A New Decomposition−Based Computer-Aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · MAY 2005

Impact Factor: 2.59 · DOI: 10.1021/ie049328h

2 AUTHORS:

Arunprakash T. Karunanithi
University of Colorado Denver, United States
28 PUBLICATIONS **266** CITATIONS

SEE PROFILE

Luke Achenie
Virginia Polytechnic Institute and State Univ…
126 PUBLICATIONS **1,292** CITATIONS

SEE PROFILE

A New Decomposition-Based Computer-Aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures

Arunprakash T. Karunanithi and Luke E. K. Achenie*

Department of Chemical Engineering, University of Connecticut, Storrs, Connecticut 06269

Rafiqul Gani

CAPEC, Department of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

This paper presents a novel computer-aided molecular/mixture design (CAMD) methodology for the design of optimal solvents and solvent mixtures. The molecular/mixture design problem is formulated as a mixed integer nonlinear programming (MINLP) model in which a performance objective is to be optimized subject to structural, property, and process constraints. The general molecular/mixture design problem is divided into two parts. For optimal single-compound design, the first part is solved. For mixture design, the single-compound design is first carried out to identify candidates and then the second part is solved to determine the optimal mixture. The decomposition of the CAMD MINLP model into relatively easy to solve subproblems is essentially a partitioning of the constraints from the original set. This approach is illustrated through two case studies. The first case study involves the design of an optimal extractant for the separation of acetic acid from water by liquid−liquid extraction. The results suggest that the new extractant would be able to perform better than the extractant being widely used for this separation. The second case study is an industrial problem involving the optimal formulation for a pharmaceutical compound. The designed formulation is able to improve the water solubility of the compound by more many fold.

1. Introduction

Solvents are used for a variety of purposes in the process industry. Some of the important applications are in (a) separation processes (crystallization, liquid−liquid extraction, and absorption), (b) chemical reactions, and (c) final product formulations. In many cases, solvent mixtures/blends are used instead of single solvents. Solvents are needed not only during the processing of products but also, in some cases, as part of the final product. Typical examples related to processing are addition of solvents (or antisolvents) for recovery of a pharmaceutical or biochemical product and for the direct/indirect influence of reaction chemistry. Solvents used in a formulated product include addition of solvents for controlled release of pesticides or drugs and as a medium (carrier) for the product. Environmental awareness and strict legislation have resulted in the search for less toxic and environmentally benign solvents and solvent formulations that have improved performance characteristics. In many cases, the search for solvents or solvent mixtures is done through experience, database searches, and bench scale synthesis. These methods are usually expensive and time-consuming. A more systematic approach which could help in identifying solvents and mixtures that can improve performance, be economical, be environmentally benign, and improve process efficiency is needed. Computer-aided molecular/mixture design (CAMD) is fast emerging as a powerful and systematic tool for efficient and reliable design of solvents and solvent mixtures.

CAMD is a reverse engineering approach, which aims to generate molecules having specific properties for the performance of desired activities. In this approach, a large number of structural molecules with desired properties are generated from a small set of structural groups. Achenie et al.[1] define computer-aided molecular design as, "Given a set of building blocks and a specified set of target properties, determine the molecule or molecular structure that matches these properties." In this technique, the reverse problem of property estimation is tackled; that is, for a specified set of properties (target properties), chemicals that satisfy the property requirements are determined.[2] To achieve this objective, we first need the solution to the forward problem; the latter consists of models for the estimation of properties given the molecular structure. Significant advances made in group contribution methods, which predict properties from molecular structure, facilitate the solution of the CAMD problem. Group contribution methods have been developed for pure component properties[3−5] as well as for solution properties.[6,7] Other approaches such as the prediction of properties using connectivity indices[8,9] can also be used.

Computer-aided mixture/blend design is defined as, "Given a set of chemicals and a specified set of property constraints, determine the optimal mixture." Here, the set of chemicals can be identified by first solving the single-compound molecular design problem. Prediction of mixture properties require pure component properties as well as mixing rules, with the simplest rule being the linear mixing rule.

* To whom correspondence should be addressed. Tel.: (860) 486-2756. Fax: (860) 486-2959. E-mail: achenie@engr.uconn.edu.

