description of the proposed formulations are presented in the next section. The concepts are applied to a case study of solvent mixture design. The optimal solvent mixture that maximizes the solubility of ibuprofen is identified under different scenarios, with the components being selected from a list of compounds. The conclusions of this work are summarized in the final section.

## Modeling Approaches

### Computer-aided mixture/blend design (CAM<sup>b</sup>D)

The many systematic approaches that have been derived for the design of compounds that exhibit desirable performance

are known collectively as CAMD methods. The CAMD concept was initially introduced by Gani and Brignole in 1983<sup>7</sup> and there has since been significant progress toward this goal.<sup>9,10,14,33,35,36,44–46</sup> In the case of mixture design applications, a CAMD problem is expanded into computer-aided mixture/blend design (CAM<sup>b</sup>D) problem, usually by including additional mixture property constraints in a “standard” MINLP CAMD problem. Achenie and coworkers<sup>10,14,35</sup> defined CAM<sup>b</sup>D as follows:

“Given a set of chemicals and a specified set of property constraints, determine the optimal mixture and/or blend.”

The main objectives of the CAM<sup>b</sup>D framework focus on optimizing the physical properties of mixtures or compounds<sup>47,48</sup> and/or on optimizing process performance, such as minimizing process cost<sup>49</sup> or maximizing production.<sup>50,51</sup>

As discussed in the introduction, most existing CAM<sup>b</sup>D methods involve formulating and solving a problem using mathematical programming techniques; usually a MINLP problem. According to Karunanithi et al.,<sup>10</sup> a general CAM<sup>b</sup>D problem can be formulated as an MINLP problem as follows)

$$\begin{aligned} & \min_{x,y} f(x,y) \\ & \text{subject to} \\ & g_1(y) \leq 0 \\ & g_2(y) \leq 0 \\ & g_3(x,y) \leq 0 \\ & g_4(x,y) = 0 \\ & x \in \mathbb{R}^n, y \in \{0, 1\}^m \end{aligned} \quad (1)$$

where  $f$  is the objective function to be optimized (minimization is assumed without loss of generality), subject to structural constraints ( $g_1(y)$ ), pure component property constraints ( $g_2(y)$ ), mixture property constraints ( $g_3(x,y)$ ), and process model constraints ( $g_4(x,y)$ ). The vector  $x$  is an  $n$ -dimensional vector of continuous variables denoting operating conditions, physical properties, or process variables (e.g., flow rate, temperature) and  $y$  is an  $m$ -dimensional vector of binary variables related to the functional groups present in the molecules and/or the identities of the molecules.<sup>10,35</sup> While the general form of a CAM<sup>b</sup>D problem is relatively simple and similar to that of a CAMD problem, the formulation of a specific instance of CAM<sup>b</sup>D as an MINLP is a challenging task because of the different types of binary or integer variables involved and their strong interactions with the often-nonlinear physical property models.

### Generalized disjunctive programming

An alternative approach for formulating discrete/continuous optimization problems, known as GDP, was introduced by Raman and Grossmann<sup>43</sup> and has since been successfully applied by Grossmann and coworkers in many different contexts including complex process networks,<sup>43,52</sup> strip-packing,<sup>53</sup> and scheduling problems.<sup>43,53,54</sup> GDP is a logic-based methodology that extends the disjunctive programming proposed by Balas<sup>55</sup> and involves Boolean and continuous variables that are related via disjunctions, algebraic equations, and logic propositions.<sup>56–60</sup>

**GDP Formulation.** The general formulation of a generalized disjunctive program can be represented as follows

$$\begin{aligned} & \min_{x,Y} f(x) \\ & \text{s.t. } g(x) \leq 0 \\ & \bigvee_{j \in J_k} \begin{bmatrix} Y_{j,k} \\ h_{j,k}(x) \leq 0 \end{bmatrix} \quad k \in K \\ & \bigvee_{j \in J_k} Y_{j,k}, \quad k \in K \\ & \Omega(Y) = \text{True} \\ & x \in [x^L, x^U] \subset \mathbb{R}^n \\ & Y_{j,k} \in \{\text{True}, \text{False}\}, j \in J_k; k \in K \end{aligned} \quad (\text{GDP})$$

where  $x$  is a vector of continuous variables,  $x^L$  and  $x^U$  are vectors denoting the lower and upper bounds on  $x$ , respectively, and the  $Y_{j,k}$ 's are the Boolean variables that indicate whether a term  $j$  in a disjunction  $k$  is True or False. The objective function  $f(x)$  depends on the continuous variables, and the logic in the continuous space is represented by a set of disjunctions. Each disjunction consists of a number of terms that are joined by the OR ( $\vee$ ) operator. Each term of the disjunction has a Boolean variable  $Y_{j,k}$  and only one of the Boolean variables can be True in each disjunction ( $\bigvee_{j \in J_k} Y_{j,k}$ ).  $K$  is the index set for the disjunctions and  $J_k$  is the index set of the terms in each disjunction  $k \in K$ . The function  $g(x)$  represents general constraints that must hold regardless of the logic, while  $h_{j,k}(x)$  are conditional constraints that hold when  $Y_{j,k}$  is True. In mixture design problems the disjunctive constraints are related to the assignment of compounds and the number of components that participate in a mixture.  $\Omega(Y)$  represents logic relations for the Boolean variables expressed as propositional logic.<sup>61</sup>

It should be mentioned that any MINLP problem can be formulated as a GDP problem and vice versa. It may be beneficial to adopt a GDP formulation because, when compared with mixed-integer programming, it provides a more structured framework for modeling discrete-continuous choices and it expresses more directly both the quantitative and the qualitative parts of the optimization task.<sup>62,63</sup> In an MINLP problem (Eq. 1) the logic needs to be expressed through the objective function and algebraic constraints, in the form  $f(x, y)$  and  $g(x, y)$ , respectively. In GDP, conversely, the logic is captured inside the disjunctions by relating Boolean variables ( $Y_{j,k}$ ) to equations in the continuous form ( $h_{j,k}(x)$ ), whereas the logic that connects the disjunctive sets is expressed through the relations  $\Omega(Y)$ .<sup>64</sup> To formulate a general mixture design problem (Eq. 1) as a GDP, several characteristics of the constraints must be taken into account. The constraints that do not depend on the logic conditions can be formulated as general constraints ( $g(x)$ ), whereas the constraints that depend on the logic conditions, such as on the assignment of compounds or on the number of components in a mixture, are formulated within the disjunctions as conditional constraints ( $h_{j,k}(x)$ ).

**Reformulation of GDP as an MINLP.** Once an appropriate GDP formulation has been obtained, it can be converted into an MINLP problem using different approaches, such as big-M or hull relaxation, that result in relaxations of varying strength.<sup>54,64,65</sup> The big-M (BM) formulation<sup>66</sup> is the simplest representation of a GDP problem in a mixed-integer form.<sup>43,60</sup>

The general reformulation of a GDP as an MINLP via big-M is given by

$$\begin{aligned}
 & \min_{x,y} f(x) \\
 & \text{s.t. } g(x) \leq 0 \\
 & h_{j,k}(x) \leq M_{j,k}(1-y_{j,k}), j \in J_k; k \in K \\
 & Ay \leq b \\
 & \sum_{j=1}^{J_k} y_{j,k} = 1, k \in K \\
 & x \in [x^L, x^U] \subset \mathbb{R}^n; y_{j,k} \in \{0, 1\}, j \in J_k; k \in K
 \end{aligned} \tag{BM}$$

where  $y$  is a vector of binary variables, which has one-to-one correspondence with the Boolean variable vector,  $Y$ ,  $A$  is an  $m \times n$  matrix,  $b$  is an  $m$ -dimensional real-valued vector and the parameter  $M_{j,k}$  is a “sufficiently large” upper bound. The logic propositions in GDP,  $\Omega(Y)=\text{True}$ , have been converted into linear inequalities,  $Ay \leq b$ .<sup>67</sup> The tightest value for  $M_{j,k}$  can be calculated as<sup>52</sup>

$$M_{j,k} = \max \{h_{j,k}(x) \mid x^L \leq x \leq x^U\}, j \in J_k; k \in K \tag{2}$$

### GDP Formulation of the CAM<sup>b</sup>D Problem

The design methodology proposed in our work integrates GDP into a computer-aided mixture/blend design (CAM<sup>b</sup>D) framework.

#### Problem definition

The aim of this study involves the generic formulation of mixture design problems to find the optimal number of mixture ingredients, the optimal identities of the components (chosen from a given list) and their compositions, such that all given specifications are satisfied and the specified performance objective is optimized.

The problem formulation is constructed in a systematic way by considering two problem statements of increasing complexity. The first class of problems involves the formulation of a restricted model, where a fixed number of components is identified from a given set of candidate compounds, subject to property constraints. The design variables are the identity of the components that participate in the mixture and their compositions. After establishing the solution of this restricted problem, we propose a second, more general, formulation, where the number of components in the mixture is not fixed but is bounded by an upper limit. The optimal components are again selected from a predefined set of compounds. In the general problem the decision variables are the number of components, their identities, and their compositions.

To develop the formulations, we define several index sets. The first is the set of components in the mixture,  $I = \{i \mid i=1, \dots, N_c\}$ , where  $N_c$  is the total number of components. The second set  $S = \{s \mid s=1, \dots, N_s\}$  defines the list of compounds from which the mixture components must be chosen, where  $N_s$  is the total number of molecules in the list. The mixture to be designed often contains components that are known *a priori* (e.g., solutes in the case of solvent design). We define  $N'$  as the number of fixed components and  $N$  as the (maximum) number of components to be designed. The total number of components in the mixture is thus  $N_c = N + N'$ . For clarity, we use the term “components” to refer to the ingre-

dients/molecules in the mixture we are designing and the term “compounds” to refer to ingredients/molecules in the set  $S$  from which we choose the components. Those components in the mixture that are not fixed (i.e., components  $N'+1$  to  $N_c$ ) are referred to as the “designed components.”

#### Formulation for known number $N_c$ of components in mixture

The discrete choices of the restricted mixture design problem with a fixed value of  $N$  are modeled using GDP. The GDP formulation of the restricted problem includes disjunctions for each choice of components and logic propositions that express the relationships between the disjunctive sets.

*Disjunctions for Assignment of Components.* In a logic-based modeling framework, the assignment of component  $i$  is determined through Boolean variable  $Y_{i,s}$ , which is True if a compound  $s$  is assigned to component  $i$  in the mixture and False if it is not, for  $i=N'+1, \dots, N_c$  and  $s=1, \dots, N_s$ . The disjunctions for the assignment of each component are given below

$$\begin{aligned}
 & \bigvee_{s \in S} \left[ \begin{array}{c} Y_{i,s} \\ h_{i,s}(x) \leq 0 \end{array} \right], \quad i=N'+1, \dots, N_c \\
 & \bigvee_{s \in S} Y_{i,s}, \quad i=N'+1, \dots, N_c \\
 & Y_{i,s} \in \{\text{True}, \text{False}\}
 \end{aligned} \tag{R-D}$$

The vector  $h_{i,s}$  in each disjunction represents the constraints that are active when compound  $s$  is assigned to  $i^{\text{th}}$  component in a mixture. The expression  $\bigvee_{s \in S} Y_{i,s}$  ensures that only one disjunction is active, which means that each component is assigned exactly one compound. It can be translated into algebraic equations by introducing a binary variable  $y_{i,s}$  as follows<sup>68</sup>

$$\sum_{s \in S} y_{i,s} = 1, \quad i=N'+1, \dots, N_c \tag{3}$$

*Logic Propositions.* Logic conditions ( $\Omega(Y)=\text{True}$ ) are included to avoid degeneracy by enforcing a specific ordering of the compounds. Degeneracy can be prevented by the following relations that ensure that the relative position of a compound in the set  $S$  is maintained in the mixture (set  $I$ ) if the compound is selected

$$\begin{aligned}
 Y_{i,s} \Rightarrow \neg Y_{i',s'}, \quad i=N'+1, \dots, N_c-1; \quad s=2, \dots, N_s \\
 i'=i+1, \dots, N_c; \quad s'=1, \dots, s
 \end{aligned} \tag{4}$$

where the symbol  $\neg$  implies negation (i.e., not  $Y_{i',s'}$  or  $Y_{i',s'}=\text{False}$ ). They are translated into algebraic equations as follows

$$\begin{aligned}
 y_{i,s} + y_{i',s'} \leq 1, \quad j=N'+1, \dots, N_c-1; \quad s=2, \dots, N_s \\
 i'=i+1, \dots, N_c; \quad s'=1, \dots, s.
 \end{aligned} \tag{5}$$

The constraints in Eq. 5 restrict the feasible space by eliminating identical degenerate solutions. Logic conditions are also derived to ensure that each candidate compound is selected at most once

$$Y_{i',s} \Rightarrow \neg \left( \bigvee_{\substack{i=N'+1, \dots, N_c-1 \\ i \neq i'}} Y_{i,s} \right), \quad i'=N'+1, \dots, N_c; \quad s \in S \tag{6}$$

This is equivalent to

$$\sum_{i=N'+1}^{N_c} y_{i,s} \leq 1, \quad s \in S \quad (7)$$

**GDP Formulation.** The GDP formulation of the restricted problem is thus written as

$$\begin{aligned} & \min_{x,Y} f(x) \\ & \text{s.t. } g(x) \leq 0 \\ & \bigvee_{s \in S} \begin{bmatrix} Y_{i,s} \\ h_{i,s}(x) \leq 0 \end{bmatrix}, \quad i=N'+1, \dots, N_c \\ & \bigvee_{s \in S} Y_{i,s}, \quad i=N'+1, \dots, N_c \\ & Y_{i,s} \Rightarrow \neg Y_{i',s'}, \quad i=N'+1, \dots, N_c-1; s=2, \dots, N_s \\ & \quad \quad \quad i'=i+1, \dots, N_c; s'=1, \dots, s \\ & Y_{i',s'} \Rightarrow \neg \left( \bigvee_{\substack{i=N'+1, \dots, N_c-1 \\ i \neq i'}} Y_{i,s} \right), \quad i'=N'+1, \dots, N_c; s \in S \\ & x_i \in [x_i^L, x_i^U] \subset \mathbb{R}; Y_{i,s} \in \{\text{True}, \text{False}\}, \quad i=N'+1, \dots, N_c; s \in S \end{aligned} \quad (\text{R-GDP})$$

**Reformulation of GDP as an MINLP.** The GDP model is reformulated as an MINLP by replacing the Boolean variables  $Y_{i,s}$  with binary variables  $y_{i,s}$ . The conditional constraints,  $h_{i,s}(x)$ , are reformulated using the big-M approach, where a large  $M$  parameter is introduced, and the logic conditions,  $\Omega(Y)$ , are converted into linear inequality constraints, following to the methodology proposed by Raman and Grossmann.<sup>68</sup> The MINLP model derived by applying the big-M reformulation is written as follows

$$\begin{aligned} & \min_{x,y} f(x) \\ & \text{s.t. } g(x) \leq 0 \\ & h_{i,s}(x) \leq M_{h_{i,s}}(1-y_{i,s}), \quad s \in S; i=N'+1, \dots, N_c \\ & \sum_{s \in S} y_{i,s} = 1, \quad i=N'+1, \dots, N_c \\ & y_{i,s} + y_{i',s'} \leq 1, \quad i=N'+1, \dots, N_c-1; s=2, \dots, N_s \\ & \quad \quad \quad i'=i+1, \dots, N_c; s'=1, \dots, s \\ & \sum_{i=N'+1}^{N_c} y_{i,s} \leq 1, \quad s=1, \dots, N_s \\ & x_i \in [x_i^L, x_i^U] \subset \mathbb{R}; y_{i,s} \in \{0, 1\}, \quad i=N'+1, \dots, N_c; s \in S \end{aligned} \quad (\text{R-BM})$$

where the binary variable  $y_{i,s}$  represents the identity of the component  $i$  in the mixture and takes the value 1 if compound  $s$  is selected as component  $i$  and 0 otherwise; the continuous variable  $x_i$  represents the mole fraction of component  $i$ , while  $x_i^L$  and  $x_i^U$  are the lower and upper bounds on the mole fraction, respectively. The objective function,  $f(x)$ , is optimized subject to general constraints,  $g(x)$ , and conditional constraints,  $h_{i,s}(x)$ .

### Formulation for unknown number $N_c$ of components in mixture

The general problem results in a more generic formulation where in addition to the identities and compositions of the components in the mixture, the optimal number of mixture ingredients also needs to be defined. As the number of selected compounds that can participate in a mixture is not fixed but is allowed to vary up to a maximum number,  $N_{\max}$ , additional

disjunctions involving property conditions that depend on the number of components are included in this model. The maximum total number of mixture ingredients can be expressed as  $N_c = N_{\max} + N'$ , where  $N'$  is the number of fixed components. The GDP formulation of the general problem includes disjunctions for the component assignments, disjunctions for the number of components in the mixture, and logic propositions that express the relationships between the disjunctive sets.

**Disjunctions for Assignment of Components.** The same Boolean variables,  $Y_{i,s}$ , as in the restricted problem, are used to establish the selection of a compound for each component for  $i=N'+1, \dots, N_c$  and  $s=1, \dots, N_s$ . The disjunctions for assigning each component are given by the expression (R-D). In the general formulation at least one designed component should be present in the mixture (i.e., exactly one Boolean variable for assigning a compound to the first component is active). The assignment of exactly one compound to the first designed component in the mixture is given by

$$\bigvee_{s \in S} Y_{N'+1,s} \quad (8)$$

and its algebraic form is

$$\sum_{s \in S} y_{N'+1,s} = 1. \quad (9)$$

The rest of the components are assigned compounds from the list only if they are participating in the mixture.

**Disjunctions for the Number of Components.** The number of components in the mixture is a discrete choice and it is expressed as a series of disjunctions. A vector  $\tilde{Y}_n$  of Boolean variables that express the logic in each disjunction is introduced. Property constraints that depend on the number of components in the mixture are included in each disjunction and they are activated when the corresponding Boolean variable  $\tilde{Y}_n$  takes the value of True. The disjunctions for the number of designed components in the mixture are given by

$$\bigvee_{n=1, \dots, N_{\max}} \begin{bmatrix} \tilde{Y}_n \\ \tilde{F}_n(x) \leq 0 \\ x_i \geq x_i^L, \quad i=N'+2, \dots, N'+n \\ x_i = 0, \quad i=N'+n+1, \dots, N_c \end{bmatrix} \quad (\text{G-D})$$

where  $\tilde{F}_n(x)$  is a vector of disjunctive constraints that depend on the number of components in the mixture and are active if  $\tilde{Y}_n = \text{True}$ ;  $x_i$  is the mole fraction of component  $i$  in the mixture and is greater than a threshold value  $x_i^L$  if the component is present in the mixture and is zero otherwise. Exactly one disjunction for the number of designed components must be selected, as shown below

$$\bigvee_{n=1, \dots, N_{\max}} \tilde{Y}_n \quad (10)$$

and it is transformed into the following constraint

$$\sum_{n=1}^{N_{\max}} \tilde{y}_n = 1 \quad (11)$$

where  $\tilde{y}_n$  is a binary variable equivalent to  $\tilde{Y}_n$ .

**Logic Propositions.** Logic conditions to avoid degeneracy are also required in the general formulation. A specific ordering of the compounds is enforced using the same logic propositions as were described in the restricted problem, i.e., Eq. 4. Equation 6 is also used in the general model to ensure that each candidate compound is selected at most once. Additional

**Table 1. Logic Propositions and Algebraic Constraints for the General Model**

Logic Expressions	Linear Inequalities
$\neg \tilde{Y}_1 \vee (\bigvee_{s \in S} Y_{N'+1,s})$	$\tilde{y}_1 \leq \sum_{s \in S} y_{N'+1,s}$
$(\neg \tilde{Y}_1 \vee \neg Y_{N'+2,s}), s \in S$	$\tilde{y}_1 + y_{N'+2,s} \leq 1, s \in S$
$\vdots$	$\vdots$
$(\neg \tilde{Y}_1 \vee \neg Y_{N'+N_{\max},s}), s \in S$	$\tilde{y}_1 + y_{N'+N_{\max},s} \leq 1, s \in S$
$\neg \tilde{Y}_2 \vee (\bigvee_{s \in S} Y_{N'+1,s})$	$\tilde{y}_2 \leq \sum_{s \in S} y_{N'+1,s}$
$\neg \tilde{Y}_2 \vee (\bigvee_{s \in S} Y_{N'+2,s})$	$\tilde{y}_2 \leq \sum_{s \in S} y_{N'+2,s}$
$(\neg \tilde{Y}_2 \vee Y_{N'+3,s}), s \in S$	$\tilde{y}_2 + y_{N'+3,s} \leq 1, s \in S$
$\vdots$	$\vdots$
$(\neg \tilde{Y}_2 \vee \neg Y_{N'+N_{\max},s}), s \in S$	$\tilde{y}_2 + y_{N'+N_{\max},s} \leq 1, s \in S$
$\vdots$	$\vdots$
$\neg \tilde{Y}_{N_{\max}} \vee (\bigvee_{s \in S} Y_{N'+1,s})$	$\tilde{y}_{N_{\max}} \leq \sum_{s \in S} y_{N'+1,s}$
$\neg \tilde{Y}_{N_{\max}} \vee (\bigvee_{s \in S} Y_{N'+2,s})$	$\tilde{y}_{N_{\max}} \leq \sum_{s \in S} y_{N'+2,s}$
$\vdots$	$\vdots$
$\neg \tilde{Y}_{N_{\max}} \vee (\bigvee_{s \in S} Y_{N'+N_{\max},s})$	$\tilde{y}_{N_{\max}} \leq \sum_{s \in S} y_{N'+N_{\max},s}$

logic conditions are required to ensure that at most one compound is assigned to components  $N'+2$  to  $N_c$

$$Y_{i,s'} \Rightarrow \neg \left( \bigvee_{\substack{s=1, \dots, N_s-1 \\ s \neq s'}} Y_{i,s} \right), \quad s' \in S; \quad i=N'+2, \dots, N_c. \quad (12)$$

This is written equivalently as

$$\sum_{s=1}^{N_s} y_{i,s} \leq 1, \quad i=N'+2, \dots, N_c \quad (13)$$

This differs from the corresponding constraint in the restricted formulation (Eq. 3) where equality was enforced. Logic propositions are also derived to relate the Boolean variables for the number of designed components in the mixture,  $\tilde{Y}_n, n=1, \dots, N_{\max}$  to the Boolean variables for the assignment of a compound to each component,  $Y_{i,s}, i=N'+1, \dots, N_c; s=1, \dots, N_s$ . We represent the number of designed components in the mixture by the Boolean variables  $\tilde{Y}_n (n=1, \dots, N_{\max})$  such that:

$$\begin{aligned} \tilde{Y}_1 &\Rightarrow 1 \text{ designed component } (N=1) \\ \tilde{Y}_2 &\Rightarrow 2 \text{ designed components } (N=2) \\ &\vdots \\ \tilde{Y}_{N_{\max}} &\Rightarrow N_{\max} \text{ designed components } (N=N_{\max}) \end{aligned}$$

Thus,  $\tilde{Y}_1$  (1 designed component) implies only  $Y_{N'+1,s}, s \in S$  (assignment of first designed component), but not  $(Y_{N'+2,s}, \dots, Y_{N'+N_{\max},s}), \tilde{Y}_2$  (2 designed components) implies  $Y_{N'+1,s}$  and  $Y_{N'+2,s}$  (assignment of first and second designed component), but not  $(Y_{N'+3,s}, \dots, Y_{N'+N_{\max},s})$ , and so forth. These relations can be expressed more formally as

$$\begin{aligned} \tilde{Y}_1 &\Rightarrow \left( \bigvee_{s \in S} Y_{N'+1,s} \right) \\ \tilde{Y}_1 &\Rightarrow \neg Y_{N'+2,s} \wedge \dots \wedge \neg Y_{N'+N_{\max},s}, \quad s \in S \\ \tilde{Y}_2 &\Rightarrow \left( \bigvee_{s \in S} Y_{N'+1,s} \right) \wedge \left( \bigvee_{s \in S} Y_{N'+2,s} \right) \\ \tilde{Y}_2 &\Rightarrow \neg Y_{N'+3,s} \wedge \dots \wedge \neg Y_{N'+N_{\max},s}, \quad s \in S \\ &\vdots \\ \tilde{Y}_{N_{\max}} &\Rightarrow \left( \bigvee_{s \in S} Y_{N'+1,s} \right) \wedge \left( \bigvee_{s \in S} Y_{N'+2,s} \right) \wedge \dots \wedge \left( \bigvee_{s \in S} Y_{N'+N_{\max},s} \right) \end{aligned}$$

The above expressions are replaced by their equivalent disjunctions and the ‘‘OR’’ operator is distributed over the ‘‘AND’’ as described by Raman and Grossmann.<sup>68</sup> As shown in Table 1, the resulting clauses can then be expressed as a set of linear inequality constraints by replacing the Boolean variables with binary ones.