10.1021/ie049328h CCC: $30.25
Published on Web 05/20/2005

CAMD has been used to identify promising candidates in a variety of applications such as polymer design,[10−12] extractants and solvents,[13−21] refrigerants,[22−24] and catalysts.[25] Constantinou et al.[2] have provided application examples of various types of CAMD problems. Many classes of solution approaches have been used to solve the CAMD problems. These approaches can be broadly classified as enumeration and knowledge-based approaches,[26−29] mathematical-programming approaches,[11,15,20−23] stochastic approaches,[10,30] and hybrid approaches.[31] Achenie et al.[1] provide a detailed review of the various solution techniques. The paper is organized as follows. Section 2 describes formulation of the CAMD problem as an MINLP model. Section 3 discusses some of the common approaches for solving an MINLP model. The proposed decomposition-based solution approach is described in section 4 followed by the validation of the solution approach through two analytical problems in section 5. The implementation of the proposed solution approach and the various tools used are described in section 6. Finally, two illustrative case studies are presented in section 7, while section 8 describes other case studies solved using this approach. Section 9 concludes the paper followed by the Appendix.

2. Problem Formulation

The general CAMD problem can be formulated as a mixed integer nonlinear programming (MINLP) problem, where a (process/product) performance index is optimized subject to constraints (molecular structural constraints, molecular property constraints, mixture property constraints, process models, etc.). Structural constraints are molecular generation rules that need to be satisfied to form a structurally feasible molecule from a collection of groups or descriptors. In most of the CAMD problems, the constraints for the combination rules of the groups proposed by Odele and Machieto[18] or Churi and Achenie[23] are used. These constraints include the following: (1) the octet rule, which ensures that the molecule has zero valency; (2) structural feasibility constraints that make sure that two adjacent groups in a molecule are not linked by more than one bond; and (3) constraints on lower and upper limits on the number of groups of a particular type and the total number of groups making up a molecule.

Pure component properties required for CAMD can be evaluated using group contribution techniques, which are based on molecular structure information. New group contribution methods, which consider higher-order groups,[4,5] improve the accuracy of the property prediction. Some examples of pure component properties are the normal melting point, the normal boiling point, and the heat of fusion at 298 K. Certain pure component properties that depend on the molecular structure as well as other parameters (e.g., temperature) are evaluated using correlations (e.g., viscosity). Also, some secondary properties (those that depend on other properties) are calculated from primary properties (those that depend on only the molecular structure), for example, the solubility parameter. The mixture property that is important in many CAMD applications related to liquid solvents is the liquid-phase activity coefficient from which properties such as solubility can be evaluated. While UNIFAC[6] is the most commonly used group contribution method for prediction of activity coefficients, newer methods such as the higher-order UNIFAC[7] provide a large application range. Process models usually consist of mass−energy balance equations, constitutive equations (property models), and constraint equations (such as phase equilibrium relationships). All these process model-related equations can be represented as equality constraints.

Considering the various types of constraint equations, a general CAMD problem (molecular/mixture design) can be formulated as the following MINLP.

$$\min/\max f_{\text{obj}}(\mathbf{X},\mathbf{Y})$$

subject to

$$\text{structural constraints: } g_1(\mathbf{Y}) \le 0$$

$$\text{pure component property constraints: } g_2(\mathbf{Y}) \le 0$$

$$\text{mixture property constraints: } g_3(\mathbf{X},\mathbf{Y}) \le 0$$

$$\text{process model constraints: } g_4(\mathbf{X},\mathbf{Y}) = 0$$

$\mathbf{Y}$ is a vector of binary variables (integers), which are related to the identities of the building blocks (groups, descriptors) and/or molecules. $\mathbf{X}$ is a vector of continuous variables, which are related to the mixture (e.g., compositions) and/or process variables (e.g., flow rates, temperatures, etc.). f_{obj} is the performance objective function, defined in terms of molecule/mixture process (performance) characteristics and/or cost that may be minimized or maximized. g_1 and g_2 are sets of structural constraints (related to the feasibility of molecular structure) and pure component property constraints (related to property−molecular structure relationships), respectively. g_3 and g_4 are mixture property (related to property−mixture relationships) constraints and process model (related to process−molecule/mixture relationships) constraints, respectively. Usually, g_1 and g_2 are linear in their arguments.

3. Solution Approaches

The MINLP model, which is combinatorial in nature, can be solved in many ways, but finding the solution is not trivial, especially when the equality constraints representing the process models are nonlinear or their number is large. Even though MINLP-based solution approaches are routinely used in the synthesis of heat/mass exchange networks, for practical solvent and mixture design problems, solving the MINLP model presents several difficulties. These arise from the need to use highly nonlinear and complex property models for the prediction of the product−process characteristics. The need to relate the product model with the process model can lead to highly nonconvex and even nonsmooth (with respect to continuous variables) problem formulations. There is also the issue of global optima for which a global optimization technique is needed.