**GDP Formulation.** The GDP formulation of the general model can be written as

$$\begin{aligned} \min_{x,Y} & f(x) \\ \text{s.t.} & g(x) \leq 0 \\ & \bigvee_{s \in S} \begin{bmatrix} Y_{i,s} \\ h_{i,s}(x) \leq 0 \end{bmatrix}, \quad i=N'+1, \dots, N_c \\ & \bigvee_{n=1, \dots, N_{\max}} \begin{bmatrix} \tilde{Y}_n \\ \tilde{F}_n(x) \leq 0 \\ x_i \geq x_i^L, \quad i=N'+2, \dots, N'+n \\ x_i=0, \quad i=N'+n+1, \dots, N_c \end{bmatrix} \end{aligned}$$

$$\bigvee_{s=1, \dots, N_s} Y_{N'+1,s}$$

$$\bigvee_{n=1, \dots, N_{\max}} \tilde{Y}_n$$

$$Y_{i,s} \Rightarrow \neg Y_{i',s'}, \quad i=N'+1, \dots, N_c-1; \quad s=2, \dots, N_s$$

$$i'=i+1, \dots, N_c; \quad s'=1, \dots, s$$

$$Y_{i',s} \Rightarrow \neg \left( \bigvee_{\substack{i=N'+1, \dots, N_c-1 \\ i \neq i'}} Y_{i,s} \right), \quad i'=N'+1, \dots, N_c; \quad s \in S$$

$$Y_{i,s'} \Rightarrow \neg \left( \bigvee_{\substack{s=1, \dots, N_s-1 \\ s \neq s'}} Y_{i,s} \right), \quad i=N'+2, \dots, N_c; \quad s' \in S$$

$$\Omega'(Y) = \text{True}$$

$$x_{N'+1} \geq x_{N'+1}^L$$

$$x_i \in [x_i^L, x_i^U] \subset \mathbb{R}, \quad i=N'+1, \dots, N_c$$

$$Y_{i,s}, \tilde{Y}_n \in \{\text{True}, \text{False}\}, \quad i=N'+1, \dots, N_c; \quad s \in S; \quad n=1, \dots, N_{\max} \quad (\text{G-GDP})$$

where  $\Omega'(Y) = \text{True}$  denotes the logic relations in Table 1. As at least one component should be present in the mixture, the mole fraction of the first designed component ( $x_{N'+1}$ ) has always a non-zero value.

**Reformulation of GDP as an MINLP.** Formulation (G-GDP) can be transformed into an MINLP problem using the big-M approach as follows

$$\begin{aligned} \min_{x,y} & f(x) \\ \text{s.t.} & g(x) \leq 0 \\ & h_{i,s}(x) \leq M_{h_{i,s}}(1-y_{i,s}), \quad s=1, \dots, N_s; \quad i=N'+1, \dots, N_c \\ & \tilde{F}_n(x) \leq M_{\tilde{F}_n}(1-\tilde{y}_n), \quad n=1, \dots, N_{\max} \\ & x_i \geq x_i^L \tilde{y}_n, \quad n=1, \dots, N_{\max}; \quad i=N'+2, \dots, N_c \\ & \sum_{s=1}^{N_s} y_{N'+1,s} = 1 \end{aligned}$$

$$\sum_{n=1}^{N_{\max}} \tilde{y}_n = 1$$

$$y_{i,s} + y_{i',s'} \leq 1, \quad i=N'+1, \dots, N_c - 1; \quad s=2, \dots, N_s$$

$$i' = i+1, \dots, N_c; \quad s' = 1, \dots, s$$

$$\sum_{i=N'+1}^{N_c} y_{i,s} \leq 1, \quad s=1, \dots, N_s$$

$$\sum_{s=1}^{N_s} y_{i,s} \leq 1, \quad i=N'+2, \dots, N_c$$

$$Ay \leq b$$

$$x_{N'+1} \geq x_{N'+1}^L$$

$$x_i \in [x_i^L, x_i^U] \subset \mathbb{R}, \quad i=N'+1, \dots, N_c$$

$$\tilde{y}_n, y_{i,s} \in \{0, 1\}, \quad n=1, \dots, N_{\max}; \quad i=N'+1, \dots, N_c; \quad s \in S$$

(G-BM)

The logic relations,  $\Omega'(Y)=\text{True}$ , of the GDP model are converted to a set of algebraic constraints,  $Ay \leq b$ , by replacing the Boolean variables with binary ones, as shown in the second column of Table 1.

The formulations described in this section are applied to a solvent mixture design case study presented in the next section.

### Case Study: Maximizing the Solubility of Ibuprofen

Ibuprofen (ibu) is a colorless anti-inflammatory compound that can be crystallized by cooling crystallisation.<sup>69</sup> Solubility is one of the key properties that determine the performance of the crystallization process.<sup>11,69</sup> Karunanithi et al.<sup>11</sup> have already addressed the problem of identifying appropriate solvents or solvent mixtures that enhance the crystallization process of ibuprofen. This well-studied application is, therefore, a suitable example to investigate the use of the proposed GDP formulations for CAM<sup>b</sup>D. The objective of the design problem considered is to identify an optimal solvent mixture to maximize the solubility of ibuprofen

$$\max x_{\text{ibu}}$$

where  $x_{\text{ibu}}$  is the mole fraction of ibuprofen in the mixture. The first important set of constraints in formulating the problem captures the relationship between the solvent mixture and the solubility of ibuprofen, via solid–liquid equilibrium. Although phase equilibrium relations involve complex nonlinear functions, in this case study only ibuprofen is at solid–liquid equilibrium, whereas all solvent molecules in the mixture are in a single liquid phase. Therefore, the solubility, which depends on the enthalpy of fusion and melting temperature of the solid and its liquid-phase activity coefficient, is expressed in terms of ibuprofen and calculated as follows<sup>70,71</sup>

$$\ln x_{\text{ibu}} + \ln \gamma_{\text{ibu}} = \frac{\Delta H_{\text{fus}}}{R} \left[ \frac{1}{T_m} - \frac{1}{T} \right] \quad (14)$$

where  $\gamma_{\text{ibu}}$  is the liquid phase activity coefficient of ibuprofen at temperature  $T$ , composition  $x$  and pressure  $P$ ,  $R$  is the gas constant,  $\Delta H_{\text{fus}}$  is the enthalpy of fusion of ibuprofen at temperature  $T_m$ , and  $T_m$  and  $T$  are the normal melting point of ibuprofen and the mixture temperature, respectively. The pressure is assumed to be atmospheric ( $P = 1$  atm). The activity coeffi-

**Table 2. Index and Sets for the Case Study**

Description	Index	Set	Value Range
Pure candidate solvents	$s, s'$	$S$	1, ..., 9
Components in mixture	$i, j$	$I$	ibu, $c_1, c_2, c_3$
Functional groups	$k, m$	$K$	1, ..., 14
$N_{\text{selected}}^{\circ}$ of solvent molecules selected	$n$	$N$	1, 2, 3

cient is evaluated using the UNIFAC<sup>70,72</sup> group contribution method, and it is calculated as the sum of two contributions, a combinatorial term (superscript  $C$ ) and a residual term (superscript  $R$ ), as shown below

$$\ln \gamma_{\text{ibu}} = \ln \gamma_{\text{ibu}}^C + \ln \gamma_{\text{ibu}}^R. \quad (15)$$

The UNIFAC model proposed by Smith et al.<sup>73</sup> in a form convenient for implementation is used in this design problem and the relevant equations are presented in Appendix A for completeness.

The mutual miscibility of the solvent molecules also needs to be examined to ensure that the final mixture is in one phase. However, algebraic relations to describe this constraint are not available for multicomponent systems and, therefore, in common with other works,<sup>10,11,40</sup> a miscibility constraint for every binary pair of solvent molecules is used in this case study (i.e., each binary pair of solvents must be miscible for the chosen relative composition, temperature, and pressure)<sup>73</sup>

$$\left[ \frac{\partial \ln \gamma_i^{i,j}}{\partial x_i^{i,j}} \right]_{T,P} + \frac{1}{x_i^{i,j}} \geq 0, \quad i < j$$

$$i=N'+1, \dots, N_c - 1$$

$$j=i+1, \dots, N_c$$

(16)

where  $\gamma_i^{i,j}$  denotes the activity coefficient of component  $i$  in a binary mixture of  $i$  and  $j$  at temperature  $T$  and pressure  $P$ , and  $x_i^{i,j}$  denotes the mole fraction of component  $i$  in the mixture of  $i$  and  $j$ . It is related to the mole fraction in the multicomponent mixture by  $x_i^{i,j} = \frac{x_i}{x_i + x_j}$ .

### Scenarios considered

We consider several instances of the case study, with varying complexity. In particular, numerical difficulties may arise due to the highly nonlinear nature of the miscibility function. To reduce the complexity of the model, two main scenarios are considered: in the first, simpler, scenario the miscibility constraint is not taken into account, whereas in the second scenario it is included in the model. For both scenarios, the two formulations proposed in the section on the GDP formulation of the CAM<sup>b</sup>D problem are applied: the restricted problem, where the number of components to be selected is fixed and the general problem, where the number of components in the mixture is unknown. With the restricted problem formulation, mixtures with one, two or three solvents are designed (i.e.,  $N_c = 2, 3$ , or 4). With the general problem formulation, a mixture with at most three solvents is identified (i.e.,  $N_{\max} = 3$ ). The solvents are to be selected from a list of nine candidate compounds.

All the design sets used in this case study are shown in Tables 2 and 3, and the list of candidate solvents is shown in Table 4. Although a list of promising pure solvents in which ibuprofen has a high solubility has sometimes been used in previous work,<sup>10,11</sup> a list of common solvents that yield a range of solubilities is used in this work to investigate mixtures where one compound is a poor performer when used on



**Table 3. Parameters**

	$N'$	$N$	$N_{\max}$	$N_c$
Restricted problem	1	1, 2 or 3	–	$1+N$
General problem	1	–	3	$1+N_{\max}$

its own, but may lead to high solubility values in a mixture. The solvent molecules in the list were chosen based on their low toxicity levels and on their liquid range, ensuring that they are liquid at the chosen conditions of  $T = 300$  K and  $P = 1$  atm. A table with experimental data for toxicity, and boiling and melting points is presented in Appendix B. The candidate solvents are often used in industrial applications and are also reported in studies of the solubility of various solid organic compounds in pure solvents.<sup>74</sup> The problem-specific parameters required are the enthalpy of fusion of ibuprofen (25.5 kJ/mol<sup>75</sup>) and its normal melting point (347.15 K<sup>75</sup>). The solvent molecules and ibuprofen are built from the set of functional groups given in Table 5. The number of groups of type  $k$  in ibuprofen ( $v_{\text{ibu},k}$ ) and in a solvent  $s$  ( $v_{s,k}$ ) is presented in Appendix C in Tables C1 and C2, respectively. The group volume parameters ( $R_k$ ), the group surface area parameters ( $Q_k$ ) and the group interaction parameters ( $a_{k,m}$ ) used in the UNI-FAC model for the prediction of the activity coefficient are obtained from Poling et al.<sup>76</sup> and listed in Appendix C (Tables C3–C5) for completeness.

### Task 1: Mixture design without miscibility constraints

The formulations of the restricted and general problems, without the miscibility constraint, are presented in this section. The objective here is to maximize the solubility of ibuprofen and the objective function, therefore, consists of the mole fraction of ibuprofen in the mixture (i.e.,  $f(x) = x_{\text{ibu}}$ ). General constraints refer to equations that do not depend on the logic choices and they include the solubility (Eq. 14) and activity coefficient (Eq. 15) relations that are expressed in terms of ibuprofen, as well as some equations of the UNIFAC model (Eqs. A2–A4, A7–A9, A12, and A13) from Appendix A. Conditional constraints, such as the identities and compositions of the selected solvent molecules that depend on the assignment and/or number of solvents in the mixture, are included in the appropriate disjunctions, as shown in the next paragraphs.

**Restricted Problem: Fixed Number of Solvents.** This problem aims to identify the optimal mixture of components for a fixed number of solvent molecules (i.e., 1, 2, or 3 selected solvents), along with the mixture composition, to maximize the solubility of ibuprofen. The formulation is presented for the selection of three solvents but it can readily be extended to any fixed number of solvents. The disjunctions for the choice of solvents are shown below

**Table 4. List of Candidate Solvents**

$s$	Compound
1	Acetone
2	Chloroform
3	Ethanol
4	Ethyl-acetate
5	Methanol
6	MIBK
7	2-Propanol
8	Toluene
9	Water

**Table 5. List of Functional Groups**

$k$	Groups	$k$	Groups
1	CH <sub>3</sub>	8	CH <sub>3</sub> COO
2	CH <sub>2</sub>	9	CH <sub>3</sub> CO
3	CH	10	CH <sub>3</sub> OH
4	aCH	11	CHCl <sub>3</sub>
5	aCCH <sub>3</sub>	12	COOH
6	aCCH <sub>2</sub>	13	OH
7	aCCH	14	H <sub>2</sub> O

$$\bigvee_{s \in S} \begin{bmatrix} Y_{i,s} \\ n_{i,k} = v_{s,k}, \quad k \in K \\ q_i = q_s \\ r_i = r_s \end{bmatrix}, \quad i = c_1, c_2, c_3 \quad (\text{R-D1})$$

where  $v_{s,k}$  defines the identity of the solvents, that is,  $v_{s,k}$  represents the number of groups of type  $k$  in solvent  $s$ ;  $n_{i,k}$  gives the identity of designed component  $i$  in the mixture, that is,  $n_{i,k}$  represents the number of groups of type  $k$  in designed component  $i$ . The variables  $r_i$  and  $q_i$  are the molecular van der Waals volume and molecular surface area of component  $i$ , respectively, while  $r_s$  and  $q_s$  are the van der Waals volume and surface area, respectively, for compound  $s$  in the solvent list. In the above disjunctions, only one of the Boolean variables  $Y_{i,s}$  can be selected, which ensures the selection of exactly three solvents. To express disjunction (R-D1) algebraically, the identity of component  $i$ ,  $n_{i,k}$ , is defined by the use of the binary variable  $y_{i,s}$ , multiplied by the identities of all available candidates,  $v_{s,k}$ . Similarly,  $q_i$  and  $r_i$  are expressed as products of  $y_{i,s}$  with  $q_s$  and  $r_s$ , respectively, as shown below

$$\begin{aligned} n_{i,k} &= \sum_{s \in S} v_{s,k} y_{i,s}, \quad i = c_1, c_2, c_3; \quad k \in K \\ q_i &= \sum_{s \in S} q_s y_{i,s}, \quad i = c_1, c_2, c_3 \\ r_i &= \sum_{s \in S} r_s y_{i,s}, \quad i = c_1, c_2, c_3 \end{aligned}$$

These relations force the above variables ( $n_{i,k}$ ,  $q_i$ ,  $r_i$ ) to become zero when a solvent is not chosen, avoiding the use of the big-M parameter in the MINLP reformulation for this simple case.

The logic conditions discussed in the section on the GDP formulation of the CAM<sup>b</sup>D problem (Eqs. 4 and 6) are derived to avoid selecting a given candidate solvent more than once, and to avoid degeneracy by ordering the solvents. After removing the disjunctions (R-D1) and including all the logic conditions expressed as algebraic equations, the resulting MINLP reformulation is given by model (R-BM1) in Appendix D.

**General Problem: Unknown number of solvents.** The formulation is extended to the general case where the number of mixture constituents is not known in advance. Hence, the design variables include the number of components in the mixture, their identities and compositions. As the number of solvent components present in the mixture is allowed to vary from one to three solvents, the maximum number of designed components in the mixture is  $N_{\max} = 3$ .

The problem includes disjunctions for the assignment of each candidate solvent and disjunctions for the number of

**Table 6. Mixture Design Problem Results Obtained from DICOPT Without Miscibility Constraint. The Maximum Solubility Achieved is Shown in Bold Font and Corresponds to a Solvent Mixture with 2 Components**

Case	$x_{ibu}$	$c_1$	$x_{c_1}$	$c_2$	$x_{c_2}$	$c_3$	$x_{c_3}$	CPU(s)
$N = 1$	0.31833	Chloroform	0.68167					0.35
$N = 2$	<b>0.34928</b>	Chloroform	0.49706	Water	0.15366			5.02
$N = 3$	0.34915	Chloroform	0.49691	Water	0.15294	Methanol	0.00100	114.49
$N \leq 3$	<b>0.34928</b>	Chloroform	0.49706	Water	0.15366			316.92

selected solvent. The former disjunctions are the same as those presented in the restricted problem (disjunctions [R-D1]). In the general problem exactly one of the Boolean variables for assigning one solvent to the first designed component in the mixture is active. The disjunctions for the number of components include constraints that depend on the number of solvents selected, as shown below

$$\bigvee_{n=1,2,3} \left[ \begin{array}{l} \tilde{Y}_n \\ \tilde{n}_{i,k} = n_{i,k}, \quad i \in I_i; k \in K \\ \tilde{q}_i = q_i, \quad i \in I_i \\ \tilde{r}_i = r_i, \quad i \in I_i \\ x_i \geq 0.001, \quad i = c_2, c_3 \\ x_i = 0, \quad i \notin (I_i \cup \{ibu\}) \end{array} \right] \quad (\text{G-D1})$$

where  $I_i = \{c_1\}$   
 $I_i = \{c_1, c_2\}$   
 $I_i = \{c_1, c_2, c_3\}$

As mentioned in the section on the GDP formulation of the CAM<sup>b</sup>D problem, at least one designed component should be present in the mixture and, therefore, the mole fraction of the first solvent component ( $x_{c_1}$ ) is always strictly greater than zero. In the general, problem only one disjunction for the number of solvents is selected. The GDP formulation is converted into an MINLP model (G-BM1) via the big-M approach as presented in Appendix D. Variables  $\tilde{n}_{i,k}$ ,  $\tilde{q}_i$ , and  $\tilde{r}_i$  in the above disjunctions are linked to disjunctions (R-D1) via  $n_{i,k}$ ,  $q_i$ , and  $r_i$ , which become zero when a solvent is not selected. It is thus sufficient, for the reformulation of the problem, to derive big-M equations only for the mole fractions ( $x_i$ ) of designed components in the mixture, based on the binary variables for the number of components. A lower bound equal to 0.001 is used to express the mole fraction via big-M approach. It should be noted that Eqs. A5, A6, and A11 of the UNIFAC model (Appendix A), also depend on the number,  $N$ , of solvent components in the mixture. These functions, however, can be placed outside of the disjunctions as the dependence on  $N$  can be captured via  $x_i$ ,  $q_i$ , and  $r_i$ , and their formulation does not lead to numerical difficulties provided that at least one solvent component is present in the final mixture and has a non-zero mole fraction. This is achieved by setting a lower bound for  $x_{c_1}$  (the first designed component) of 0.001. We also set  $x_i^L = 0.001$  for  $i = c_2, c_3$ . Following the formulation steps outlined in the section on the GDP formulation of the CAM<sup>b</sup>D problem,

logic conditions that establish the relationships between the disjunctive sets (Eqs. 4, 6, 12, and  $\Omega'(Y)$  in Table 1) are also derived. Valid upper bounds can readily be derived for the big-M parameters ( $M_{his}$  and  $M_{F_n}$ ) using Eq. 2. The upper bounds used in this case study are relaxed bounds rather than exact bounds, to avoid numerical difficulties arising from tight bounds and machine precision.

### Task 2: Mixture design with the miscibility constraint

In this task, the introduction of the highly nonlinear and nonconvex miscibility function increases the complexity of the formulations and makes their solution quite challenging.

*Restricted Problem: Fixed Number of Solvents.* The restricted problem is formulated in the same fashion as for the first task and it consists of the objective function ( $x_{ibu}$ ), solubility (Eq. 14), activity coefficient (Eq. 15) and miscibility (Eq. 16) constraints, and the logic relations Eqs. 4 and 6. The miscibility function is calculated for every binary pair of designed components, that is, for the pairs ( $c_1, c_2$ ), ( $c_1, c_3$ ), and ( $c_2, c_3$ ) for  $N = 3$ , using the composition of the binary mixture derived from the overall mixture composition. Specifically, the miscibility constraint for a mixture of solvents  $i$  and  $j$  is given as

$$d\gamma_i^{i,j} + \frac{1}{x_i^{i,j}} \geq 0, \quad i < j; \quad i = c_1, c_2; \quad j = c_2, c_3 \quad (17)$$

where  $d\gamma_i^{i,j}$  is the derivative of the natural logarithm of the activity coefficient of component  $i$  with respect to the mole fraction of  $i$  in the binary mixture (i.e.,  $d\gamma_i^{i,j} = \frac{\partial \ln \gamma_i^{i,j}}{\partial x_i^{i,j}}$ ) and it is calculated from the UNIFAC model (i.e.,  $d\gamma_i^{i,j} = (d\gamma_i^{i,j})^C + (d\gamma_i^{i,j})^R$ ). The mole fraction of the solvent  $i$  in the binary mixture ( $x_i^{i,j}$ ) is expressed as

$$x_i^{i,j} = \frac{x_i}{x_i + x_j}, \quad i = c_1, c_2; \quad j = c_2, c_3 \quad (18)$$

where  $x_i$  and  $x_j$  are the mole fractions of components  $i$  and  $j$ , respectively, in the overall mixture. Because the total number of components in the mixture is fixed, Eqs. 17 and 18 can be treated as general equations and included outside the disjunctions. The resulting MINLP formulation is given by model (R-BM2) in Appendix D.

*General Problem: Unknown Number of Solvents.* The more general problem formulation requires further adaptation to include the miscibility constraints. Recalling that the problem includes disjunctions for the assignment of the candidate solvents and disjunctions for the number of the solvents selected,

**Table 7. Mixture Design Problem Results Obtained from DICOPT with Miscibility Constraint. The Maximum Solubility Achieved is Shown in Bold Font and Corresponds to a Solvent Mixture with 2 Components**

Case	$x_{ibu}$	$c_1$	$x_{c_1}$	$c_2$	$x_{c_2}$	$c_3$	$x_{c_3}$	CPU(s)
$N = 2$	<b>0.33383</b>	Chloroform	0.52292	Methanol	0.14325			4.87
$N = 3$	0.33375	Chloroform	0.52263	Methanol	0.14262	Ethanol	0.00100	107.90
$N \leq 3$	<b>0.33383</b>	Chloroform	0.52292	Methanol	0.14325			369.06

**Table 8. Mixture Design Problem Results Obtained from BARON Without Miscibility Constraint. The Maximum Solubility Achieved is Shown in Bold Font and Corresponds to a Solvent Mixture with 2 Components**

Case	$x_{ibu}$	$c_1$	$x_{c_1}$	$c_2$	$x_{c_2}$	$c_3$	$x_{c_3}$	CPU(s)
$N = 1$	0.31833	Chloroform	0.68167					8.14
$N = 2$	<b>0.34929</b>	Chloroform	0.49662	Water	0.15409			210.25
$N = 3$	0.34915	Chloroform	0.49691	Water	0.15294	Methanol	0.00100	4527.36
$N \leq 3$	<b>0.34928</b>	Chloroform	0.49793	Water	0.15279			5715.37

we first note that the disjunctions for assigning solvents from the list to mixture components are unchanged from task 1.