The most direct approach for solving the MINLP model is to completely enumerate the MINLP model to give several nonlinear programs (NLP) corresponding to fixed values of the integer variables. Each NLP then needs to be solved to optimality. For all practical purposes, generate and test methods can be classified under enumeration approaches. The basic technique for the molecular synthesis stage in the generate and test methods follows a combinatorial approach. That is enumerating the possible combinations from the building blocks and testing for structural and property

constraints. Achenie et al.[1] (Chapter 6) report developments based on a multilevel approach that avoids complete enumeration.

Gradient-based optimization algorithms for solving MINLP models such as outer approximation algorithms[32] have been used to solve the CAMD problem.[21,22] These algorithms involve solving a finite sequence of NLP and MILP problems. An MILP master problem is solved first, and the integer solution from the MILP is used in the NLP subproblem in which the binary variables are fixed. The MILP and NLP subproblems are alternatively solved. Because of the nonconvexity in the NLP subproblem, there is no guarantee that this solution is globally optimal. A common strategy is to try several initial guesses, to see if a consistent solution is obtained. At any rate, global optimality is not assured.

Stochastic search methods such as simulated annealing[30] and genetic algorithms[10] have been used to solve the CAMD problem. These methods often do find, although not provable, near-global optimal solutions. Genetic algorithms perform a guided stochastic search where improved solutions are achieved by sampling areas of the search space that have a higher probability for good solutions.[1] The simulated annealing approach is based on randomized evolution of states through stepwise modification.[30]

There have been some instances where global optimization algorithms based on interval analysis[11,33] have been employed to solve relatively small size CAMD problems. Interval analysis-based optimization continuously deletes portions of the search space until one obtains final boxes of specified width that contain global solutions. Interval-based global optimization is in general computationally intensive, and for large CAMD problems, this could be a big disadvantage. Achenie and Sinha[33] have developed a global optimization algorithm for solving CAMD problems that exploits the problem structure. Although in recent years there has been a significant improvement in global optimization algorithms, not many of them have been used for solving CAMD problems. This is primarily because most of these methods can be used only if the problem is small or medium in size. Another approach is to identify sources of nonconvexity and convexify[34] the objective function and constraints for faster convergence and to guarantee globally optimal solution. Often nonlinearities in the continuous variables of the property and process models do not readily lend themselves to convexification.

Problem Decomposition. In most CAMD problems, due to the large number of constraints involved, the feasible region can be very small compared to the search space. All of the feasible solutions to the problem may lie in a relatively small portion of the search space. The ability to solve such problems lies in quickly identifying and avoiding the infeasible portion of the search space. One way to do this is by first solving subproblems consisting of a set of property constraints only.

The process model constraints usually involve equilibrium relationships, which are modeled using highly nonlinear methods for predicting activity coefficients (e.g., UNIFAC). The complexities in the property models lead to nonlinear behavior of the process model equations causing difficulties in convergence and computational efficiency. A way to avoid some of these problems is to decompose the problem into subproblems, which are relatively easy to solve.

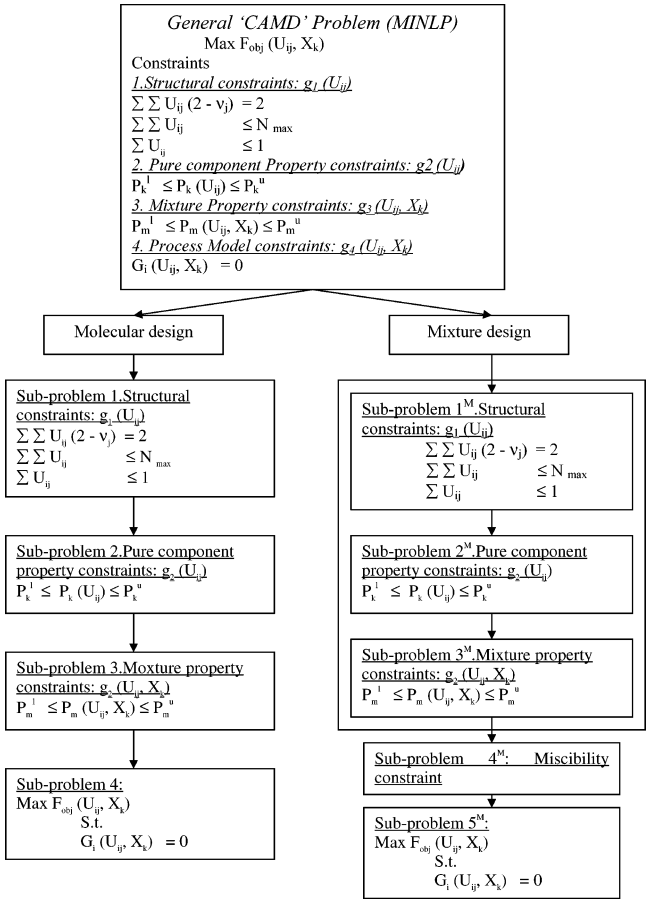

Figure 1. Decomposition methodology.

In this paper, we present a new CAMD methodology where the general molecule/mixture design problem (formulated as an MINLP model) is decomposed into an ordered set of subproblems. Each subproblem (except the final) requires only the solution of a subset of the constraints from the original set. The final subproblem contains the objective function and the remaining constraints. In this way, the solution of the decomposed set of subproblems is equivalent to that of the original MINLP problem. The advantage is a more flexible solution approach together with relatively easy to solve subproblems and a solvable "final" MINLP subproblem. As each subproblem is being solved, a large portion of the infeasible part of the search space is deleted, thereby leading to a final subproblem that is a significantly smaller MINLP or NLP problem, which can be solved more easily. As all the subproblems except the final subproblem are constraint satisfaction problems, global optimality can be guaranteed if a global optimization algorithm is used to solve the final subproblem.

4. Decomposition-Based Solution Strategy

The constraints in the general CAMD problem are first decomposed into two parts, namely, a molecular design (pure component design) part and a mixture design part as shown in Figure 1. If we are interested in single-compound solvent design, only the first part is needed, while if we are interested in mixture design, both parts are needed.

4.1. Single-Compound Design. In the first part (single compound), the problem is decomposed into four subproblems.

(a) Subproblem 1. This subproblem considers the structural constraints that result in generation of feasible molecular structures. Odele and Machietto structural feasibility constraints[18] are used. The subproblem is a function of the binary variables alone. These feasibility rules were developed for designing acyclic, monocyclic, and bicyclic compounds. If we are interested in designing complex multifunctional structures, then suitable feasibility rules need to be developed which can then be incorporated into this subproblem.

(b) Subproblem 2. This subproblem considers the pure component properties. This subproblem is also a function of binary variables alone (because these constraints handle only primary structure-based properties). The feasible molecular structures from subproblem 1 are solved for the pure component properties. Those molecules, which satisfy the pure component property constraints, are then passed into subproblem 3.

(c) Subproblem 3. This subproblem considers the mixture properties. Mixture properties can be categorized into two types. Properties such as selectivity, solvent power, etc., are based on infinite dilution activity coefficients, which are independent of composition, and hence, only structural information is needed for their calculation. Properties such as complete or partial miscibility of a solvent with another constituent are handled by discritizing the composition range from 0 to 1 into n divisions and verifying the miscibility criterion at those points. In this approach, only structural information is needed as input. The difference between pure component property constraints and mixture property constraints is that the former are linear and the later are nonlinear. Those satisfying the mixture property constraints are passed on to subproblem 4.

(d) Subproblem 4. In subproblem 4, the process model constraints (function of both integer and continuous variables) are considered along with the objective function. The optimal solvent is identified by solving either a smaller MINLP problem (if the number of feasible solution is large) or a set of NLP problems (if the number of feasible solutions is small) by fixing the values of integer variables, in subproblem 4.

4.2. Mixture Design. The second part deals with mixture design, which is also solved as a series of subproblems. In this part, promising pure component solvents are designed first and then candidate solvent mixtures are identified. The first three subproblems deal with the design of pure component solvents.

(a) Subproblem 1^{M}. As in subproblem 1, structural constraints are considered in subproblem 1^{M}. All the feasible molecular structures are passed to subproblem 2^{M}.

(b) Subproblem 2^{M}. Subproblem 2^{M} considers the pure component property constraints where the compounds from subproblem 1^{M} are evaluated for the pure component properties. All the molecules satisfying these constraints are passed onto subproblem 3^{M}.

(c) Subproblem 3^{M}. Subproblem 3^{M} considers the mixture property constraints. The molecules from subproblem 2^{M} are considered in this subproblem. The starting point is a list of promising solvents. From this list of solvents, the optimal mixture and the compositions of the constituents are identified by solving subproblem 4^{M} and subproblem 5^{M}. Since the first three subproblems in the mixture design involve the design of pure component solvents, these subproblems are essentially the same as the first three subproblems in single-compound design; see the highlighted sections in Figure 1.