Conversely, the disjunctions for the number of the solvents selected now include variables and functions that depend on the number of the components in the mixture, such as the composition, the miscibility function and the UNIFAC model equations for evaluating the solvent properties. These functions and the relevant variables are placed in the appropriate disjunctions and are presented as formulation (G-D2) in Appendix D. The logic propositions used in the general problem of task 1 are also included in this task. The disjunctions (G-D2) are reformulated via the big-M approach and the resulting MINLP problem is also given as model (G-BM2) in Appendix D.

## Results and Discussion

All models were implemented and solved in GAMS<sup>77</sup> version 24.2.3, using DICOPT,<sup>78–80</sup> which is a local MINLP solver. The models were run on a single core of a dual 6 core Intel Xeon X5675 machine at 3.07 GHz with 48 GB of memory. Due to the highly nonlinear nature of the equations in the models, multiple initial guesses were used to identify good solutions. The best solutions obtained in the first task (i.e., without the miscibility constraint) and in the second task (i.e., with the miscibility constraint) are summarized in Tables 6 and 7, respectively. It can be observed that in the restricted problem for both tasks, the best solution is yielded by a mixture of two solvents. In the formulations where the miscibility function was not included, a mixture of chloroform and water was identified as the optimal solvent mixture, whereas a binary mixture of chloroform and methanol was identified as optimal when the miscibility constraint was added. Indeed, the miscibility constraint is not satisfied for the pair of chloroform and water at the optimal composition of task 1. Chloroform and methanol, conversely, are fully miscible. The mixtures with three components give slightly lower solubility than the mixtures with two components. In these cases, the mole fraction of the third solvent component is at the lower bound. The results obtained when solving the general problem, with  $N$  unknown, validate the solutions obtained in solving the three restricted problems, by confirming that the highest solubility is achieved by a binary mixture, in which the composition of ibuprofen is 0.34928 and 0.33383 in the first and second task, respectively. The results of both tasks show that a higher solubility can be obtained in a mixture of two or three components rather than in a pure solvent.

The problems of the first task, where the miscibility constraint was not included in the formulation, were also solved globally in GAMS, using BARON<sup>81</sup> which is a global MINLP solver. The results validate the solutions obtained with DICOPT, by proving that the highest solubility is achieved by the mixture of chloroform and water, as shown in Table 8. The slightly different values of mole fraction observed between DICOPT and BARON are due to the convergence criteria used in each solver. The problems of the second task, where the miscibility of the solvents was taken into account, are more complex and could not be solved to global optimality.

It is instructive to consider the computational requirements of the different problem formulations. The values of the CPU time presented in Tables 6–8 correspond to the runs where the best solutions were found and they are representative of all attempts with different initial points. For the problems with a fixed number of mixture components, the CPU time increases rapidly with the number of components, as a result of the increase in the size and complexity of the problems. The introduction of the nonlinear miscibility constraints has little effect on the cost of solving the restricted and general problems locally. Furthermore, enumerating all options for the number of components ( $N = 1, 2, \text{ or } 3$ ) appears to be a more effective strategy than the solution of the general problem from the perspective of computational cost. Further, computational studies are required to investigate whether this finding holds for larger number of components.

## Conclusions

Computer-aided mixture design is an important tool that has the potential to improve process and product design, but that often leads to challenging mixed integer optimization problems due to nonconvexities in the space of the continuous variables and a large combinatorial solution space. The number of interlinked decisions to be considered makes it difficult to formulate the problem in a way which can be easily understood, modified and solved. A general modeling framework for mixture design problems has been proposed in this work to address these difficulties. Several problem formulations based on the GDP formalism have been presented. They provide a systematic approach to posing CAM<sup>b</sup>D problems in which the number of mixture components, the identities of the components and their compositions are to be determined. The proposed approach has been applied successfully to a solvent mixture design problem for maximizing the solubility of ibuprofen. The methodology adopted in this case study included two problem formulations: (1) with fixed number of solvents (restricted problem) and (2) with unknown number of components (general problem). Both problems were first solved without taking any miscibility constraint into account in the problem formulation and then including a miscibility constraint for every binary solvent pairs. Logic conditions between the disjunctive sets were expressed as algebraic constraints, whereas disjunctions for the assignment and number of solvent molecules were transformed into mixed-integer constraints using the big-M approach. High quality solutions of all problems were obtained using a local MINLP solver.

The findings from the case study provide evidence of the usefulness and versatility of a GDP-based approach to optimal mixture design. Integrating GDP techniques into the CAM<sup>b</sup>D framework can facilitate the formulation of the design problem, making it possible to optimize simultaneously the number, identities, and compositions of components in the mixture. Numerical difficulties associated with the absence of components in the final mixture, which are a concern when miscibility constraints are included in the formulation, can be avoided,

leading to computationally efficient solutions. In the case study, it was found that mixtures outperform pure solvents.

Future perspectives for this work include developing algorithms to solve these complex design problems globally and using alternative logic-based optimization techniques. The Big-M formulation used in this study is the most common relaxation technique but it is known to give weak lower bounds for a minimization problem.<sup>43,82</sup> Other techniques, such as Hull Relaxation, can be used. In the case of convex problems, the resulting bounds are at least as tight or tighter<sup>52,65</sup> but the case of nonconvex problems<sup>83,84</sup> presents additional challenges. Finally, the formulation of the design problems could be extended to the design of molecules from the basic building blocks (UNIFAC groups) so that the preselection of promising molecules to include in the list of candidates can be avoided. In this way, a comprehensive approach to mixture design problems can be adopted, where the optimal number of molecules, their identities and compositions are optimized simultaneously.

## Acknowledgments

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).

## Data Statement

Data underlying this article can be accessed on Zenodo at <https://zenodo.org/record/35234>, and used under the Creative Commons Attribution license.

## Literature Cited

1. Duvedi A, Achenie LEK. On the design of environmentally benign refrigerant mixtures: a mathematical programming approach. *Comput Chem Eng.* 1997;21:915–923.
2. Churi N, Achenie LEK. The optimal design of refrigerant mixtures for a two-evaporator refrigeration system. *Comput Chem Eng.* 1997; 21:S349–S354.
3. Vaidyaraman S, Maranas CD. Synthesis of mixed refrigerant cascade cycles. *Chem Eng Commun.* 2002;189:1057–1078.
4. Vaidyanathan R, El-Halwagi M. Computer-aided synthesis of polymers and blends with target properties. *Ind Eng Chem Res.* 1996;35:627–634.
5. Solvason CC, Chemmangattuvalappil NG, Eljack FT, Eden MR. Efficient visual mixture design of experiments using property clustering techniques. *Ind Eng Chem Res.* 2009;48:2245–2256.
6. Adjiman CS. Optimal solvent design approaches. In: Floudas CA, Pardalos PM, editors. *Encyclopedia of Optimization*. US: Springer, 2009:2750–2757.
7. Brignole EA, Gani R. Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilib.* 1983;13:331–340.
8. Brignole EA, Bottini S, Gani R. A strategy for the design and selection of solvents for separation processes. *Fluid Phase Equilib.* 1986; 29:125–132.
9. Pistikopoulos EN, Stefanis SK. Optimal solvent design for environmental impact minimization. *Comput Chem Eng.* 1998;22:717–733.
10. Karunanithi AT, Achenie LEK, Gani R. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Ind Eng Chem Res.* 2005;44:4785–4797.
11. Karunanithi AT, Achenie LEK, Gani R. A computer-aided molecular design framework for crystallization solvent design. *Chem Eng Sci.* 2006;61:1247–1260.
12. Folić M, Adjiman CS, Pistikopoulos EN. Design of solvents for optimal reaction rate constants. *AIChE J.* 2007;53:1240–1256.
13. Adjiman CS, Clarke AJ, Cooper G, Taylor PC. Solvents for ring-closing metathesis reactions. *Chem Commun.* 2008;2806–2808.
14. Gani R. Chemical product design: challenges and opportunities. *Comput Chem Eng.* 2004;28:2441–2457.
15. Gani R. Computer-aided methods and tools for chemical product design. *Chem Eng Res Des.* 2004;82:1494–1504.

16. Registration, evaluation, authorisation & restriction of chemicals (REACH). 2014. Available online: <http://www.hse.gov.uk/reach/legislation.htm> [Accessed January, 2015].
17. Granberg RA, Rasmuson AC. Solubility of paracetamol in binary and ternary mixtures of water + acetone + toluene. *J Chem Eng Data.* 2000;45:478–483.
18. Pacheco DP, Manrique YJ, Martínez F. Thermodynamic study of the solubility of ibuprofen and naproxen in some ethanol+propylene glycol mixtures. *Fluid Phase Equilib.* 2007;262:23–31.
19. Meindersma GW, Podt AJG, de Haan AB. Ternary liquid-liquid equilibria for mixtures of toluene + n-heptane + an ionic liquid. *Fluid Phase Equilib.* 2006;247:158–168.
20. Nagata I, Ohtsubo K. Thermodynamics of alcohol solutions: phase equilibria of binary and ternary mixtures containing two alcohols. *Thermochim Acta.* 1986;102:185–205.
21. Gani R, Jiménez-González C, ten Kate A, Crafts PA, Jones M, Powell L, Atherton JH, Cordiner JL. A modern approach to solvent selection. *Chem Eng.* 2006;113:30–43.
22. CAPEC database. Computer aided process engineering center. Department of chemical and biochemical engineering, Technical University of Denmark. Available online: <http://www.capec.kt.dtu.dk/Software> [Accessed March, 2015]
23. CHEMSAFE numerical database. Gesellschaft für chemische technik und biotechnologie e.V. 1989. Available online: <http://www.dechema.de/en/chemsafe.html> [Accessed March, 2015]
24. DECHEMA chemistry data series (CDS). Gesellschaft für chemische technik und biotechnologie e.V. 1979. Available online: <http://www.dechema.com/en/CDS.html> [Accessed March, 2015]
25. DETHERM database. Thermophysical properties of pure substances & mixtures. Gesellschaft für chemische technik und biotechnologie e.V. Available online: <http://www.dechema.de/en/detherm.html> [Accessed March, 2015]
26. DIPPR - Thermophysical data. Design institute for physical properties. American institute of chemical engineers. Available online: <http://www.aiche.org/dippr> [Accessed March, 2015]
27. HSSDS - Hazardous solvent substitution data system. US environmental protection agency. Available online: <http://es.epa.gov/ssds/ssds.html> [Accessed March, 2015]
28. NIST chemistry webbook. National institute of standards and technology. Available online: <http://webbook.nist.gov/> [Accessed March, 2015]
29. SOLV-DB solvents database. National center for manufacturing sciences. Available online: <http://solvdb.ncms.org/index.html> [Accessed March, 2015]
30. SSDS - Solvent substitution data systems. US environmental protection agency. Available online: <http://es.epa.gov/ssds/ssds.html> [Accessed March, 2015]
31. TAPP database for thermochemical and physical properties. Available online: <https://www.chempute.com/tapp.htm> [Accessed March, 2015]
32. Akula PT, Kleniati PM, Adjiman CS. On the design of optimal solvent mixtures using generalised disjunctive programming. In: *22nd European Symposium on Computer Aided Process Engineering*. 2012. Available online: <http://booksite.elsevier.com/9780444594310/downloads/ESC.311-Onthedesigndofoptimalsolventmixturesusinggeneraliseddisjunctiveprogramming.pdf>.
33. Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward optimum working fluid mixtures for organic Rankine cycles using molecular design and sensitivity analysis. *Ind Eng Chem Res.* 2013;52:12116–12133.
34. Yunus NA, Gernaey KV, Woodley JM, Gani R. A systematic methodology for design of tailor-made blended products. *Comput Chem Eng.* 2014;66:201–213.
35. Achenie LEK, Gani R, Venkatasubramanian V. *Computer Aided Molecular Design: Theory and Practice*. Amsterdam, The Netherlands: Elsevier, 2003.
36. Harper PM, Hostrup M, Gani R. A hybrid CAMD method. In: Achenie LEK, Gani R, Venkatasubramanian V, editors. *Computer Aided Molecular Design: Theory and Practice*. Amsterdam, The Netherlands: Elsevier, 2003:129–165.
37. Sioukrou E, Galindo A, Adjiman CS. On the optimal design of gas-expanded liquids based on process performance. *Chem Eng Sci.* 2014;115:19–30.
38. Sinha M, Achenie LEK, Gani R. Blanket wash solvent blend design using interval analysis. *Ind Eng Chem Res.* 2003;42:516–527.
39. Achenie LEK, Sinha M. Interval global optimization in solvent design. *Reliab Comput.* 2003;9:317–338.
40. Buxton A, Livingston AG, Pistikopoulos EN. Optimal design of solvent blends for environmental impact minimization. *AIChE J.* 1999; 45:817–843.