(d) Subproblem 4^{M}. In subproblem 4^{M}, the miscibility of the solvents among themselves in the mixture (e.g., miscibility of two solvents if a binary mixture is designed) is considered.

(e) Subproblem 5^{M}. In subproblem 5^{M}, the process model constraints are considered along with the objective function and the optimal mixture is identified by solving a smaller MINLP problem (if the number of feasible solutions is large) or a set of NLP problems (if the number of feasible solutions is small).

In some mixture design problems (such as formulations), it may not be necessary to consider processing issues; hence, we would not have the process model constraints, and the problem would have become a simple mixing problem, which would already have been addressed by the miscibility criteria in subproblem 4^{M}. Hence, for these problems, we will not need subproblem 5^{M}. Also, in some cases, we might have to identify a mixture whose constituents perform different functions such as solvents and antisolvents for crystallization.[35] In such cases, we would have to formulate and solve more than one single-compound design problem to identify the constituents and then solve the final two subproblems to identify the optimal mixture. In certain single-compound and mixture design problems, we may not have process model constraints; however, we may still have to solve an optimization problem with other constraints, in subproblem 4 and subproblem 5^{M}, respectively.

5. Conceptual Validation

In this section, the solution through the proposed decomposition methodology is illustrated with the help of two analytical examples. The objective here is to highlight the applicability of the decomposition methodology. The first example is a small MINLP problem, which is solved through the decomposition approach. The second example is a small MINLP model with CAMD characteristics.

5.1. Example 1

$$\min\ 2x_1 + 3x_2 + 1.5y_1 + 2y_2 - 0.5y_3 \tag{1}$$

subject to

$$x_1^{\,2} + y_1 = 1.25 \tag{2}$$

$$x_2^{\,1.5} + 1.5y_2 = 3.0 \tag{3}$$

$$x_1 + y_1 \le 1.60 \tag{4}$$

$$1.333x_2 + y_2 \le 3.00 \tag{5}$$

$$-y_1 - y_2 + y_3 \le 0 \tag{6}$$

$$y_1 y_2 = 1 \tag{7}$$

$$x_1, x_2 \geq 0 \tag{8}$$

$$y_1, y_2, y_3 = \{0,1\} \tag{9}$$

This problem was decomposed into four subproblems.

Subproblem 1. First, all possible solutions in the decision (binary) variables are enumerated in subproblem 1. They are

$$(y_1, y_2, y_3) = (1,1,1),(0,0,0),(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1)$$

Subproblem 2. The linear constraint $-y_1 - y_2 + y_3 \leq 0$ is considered, and the feasible solutions from subproblem 1 satisfying this constraint are selected. They are (1,1,1),(0,0,0),(1,0,0),(0,1,0),(1,1,0),(1,0,1),(0,1,1)

Subproblem 3. The nonlinear constraint $y_1 y_2 = 1$ is considered, and the candidates from subproblem 2 satisfying this constraint, (1,1,1) and (1,1,0), are selected.

Subproblem 4. The problem is reduced to an NLP problem for each set of feasible solutions selected (i.e., by fixing the binary variables) in subproblem 3.

For example, for the first feasible solution, (1,1,1), the problem reduces to

$$\min 2x_1 + 3x_2 + 1.5 + 2 - 0.5$$

subject to

$$x_1^{\,2} + 1 = 1.25$$

$$x_2^{\,1.5} + 1.5 = 3.0$$

$$x_1 + 1 \leq 1.60$$

$$1.333x_2 + 1 \leq 3.00$$

$$x_1, x_2 \geq 0$$

For the two feasible solutions, two NLP problems are solved, and the solution having the minimum objective function value is the optimal solution for the MINLP problem.

The solutions of the two NLP problems are given in Table 1.

The smallest objective function value is 7.931111, corresponding to (1,1,1).

The optimal solution for the MINLP problem, which could also have been obtained by other means, is therefore

$$(y_1, y_2, y_3, x_1, x_2, f_{obj}) = (1,1,1,0.499999,1.310371,7.931111)$$

5.2. Example 2. This example is to illustrate the solution of a fictitious CAMD problem through the decomposition approach. In this binary mixture design problem, one of the compounds is known while the other compound, which is the solvent that needs to be designed, and the optimal composition of the two components need to be identified. The objective is to maximize some fictitious property, which is a function of the binary variable (subgroups) and continuous variable (composition) subject to structural and property constraints. The MINLP model of the above problem is shown below.

Table 1. Solution of NLP Problems

candidate	x_1	x_2	f_{obj}
(1,1,1)	0.499999	1.310371	7.931111
(1,1,0)	0.499999	1.310371	8.431111

Objective function:

$$\max 2x_1 + 20\sum_i U_{i1} + 18\sum_i U_{i2} + 12\sum_i U_{i3} \tag{10}$$

subject to

Structural constraints (subproblem 1^M)

$$\sum_i U_{i1}(2-1) + \sum_i U_{i2}(2-2) + \sum_i U_{i3}(2-1) = 2 \tag{11}$$

$$\sum_i U_{i1} + \sum_i U_{i2} + \sum_i U_{i3} - 3 \leq 0 \tag{12}$$

$$\sum_i U_{i1} + \sum_i U_{i2} + \sum_i U_{i3} - 2 \geq 0 \tag{13}$$

Pure component property constraints (subproblem 2^M)

$$198 + \sum_i U_{i1} \times 23.58 + \sum_i U_{i2} \times 22.58 + \sum_i U_{i3} \times 92.88 \geq 280 \tag{14}$$

$$122.5 + \sum_i U_{i1} \times -5.10 + \sum_i U_{i2} \times 11.27 + \sum_i U_{i3} \times 44.45 \leq 220 \tag{15}$$

Binary mixture property constraint (subproblem 5^M)

$$x_1\left(\sum_i U_{i1} \times 10 + \sum_i U_{i2} \times 12 + \sum_i U_{i3} \times 13\right) + x_2 \times 250 - 125 = 0 \tag{16}$$

Binary mixture constraint (subproblem 5^M)

$$x_1 + x_2 = 1 \tag{17}$$

$$x_1, x_2 \geq 0 \tag{18}$$

Note that subproblems 3^M and 4^M do not exist because the corresponding constraints are not needed in this problem. Subproblem 5^M combines the objective function and the corresponding constraints. In the binary variable U_{ij}, j represents the functional group in the basis set and i represents the position in the molecule. In this problem, the maximum number of groups (and hence the maximum number of positions in the molecule) allowed is 3 (eq 12) and the minimum number of groups (and hence the minimum number of positions allowed) is 2 (eq 13). Hence, $i = 1$, 2, or 3 and $j = 1$, 2, or 3. By this representation, we allow the same functional group to occur more than once in the molecule. x_1 and x_2 are compositions of the solvent and known compound, respectively.

The $\{CH_3, CH_2, OH\}$ basis set is chosen. This is the set of distinct functional groups from which molecules are formed.

The corresponding valency set is {1,2,1}. These numbers are used in constraint 1.

The numerals, which are coefficients to the binary variable in the objective function, are fictitious contribu-

Table 2. NLP Solution of Subproblem 4

	CH_3CH_2OH	CH_3OH
x_1	0.581	0.55
x_2	0.419	0.449
f_{obj}	51.163	33.101

tions of the groups to the property. Equations 14 and 15 represent constraints on the boiling point and melting point estimated using group contribution models.

Equation 16 is a binary mixture property constraint.

Subproblem 1^M. Equations 11−13 are typical structural constraints in a CAMD problem. Equation 11 is the octet rule for designing acyclic compounds. Equation 12 sets an upper limit of three subgroups for the designed molecule, and eq 13 sets a lower limit of two functional groups. These constraints are similar to the ones in subproblem 1 of Figure 1. The feasible molecules satisfying these constraints are as follows. (1) $U_{11},U_{22},U_{33} = 1$; all other U's = 0, i.e., CH_3CH_2OH. (2) $U_{11},U_{21} = 1$; all other U's = 0, i.e., CH_3CH_3. (3) $U_{11},U_{23} = 1$; all other U's = 0, i.e., CH_3OH. (4) $U_{11},U_{22},U_{31} = 1$; all other U's = 0, i.e., $CH_3CH_2CH_3$.

Subproblem 2^M. Equations 14 and 15 represent two examples of property constraints, the normal boiling point and normal melting point, estimated using the method of Joback and Reid.[3]

The feasible molecules from subproblem 1 satisfying constraints 14 and 15 are as follows. (1) $U_{11},U_{22},U_{33} = 1$; all other U's = 0. (3) $U_{11},U_{23} = 1$; all other U's = 0.

Subproblems 4^M and 5^M. In this problem, it is assumed that all candidate molecules will be completely miscible when mixed. Therefore, a miscibility constraint is not necessary. The remaining constraints and the objective function form the final MINLP problem.