41. Klein JA, Wu DT, Gani R. Computer aided mixture design with specified property constraints. *Comput Chem Eng.* 1992;16:S229–S236.
42. Ng LY, Andiappan V, Chemmangattuvalappil NG, Ng DKS. A systematic methodology for optimal mixture design in an integrated biorefinery. *Comput Chem Eng.* 2015;81:288–309.
43. Raman R, Grossmann IE. Modelling and computational techniques for logic based integer programming. *Comput Chem Eng.* 1994;18:563–578.
44. Sinha M, Achenie LEK, Ostrovsky GM. Environmentally benign solvent design by global optimization. *Comput Chem Eng.* 1999;23:1381–1394.
45. Sinha M, Achenie LEK, Ostrovsky GM. Optimization methods in CAMD-I. In: Achenie LEK, Gani R, Venkatasubramanian V, editors. *Computer Aided Molecular Design: Theory and Practice.* Amsterdam, The Netherlands: Elsevier, 2003:43–61.
46. Apostolou A, Adjiman CS. Optimization methods in CAMD-II. In: Achenie LEK, Gani R, Venkatasubramanian V, editors. *Computer Aided Molecular Design: Theory and Practice.* Amsterdam, The Netherlands: Elsevier, 2003:63–93.
47. Maranas CD. Optimal computer-aided molecular design: a polymer design case study. *Ind Eng Chem Res.* 1996;35:3403–3414.
48. Sheldon TJ, Folić M, Adjiman CS. Solvent design using a quantum mechanical continuum solvation model. *Ind Eng Chem Res.* 2006;45:1128–1140.
49. Giovanoglou A, Barlatier J, Adjiman CS, Pistikopoulos EN, Cordiner JL. Optimal solvent design for batch separation based on economic performance. *AIChE J.* 2003;49:3095–3109.
50. Folić M, Adjiman CS, Pistikopoulos EN. Computer-aided solvent design for reactions: maximizing product formation. *Ind Eng Chem Res.* 2008;47:5190–5202.
51. Cheng HC, Wang FS. Computer-aided biocompatible solvent design for an integrated extractive fermentation-separation process. *Chem Eng J.* 2010;162:809–820.
52. Vecchiotti A, Lee S, Grossmann IE. Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations. *Comput Chem Eng.* 2003;27:433–448.
53. Sawaya NW, Grossmann IE. A cutting plane method for solving linear generalized disjunctive programming problems. *Comput Chem Eng.* 2005;29:1891–1913.
54. Castro PM, Grossmann IE. Generalized disjunctive programming as a systematic modeling framework to derive scheduling formulations. *Ind Eng Chem Res.* 2012;51:5781–5792.
55. Balas E. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM J Algebr Discrete Methods.* 1985;6:466–486.
56. Beaumont N. An algorithm for disjunctive programs. *Eur J Oper Res.* 1991;48:362–371.
57. Turkay M, Grossmann IE. Logic-based MINLP algorithms for the optimal synthesis of process networks. *Comput Chem Eng.* 1996;20:959–978.
58. Hooker JN, Osorio MA. Mixed logical-linear programming. *Discrete Appl Math.* 1999;96–97:395–442.
59. Hooker JN. *Logic-Based Methods for Optimization: Combining Optimization and Constraint Satisfaction*, 1st ed. USA: Wiley & Sons, Inc., 2000.
60. Lee S, Grossmann IE. New algorithms for nonlinear generalized disjunctive programming. *Comput Chem Eng.* 2000;24:2125–2141.
61. Ruiz JP, Jagla JH, Grossmann IE, Meeraus A, Vecchiotti A. Generalized disjunctive programming: solution strategies. In: Kallrath J, editor. *Algebraic Modeling Systems.* Germany: Springer Science+Business Media, 2012:82–99.
62. Vecchiotti A, Grossmann IE. LOGMIP: a disjunctive 0-1 non-linear optimizer for process system models. *Comput Chem Eng.* 1999;23:555–565.
63. Vecchiotti A, Grossmann IE. Modeling issues and implementation of language for disjunctive programming. *Comput Chem Eng.* 2000;24:2143–2155.
64. Grossmann IE, Trespacios F. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. *AIChE J.* 2013;59:3276–3295.
65. Lee S, Grossmann IE. Generalized convex disjunctive programming: nonlinear convex hull relaxation. *Comput Optim Appl.* 2003;26:83–100.
66. Nemhauser GL, Wolsey LA. *Integer and Combinatorial Optimization*, vol. 41. New York: Wiley & Sons, Inc., 1999.
67. Williams HP. *Model Building in Mathematical Programming*, 2nd ed. California: Wiley & Sons, 1985.
68. Raman R, Grossmann IE. Relation between MILP modelling and logical inference for chemical process synthesis. *Comput Chem Eng.* 1991;15:73–84.
69. Gordon RE, Amin SI. Crystallization of ibuprofen. US Patent US4476248 A. 1984. Available online: <http://www.google.com/patents/US4476248> [Accessed May, 2015].
70. Gmehling J, Anderson TF, Prausnitz JM. Solid-liquid equilibria using UNIFAC. *Ind Eng Chem Fundam.* 1978;17:269–273.
71. Sandler SI. *Chemical and Engineering Thermodynamics*, 3rd ed. New York, Chichester: John Wiley & Sons, 1999.
72. Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* 1975;21:1086–1099.
73. Smith JM, Van Ness HC, Abbott MM. *Introduction to Chemical Engineering Thermodynamics*, 6th ed. Boston, London: McGraw-Hill, 2001.
74. Gracin S, Brinck T, Rasmuson AC. Prediction of solubility of solid organic compounds in solvents by UNIFAC. *Ind Eng Chem Res.* 2002;41:5114–5124.
75. Gracin S, Rasmuson AC. Solubility of phenylacetic acid, p-hydroxyphenylacetic acid, p-aminophenylacetic acid, p-hydroxybenzoic acid, and ibuprofen in pure solvents. *J Chem Eng Data.* 2002;47:1379–1383.
76. Poling BE, Prausnitz JM, O'Connell JP. *The Properties of Gases and Liquids*, 5th ed. New York, London: McGraw-Hill, 2001.
77. GAMS Development Corporation. General Algebraic Modeling System (GAMS). 2014. Available online: <http://www.gams.com> [Accessed April, 2015]
78. Duran MA, Grossmann IE. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math Program.* 1986;36:307–339.
79. Kocis GR, Grossmann IE. Computational experience with DICOPT solving problems in process systems engineering. *Comput Chem Eng.* 1989;13:307–315.
80. Grossmann IE, Viswanathan J, Vecchiotti A, Raman R, Kalvelagen E. GAMS/DICOPT: a discrete continuous optimization package. 2002. Available online: <http://www.gams.com/dd/docs/solvers/dicopt/index.html> [Accessed April, 2015]
81. Tawarmalani M, Sahinidis NV. A polyhedral branch-and-cut approach to global optimization. *Math Program.* 2005;103:225–249.
82. Grossmann IE. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optim Eng.* 2002;3:227–252.
83. Ruiz JP, Grossmann IE. Strengthening of lower bounds in the global optimization of bilinear and concave generalized disjunctive programs. *Comput Chem Eng.* 2010;34:914–930.
84. Ruiz JP, Grossmann IE. Using convex nonlinear relaxations in the global optimization of nonconvex generalized disjunctive programs. *Comput Chem Eng.* 2013;49:70–84.
85. Martin TM, Young DM. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (*Pimephales promelas*) using a group contribution method. *Chem Res Toxicol.* 2001;14:1378–1385.
86. Sigma-Aldrich. Chemistry-Solvents. 2014. Available online: <https://www.sigmaaldrich.com> [Accessed January, 2015]

## Appendix A UNIFAC Model

These equations are proposed by Smith et al.<sup>73</sup> in a form convenient for programming and they are slightly changed to avoid some numerical difficulties when the activity coefficient of ibuprofen is calculated.

### Activity coefficient

$$\ln \gamma_{ibu} = \ln \gamma_{ibu}^C + \ln \gamma_{ibu}^R \quad (A1)$$

### Combinatorial part of activity coefficient

$$\ln \gamma_{ibu}^C = 1 - J_{ibu} + \ln J_{ibu} - 5q_{ibu} \left( 1 - \frac{J_{ibu}}{L_{ibu}} + \ln \frac{J_{ibu}}{L_{ibu}} \right) \quad (A2)$$

$$r_s = \sum_k v_{s,k} R_k \quad (A3)$$

$$q_s = \sum_k v_{s,k} Q_k \quad (A4)$$

$$J_{ibu} = \frac{r_{ibu}}{N_c} \sum_{i=1}^{N_c} r_i X_i \quad (A5)$$

$$L_{\text{ibu}} = \frac{q_{\text{ibu}}}{\sum_{i=1}^{N_c} q_i x_i} \quad (\text{A6})$$

Residual part of activity coefficient

$$\ln \gamma_{\text{ibu}}^R = q_{\text{ibu}} \left[ 1 - \sum_{k \in K} \left( \theta_k \frac{\beta_{\text{ibu},k}}{\omega_k} - e_{\text{ibu},k} \ln \frac{\beta_{\text{ibu},k}}{\omega_k} \right) \right] \quad (\text{A7})$$

$$e_{\text{ibu},k} = \frac{n_{\text{ibu},k} Q_k}{q_{\text{ibu}}}, \quad k \in K \quad (\text{A8})$$

$$\beta_{\text{ibu},k} = \sum_{m \in K} \frac{n_{\text{ibu},m} Q_m \psi_{m,k}}{q_{\text{ibu}}}, \quad k \in K \quad (\text{A9})$$

$$b_{i,k} = \sum_{m \in K} n_{i,m} Q_m \psi_{m,k}, \quad i \in I; \quad k \in K \quad (\text{A10})$$

$$\theta_k = \frac{\sum_{i=1}^{N_c} x_i n_{i,k} Q_k}{\sum_{j=1}^{N_c} x_j Q_j}, \quad k \in K \quad (\text{A11})$$

$$\omega_k = \sum_{m \in K} \theta_m \psi_{m,k}, \quad k \in K \quad (\text{A12})$$

$$\psi_{m,k} = \exp\left(\frac{-a_{m,k}}{T}\right), \quad m \in K; \quad k \in K \quad (\text{A13})$$

## B Solvent Properties

Experimental data for toxicity, and boiling and melting temperatures of the candidate solvents are presented in Table B1.