Now the binary variables are fixed. We have two sets of fixed binary variables ($U_{11},U_{22},U_{33} = 1$ and $U_{11},U_{23} = 1$) representing two solvents. One of these solvents will form the mixture along with the known compound. Therefore, when we solve the final optimization problem, we will identify the optimal solvent (among the two solvents) and the optimal composition of the solvent−known compound pair.

The problem is reduced into an NLP problem for each set of feasible solutions that is selected (i.e., by fixing the binary variables) in subproblem 3. Hence, we have two NLP problems to solve.

For example, for the first feasible solution, CH_3CH_2OH, the problem reduces to

$$\max 2x_1 + (20 + 18 + 12)$$

subject to

$$x_1(10 + 12 + 13) + x_2 \times 250 - 125 = 0$$

$$x_1 + x_2 = 1$$

$$x_1,x_2 \geq 0$$

The solution for the two NLP problems is given in Table 2.

From the two solutions, CH_3CH_2OH which has the maximum objective function value is the optimal compound.

6. Software Implementation Details

The requirements for the proposed methodology are an algorithm to generate molecular structures (subproblems 1 and 1^M), an algorithm for evaluating the various pure component and mixture properties and for screening the feasible structures (subproblems 2, 3, 2^M, and 3^M), and an algorithm for the solution of the final optimization problem (subproblem 4 or subproblem 5^M). In the case of mixture design, we need an additional step to verify the miscibility criterion (subproblem 4^M). The methodology is implemented in two linked parts. The tools employed in these two parts are ProCAMD[36] and OPT-CAMD$^+$, which was developed for this work. Molecule generation and property evaluations are carried out in ProCAMD, while the optimization part is carried out in OPT-CAMD$^+$. ProCAMD uses a multilevel algorithm for generation of molecular structure and property predictions, the details of which can be found in Harper et al.[37] In ProCAMD, constraints can be provided for 26 non-temperature-dependent pure component properties, six temperature-dependent pure component properties, and nine mixture properties. Also, the miscibility calculation can be performed. The results of ProCAMD are fed into OPT-CAMD$^+$, which solves the final optimization part, where the performance index is optimized subject to the process model constraints. The flowchart of the OPT-CAMD$^+$ algorithm is shown in Figure 2.

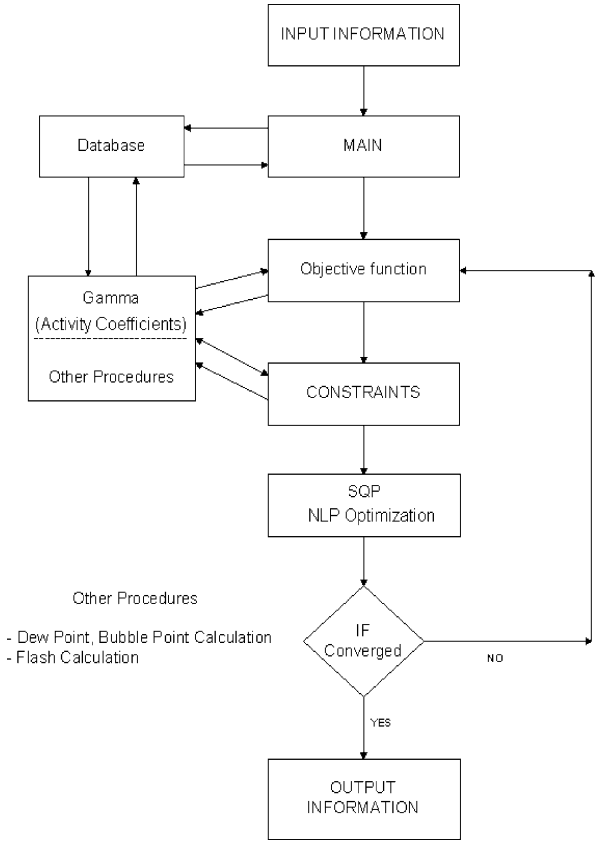

Figure 2. Flow diagram for OPT-CAMD$^+$.

In OPT-CAMD$^+$, the input information to the main program includes the identities of the nonsolvent components, identities of all the solvent molecules from subproblem 3 or 4^M (binary variables) of ProCAMD, and other known parameters such as temperature, pressure, and feed compositions. Information about the choice of the particular type of UNIFAC or KT UNIFAC method and the corresponding group parameters must be provided. Also, parameters for the optimization algorithm such as the tolerance, convergence criterion, maximum number of iterations, etc., need to be given.