**Table B1. Experimental Values for Toxicity,<sup>85</sup> Melting and Boiling Temperatures<sup>86</sup> of the Candidate Solvents**

Solvents	Toxicity ( $-\log LC_{50}$ )	Normal Melting Point (K)	Normal Boiling Point (K)
Acetone	0.85	179.15	329.15
Chloroform	3.06	210.15	334.15
Ethanol	0.52	159.15	351.15
Ethylacetate	2.58	189.15	350.15
Methanol	0.05	175.15	337.85
MIBK	2.27	193.15	390.65
2-Propanol	0.78	183.65	355.15
Toluene	3.42	180.15	383.65
Water	0	273.15	373.15
	largest value = 3.42	Mixture temperature = 300K	

**Table C1.  $v_{\text{ibu},k}$ , Number Each Group  $k$  in Ibuprofen**

	CH <sub>3</sub>	CH	aCH	aCCH <sub>2</sub>	aCCH	COOH
$v_{\text{ibu},k}$	3	1	4	1	1	1

**Table C2.  $v_{s,k}$ , Number Each Group  $k$  in a Solvent  $s$**

$v_{s,k}$	CH <sub>3</sub>	CH <sub>2</sub>	CH	aCH	aCCH <sub>3</sub>	CH <sub>3</sub> OH	OH	CH <sub>3</sub> COO	CH <sub>3</sub> CO	CHCl <sub>3</sub>	H <sub>2</sub> O
Acetone	1								1		
Chloroform										1	
Ethanol	1	1					1				
Ethylacetate	1	1						1			
Methanol						1					
MIBK	2	1	1						1		
2-Propanol	2	1					1				
Toluene				5	1						
Water											1

**Table C3.  $R_k$ , Group Volume Parameters<sup>76</sup>**

Groups	$R_k$
CH <sub>3</sub>	0.9011
CH <sub>2</sub>	0.6744
CH	0.4469
aCH	0.5313
aCCH <sub>3</sub>	1.2663
aCCH <sub>2</sub>	1.0396
aCCH	0.8121
CH <sub>3</sub> OH	1.4311
OH	1.0000
CH <sub>3</sub> COO	1.9031
CH <sub>3</sub> CO	1.6724
COOOH	1.3013
CHCl <sub>3</sub>	2.8700
H <sub>2</sub> O	0.9200

**Table C4.  $Q_k$ , Group Surface Area Parameters<sup>76</sup>**

Groups	$Q_k$
CH <sub>3</sub>	0.848
CH <sub>2</sub>	0.540
CH	0.228
aCH	0.400
aCCH <sub>3</sub>	0.968
aCCH <sub>2</sub>	0.660
aCCH	0.348
CH <sub>3</sub> OH	1.432
OH	1.200
CH <sub>3</sub> COO	1.728
CH <sub>3</sub> CO	1.488
COOOH	1.224
CHCl <sub>3</sub>	2.410
H <sub>2</sub> O	1.400

## C Parameters of the UNIFAC Model used in this Case Study

The number of groups of type  $k$  in ibuprofen ( $v_{\text{ibu},k}$ ) and in a solvent  $s$  ( $v_{s,k}$ ) are presented in Tables C1 and C2, respectively; the group volume parameters ( $R_k$ ), the group surface area parameters ( $Q_k$ ) and the group interaction parameters ( $a_{k,m}$ ) used in the UNIFAC model for the prediction of the activity coefficient are listed in Tables C3–C5, respectively.

## D Problem Formulations

For definition of indices and sets see Table 2.

### MINLP Formulations for Task 1

*Restricted problem ( $N = 3$ )*

$$\max x_{\text{ibu}}$$

Table C5.  $a_{k,m}$ , Group Interaction Parameters<sup>76</sup>

$k/m$	CH <sub>3</sub>	CH <sub>2</sub>	CH	aCH	aCCH <sub>3</sub>	aCCH <sub>2</sub>	aCCH	CH <sub>3</sub> OH	OH	CH <sub>3</sub> COO	COOH	CH <sub>3</sub> CO	CHCl <sub>3</sub>	H <sub>2</sub> O
CH <sub>3</sub>	0	0	0	61.13	76.5	76.5	76.5	986.5	697.2	1318	476.4	232.1	663.5	24.9
CH <sub>2</sub>	0	0	0	61.13	76.5	76.5	76.5	986.5	697.2	1318	476.4	232.1	663.5	24.9
CH	0	0	0	61.13	76.5	76.5	76.5	986.5	697.2	1318	476.4	232.1	663.5	24.9
aCH	-11.12	-11.12	-11.12	0	167	167	167	636.1	637.4	903.8	25.77	5.994	537.4	-231.9
aCCH <sub>3</sub>	-69.7	-69.7	-69.7	-146.8	0	0	0	803.2	603.3	5695	-52.1	5688	872.3	-80.25
aCCH <sub>2</sub>	-69.7	-69.7	-69.7	-146.8	0	0	0	803.2	603.3	5695	-52.1	5688	872.3	-80.25
aCCH	-69.7	-69.7	-69.7	-146.8	0	0	0	803.2	603.3	5695	-52.1	5688	872.3	-80.25
CH <sub>3</sub> OH	16.51	16.51	16.51	-50	-44.5	-44.5	-44.5	249.1	0	-181	23.39	-10.72	-202	-139.4
OH	156.4	156.4	156.4	89.6	25.82	25.82	25.82	0	-137.1	353.5	84	101.1	199	-98.12
CH <sub>3</sub> COO	114.8	114.8	114.8	85.84	-170	-170	-170	245.4	249.6	200.8	372.2	0	660.2	-209.7
COOH	315.3	315.3	315.3	62.32	89.86	89.86	89.86	-151	339.8	-66.17	-297.8	-256.3	0	39.63
CH <sub>3</sub> CO	26.76	26.76	26.76	140.1	365.8	365.8	365.8	164.5	108.7	472.5	0	-213.7	669.4	-354.6
CHCl <sub>3</sub>	36.7	36.7	36.7	288.5	69.9	69.9	69.9	742.1	649.1	826.8	552.1	176.5	504.2	0
H <sub>2</sub> O	300	300	300	362.3	377.6	377.6	377.6	-229.1	289.6	0	-195.4	72.87	-14.09	353.7

subject to

$$\ln x_{ibu} + \ln \gamma_{ibu} = \frac{\Delta H_{fus}}{R} \left[ \frac{1}{T_m} - \frac{1}{T} \right]$$

$$\ln \gamma_{ibu} = \ln \gamma_{ibu}^C + \ln \gamma_{ibu}^R$$

$$n_{i,k} = \sum_{s \in S} v_{s,k} y_{i,s}, \quad i = c_1, c_2, c_3; \quad k \in K$$

$$q_i = \sum_{s \in S} q_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$r_i = \sum_{s \in S} r_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$\sum_{i \in I} x_i = 1$$

select exactly 3 solvent molecules :

$$\sum_{s \in S} y_{i,s} = 1, \quad i = c_1, c_2, c_3$$

logic relations :

select each candidate solvent at most once :

$$\sum_{i \in I} y_{i,s} \leq 1, \quad s \in S; \quad I = \{c_1, c_2, c_3\}$$

solvent ordering :

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

$$\forall s' < s; \quad s = 2, \dots, 9; \quad s' = 1, \dots, s$$

$$0.001 \leq x_i \leq 1, \quad y_{i,s} \in \{0, 1\}, \quad i \in I; \quad s \in S \quad (\text{R-BM1})$$

General problem ( $N_{max}=3$ )

$$\max x_{ibu}$$

subject to

$$\ln x_{ibu} + \ln \gamma_{ibu} = \frac{\Delta H_{fus}}{R} \left[ \frac{1}{T_m} - \frac{1}{T} \right]$$

$$\ln \gamma_{ibu} = \ln \gamma_{ibu}^C + \ln \gamma_{ibu}^R$$

$$n_{i,k} = \sum_{s \in S} v_{s,k} y_{i,s}, \quad i = c_1, c_2, c_3; \quad k \in K$$

$$q_i = \sum_{s \in S} q_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$r_i = \sum_{s \in S} r_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$\sum_{i \in I} x_i = 1$$

select only one disjunction :

$$\sum_{n=1}^3 \tilde{y}_n = 1$$

select exactly one solvent for the first designed component :

$$\sum_{s \in S} y_{c_1,s} = 1$$

logic relations :

select at most 2 more solvents :

$$\sum_{s \in S} y_{i,s} \leq 1, \quad i = c_2, c_3$$

select each candidate solvent at most once :

$$\sum_{i \in I} y_{i,s} \leq 1, \quad s \in S; \quad I = \{c_1, c_2, c_3\}$$

solvent ordering :

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

$$\forall s' < s; \quad s = 2, \dots, 9; \quad s' = 1, \dots, s$$

using  $\tilde{y}_n$  to control  $y_{i,s}$  :

(G-BM1)

$$\tilde{y}_1 + y_{i,s} \leq 1, \quad s \in S, \quad i = c_2, c_3$$

$$\tilde{y}_2 + y_{i,s} \leq 1, \quad s \in S, \quad i = c_3$$

$$\tilde{y}_2 \leq \sum_{s \in S} y_{i,s}, \quad i = c_2$$

$$\tilde{y}_3 \leq \sum_{s \in S} y_{i,s}, \quad i = c_2, c_3$$

reformulation of the disjunctive constraints via big-M :

$$0.001(1 - \tilde{y}_1) \leq x_{c_2} \leq 0.999(1 - \tilde{y}_1)$$

$$0.001\tilde{y}_3 \leq x_{c_3} \leq 0.999\tilde{y}_3$$

$$0.001 \leq x_{ibu}, x_{c_1} \leq 1$$

$$0 \leq x_{c_2}, x_{c_3} \leq 1, \quad y_{i,s}, \tilde{y}_n \in \{0, 1\}, \quad i \in I; \quad s \in S; \quad n \in N$$

## Disjunctions for Number of Components in the Mixture for Task 2

General problem ( $N_{max} = 3$ )

$$\begin{array}{c}
 \left[ \begin{array}{c}
 \tilde{Y}_1 \\
 x_{c_2} = x_{c_3} = 0 \\
 \tilde{n}_{c_1,k} = n_{c_1,k} \\
 \tilde{n}_{c_2,k} = \tilde{n}_{c_3,k} = 0 \\
 \text{for } k \in K \\
 \tilde{q}_{c_1} = q_{c_1} \\
 \tilde{q}_{c_2} = \tilde{q}_{c_3} = 0 \\
 \tilde{r}_{c_1} = r_{c_1} \\
 \tilde{r}_{c_2} = \tilde{r}_{c_3} = 0 \\
 b_{i,k} = 0 \\
 \text{for } i = c_1, c_2, c_3; k \in K
 \end{array} \right] \vee \left[ \begin{array}{c}
 \tilde{Y}_2 \\
 x_{c_2} \geq 0.001 \\
 x_{c_3} = 0 \\
 \tilde{n}_{i,k} = n_{i,k} \\
 \tilde{n}_{c_3,k} = 0 \\
 \text{for } i = c_1, c_2; k \in K \\
 \tilde{q}_i = q_i \\
 \tilde{q}_{c_3} = 0 \\
 \text{for } i = c_1, c_2 \\
 \tilde{r}_i = r_i \\
 \tilde{r}_{c_3} = 0 \\
 \text{for } i = c_1, c_2 \\
 b_{c_1,k} = \sum_m n_{c_1,m} Q_m \psi_{m,k}, \\
 b_{c_2,k} = b_{c_3,k} = 0 \\
 \text{for } k \in K \\
 \text{miscibility :} \\
 d\gamma_{c_1}^{c_1,c_2} + \frac{1}{x_{c_1}^{c_1,c_2}} \geq 0 \\
 d\gamma_{c_1}^{c_1,c_2} = (d\gamma_{c_1}^{c_1,c_2})^C + (d\gamma_{c_1}^{c_1,c_2})^R \\
 (d\gamma_{c_1}^{c_1,c_2})^C = \frac{\partial(\ln \gamma_{c_1}^{c_1,c_2})^C}{\partial x_{c_1}^{c_1,c_2}} \\
 (d\gamma_{c_1}^{c_1,c_2})^R = \frac{\partial(\ln \gamma_{c_1}^{c_1,c_2})^R}{\partial x_{c_1}^{c_1,c_2}} \\
 x_{c_1}^{c_1,c_2} = \frac{x_{c_1}}{x_{c_1} + x_{c_2}}
 \end{array} \right] \vee \left[ \begin{array}{c}
 \tilde{Y}_3 \\
 x_i \geq 0.001 \\
 \text{for } i = c_2, c_3 \\
 \tilde{n}_{i,k} = n_{i,k} \\
 \text{for } i = c_1, c_2, c_3; k \in K \\
 \tilde{q}_i = q_i \\
 \text{for } i = c_1, c_2, c_3 \\
 \tilde{r}_i = r_i \\
 \text{for } i = c_1, c_2, c_3 \\
 b_{i,k} = \sum_m n_{i,m} Q_m \psi_{m,k} \\
 b_{c_3,k} = 0 \\
 \text{for } i = c_1, c_2; k \in K \\
 \text{miscibility for } i = c_1, c_2; \\
 j = c_2, c_3; i < j : \\
 d\gamma_i^{i,j} + \frac{1}{x_i^{i,j}} \geq 0 \\
 d\gamma_i^{i,j} = (d\gamma_i^{i,j})^C + (d\gamma_i^{i,j})^R \\
 (d\gamma_i^{i,j})^C = \frac{\partial(\ln \gamma_i^{i,j})^C}{\partial x_i^{i,j}} \\
 (d\gamma_i^{i,j})^R = \frac{\partial(\ln \gamma_i^{i,j})^R}{\partial x_i^{i,j}} \\
 x_i^{i,j} = \frac{x_i}{x_i + x_j}
 \end{array} \right] \quad (G-D2)
 \end{array}$$

where variable  $b_{i,k}$  is used to evaluate the miscibility constraint for binary pairs of designed components and it is non-zero only when a mixture of 2 or 3 solvents is designed.

### MINLP Formulations for Task 2

Restricted problem ( $N = 3$ )

$$\begin{array}{l}
 \max x_{ibu} \\
 \text{subject to} \\
 \ln x_{ibu} + \ln \gamma_{ibu} = \frac{\Delta H_{fus}}{R} \left[ \frac{1}{T_m} - \frac{1}{T} \right]
 \end{array}$$

$$\begin{array}{l}
 \ln \gamma_{ibu} = \ln \gamma_{ibu}^C + \ln \gamma_{ibu}^R \\
 n_{i,k} = \sum_{s \in S} v_{s,k} y_{i,s}, \quad i = c_1, c_2, c_3; k \in K \\
 q_i = \sum_{s \in S} q_s y_{i,s}, \quad i = c_1, c_2, c_3 \\
 r_i = \sum_{s \in S} r_s y_{i,s}, \quad i = c_1, c_2, c_3 \\
 \sum_{i \in I} x_i = 1
 \end{array}$$



select exactly 3 solvent molecules :

$$\sum_{s \in S} y_{i,s} = 1, \quad i = c_1, c_2, c_3$$

miscibility constraint :

$$d\gamma_i^{i,j} + \frac{1}{x_i^{i,j}} \geq 0, \quad i < j; i = c_1, c_2; j = c_2, c_3$$

$$d\gamma_i^{i,j} = (d\gamma_i^{i,j})^C + (d\gamma_i^{i,j})^R \quad (\text{R-BM2})$$

$$(d\gamma_i^{i,j})^C = \frac{\partial}{\partial x_i^{i,j}} \left[ 1 - J_i + \ln J_i - 5q_i \left( 1 - \frac{J_i}{L_i} + \ln \frac{J_i}{L_i} \right) \right]$$

$$(d\gamma_i^{i,j})^R = \frac{\partial}{\partial x_i^{i,j}} \left[ q_i - \sum_k \left( \theta_k \frac{b_{i,k}}{\omega_k} - q_i e_{k,i} \ln \frac{\beta_{i,k}}{\omega_k} \right) \right]$$

$$x_i^{i,j} = \frac{x_i}{x_i + x_j}, \quad i < j; i = c_1, c_2; j = c_2, c_3$$

**logic relations :**

select each candidate solvent at most once :

$$\sum_{i \in I} y_{i,s} \leq 1, \quad s \in S; I = \{c_1, c_2, c_3\}$$

solvent ordering :

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

$$\forall s' < s; s = 2, \dots, 9; s' = 1, \dots, s$$

$$0.001 \leq x_i \leq 1; y_{i,s} \in \{0, 1\}, \quad i \in I; s \in S$$

**General problem ( $N_{\max} = 3$ )**

$$\max x_{\text{ibu}}$$

subject to

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

$$\ln \gamma_{\text{ibu}} = \ln \gamma_{\text{ibu}}^C + \ln \gamma_{\text{ibu}}^R$$

$$n_{i,k} = \sum_{s \in S} v_{s,k} y_{i,s}, \quad i = c_1, c_2, c_3; k \in K$$

$$q_i = \sum_{s \in S} q_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$r_i = \sum_{s \in S} r_s y_{i,s}, \quad i = c_1, c_2, c_3$$

$$\sum_{i \in I} x_i = 1$$

select only one disjunction :

$$\sum_{n=1}^3 \tilde{y}_n = 1$$

select exactly one solvent for the first designed component :

$$\sum_{s \in S} y_{c_1,s} = 1$$

**logic relations :**

select at most 2 more solvents :

$$\sum_{s \in S} y_{i,s} \leq 1, \quad i = c_2, c_3$$

select each candidate solvent at most once :

$$\sum_{i \in I} y_{i,s} \leq 1, \quad s \in S; I = \{c_1, c_2, c_3\} \quad (\text{G-BM2})$$

solvent ordering :

$$y_{c_1,s} + y_{c_2,s'} \leq 1$$

$$y_{c_1,s} + y_{c_3,s'} \leq 1$$

$$y_{c_2,s} + y_{c_3,s'} \leq 1$$

$$\forall s' < s; s = 2, \dots, 9; s' = 1, \dots, s$$

using  $\tilde{y}_n$  to control  $y_{i,s}$  :

$$\tilde{y}_1 + y_{i,s} \leq 1, \quad s \in S, i = c_2, c_3$$

$$\tilde{y}_2 + y_{i,s} \leq 1, \quad s \in S, i = c_3$$

$$\tilde{y}_2 \leq \sum_{s \in S} y_{i,s}, \quad i = c_2$$

$$\tilde{y}_3 \leq \sum_{s \in S} y_{i,s}, \quad i = c_2, c_3$$

**reformulation of the disjunctive constraints via big-M :**

$$0.001(1 - \tilde{y}_1) \leq x_{c_2} \leq 0.999(1 - \tilde{y}_1)$$

$$0.001\tilde{y}_3 \leq x_{c_3} \leq 0.999\tilde{y}_3$$

$$-100\tilde{y}_1 \leq b_{c_1,k} - \sum_m n_{m,c_1} Q_m \psi_{m,k} \leq 100\tilde{y}_1, \quad k \in K$$

$$-100(1 - \tilde{y}_3) \leq b_{c_2,k} - \sum_m n_{m,c_2} Q_m \psi_{m,k} \leq -100(1 - \tilde{y}_3), \quad k \in K$$

miscibility for the binary pair  $(c_1, c_2)$

$$-d\gamma_{c_1}^{c_1,c_2} - \frac{1}{x_{c_1}^{c_1,c_2}} \leq 100\tilde{y}_1$$

$$-100\tilde{y}_1 \leq d\gamma_{c_1}^{c_1,c_2} - (d\gamma_{c_1}^{c_1,c_2})^C - (d\gamma_{c_1}^{c_1,c_2})^R \leq 100\tilde{y}_1$$

$$-100\tilde{y}_1 \leq (d\gamma_{c_1}^{c_1,c_2})^C - \frac{\partial}{\partial x_{c_1}^{c_1,c_2}} \left[ 1 - J_{c_1} + \ln J_{c_1} - 5q_{c_1} \left( 1 - \frac{J_{c_1}}{L_{c_1}} + \ln \frac{J_{c_1}}{L_{c_1}} \right) \right] \leq 100\tilde{y}_1$$

$$-100\tilde{y}_1 \leq (d\gamma_{c_1}^{c_1,c_2})^R - \frac{\partial}{\partial x_{c_1}^{c_1,c_2}} \left[ q_{c_1} - \sum_k \left( \theta_k \frac{b_{c_1,k}}{\omega_k} - q_{c_1} e_{k,c_1} \ln \frac{\beta_{c_1,k}}{\omega_k} \right) \right] \leq 100\tilde{y}_1$$

$$-10\tilde{y}_1 \leq x_{c_1}^{c_1,c_2} - \frac{x_{c_1}}{x_{c_1} + x_{c_2}} \leq 10\tilde{y}_1$$

miscibility for the binary pair  $(c_1, c_3)$

$$-d\gamma_{c_1}^{c_1,c_3} - \frac{1}{x_{c_1}^{c_1,c_3}} \leq 100(1 - \tilde{y}_3)$$

$$-100(1 - \tilde{y}_3) \leq d\gamma_{c_1}^{c_1,c_3} - (d\gamma_{c_1}^{c_1,c_3})^C - (d\gamma_{c_1}^{c_1,c_3})^R \leq 100(1 - \tilde{y}_3)$$

$$-100(1 - \tilde{y}_3) \leq (d\gamma_{c_1}^{c_1,c_3})^C - \frac{\partial}{\partial x_{c_1}^{c_1,c_3}} \left[ 1 - J_{c_1} + \ln J_{c_1} - 5q_{c_1} \left( 1 - \frac{J_{c_1}}{L_{c_1}} + \ln \frac{J_{c_1}}{L_{c_1}} \right) \right] \leq 100(1 - \tilde{y}_3)$$

$$-100(1 - \tilde{y}_3) \leq (d\gamma_{c_1}^{c_1,c_3})^R - \frac{\partial}{\partial x_{c_1}^{c_1,c_3}} \left[ q_{c_1} - \sum_k \left( \theta_k \frac{b_{c_1,k}}{\omega_k} - q_{c_1} e_{k,c_1} \ln \frac{\beta_{c_1,k}}{\omega_k} \right) \right] \leq 100(1 - \tilde{y}_3)$$

$$-10(1-\tilde{y}_3) \leq x_{c_1}^{c_1, c_3} - \frac{x_{c_1}}{x_{c_1} + x_{c_3}} \leq 10(1-\tilde{y}_3)$$

miscibility for the binary pair  $(c_2, c_3)$

$$-d\gamma_{c_2}^{c_2, c_3} - \frac{1}{x_{c_2}^{c_2, c_3}} \leq 100(1-\tilde{y}_3)$$

$$-100(1-\tilde{y}_3) \leq d\gamma_{c_2}^{c_2, c_3} - (d\gamma_{c_2}^{c_2, c_3})^C - (d\gamma_{c_2}^{c_2, c_3})^R \leq 100(1-\tilde{y}_3)$$

$$-100(1-\tilde{y}_3) \leq (d\gamma_{c_2}^{c_2, c_3})^C$$

$$-\frac{\partial}{\partial x_{c_2}^{c_2, c_3}} \left[ 1 - J_{c_2} + \ln J_{c_2} - 5q_{c_2} \left( 1 - \frac{J_{c_2}}{L_{c_2}} + \ln \frac{J_{c_2}}{L_{c_2}} \right) \right] \leq 100(1-\tilde{y}_3)$$

$$-100(1-\tilde{y}_3) \leq (d\gamma_{c_2}^{c_2, c_3})^R$$

$$-\frac{\partial}{\partial x_{c_2}^{c_2, c_3}} \left[ q_{c_2} - \sum_k \left( \theta_k \frac{b_{c_2, k}}{\omega_k} - q_{c_2} e_{k, c_2} \ln \frac{\beta_{c_2, k}}{\omega_k} \right) \right] \leq 100(1-\tilde{y}_3)$$

$$-10(1-\tilde{y}_3) \leq x_{c_2}^{c_2, c_3} - \frac{x_{c_2}}{x_{c_2} + x_{c_3}} \leq 10(1-\tilde{y}_3)$$

$$0.001 \leq x_{\text{ibu}}, x_{c_1} \leq 1; 0 \leq x_{c_2}, x_{c_3} \leq 1; y_{i, s}, \tilde{y}_n$$

$$\in \{0, 1\}, i \in I; s \in S; n \in N$$

Manuscript received June 30, 2015, and revision received Sep. 17, 2015.