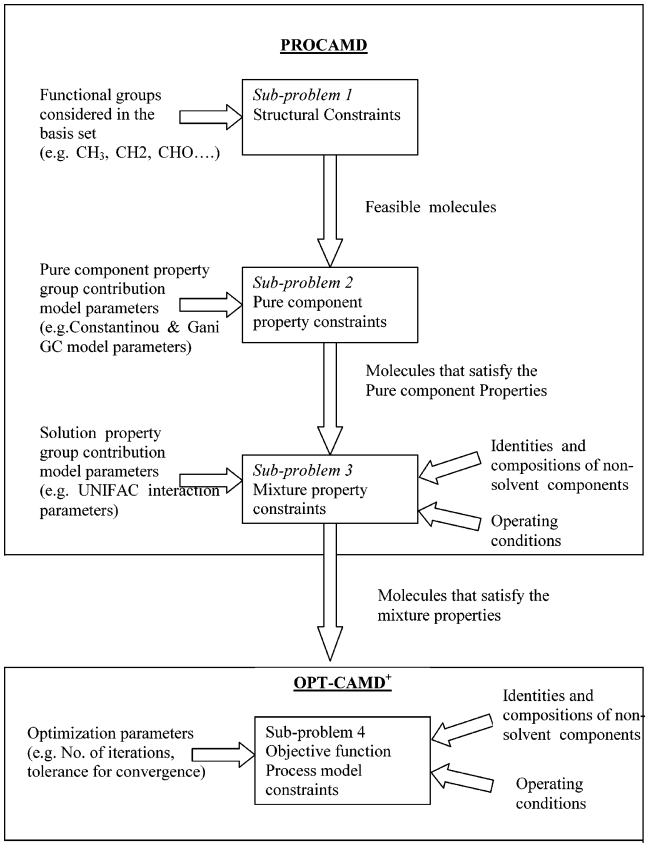

Figure 3. Data flow diagram (single-compound design).

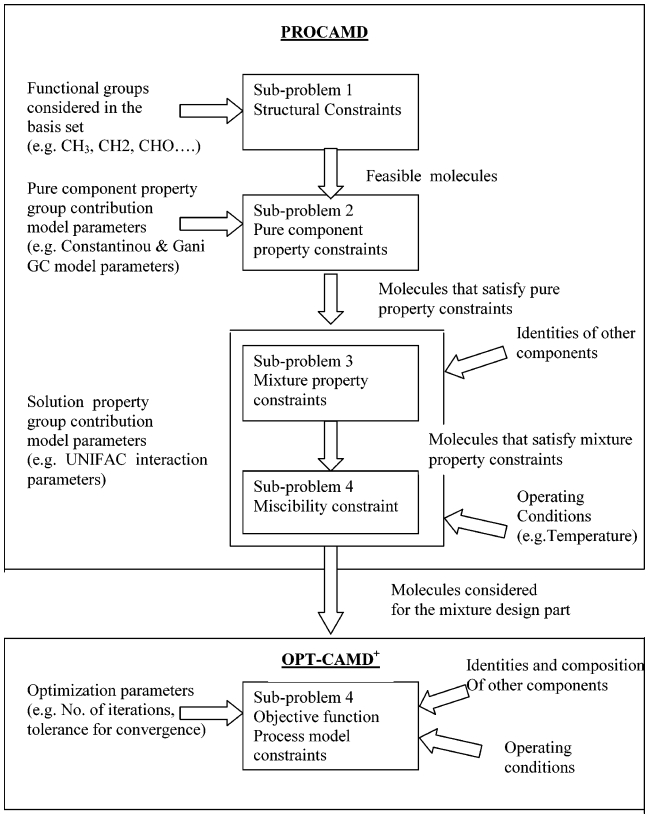

Figure 4. Data flow diagram (mixture design).

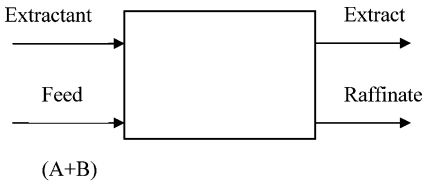

Figure 5. Single-stage extraction unit.

The process model constraints (phase equilibrium) and the objective function usually require computation of liquid-phase activity coefficients. For the computation of the constraints and the objective function, special routines for the calculation of activity coefficients (UNIFAC or KT UNIFAC) and other property models and phase equilibrium (including saturation points) are called. Instead, any other procedure for calculating activity coefficients on the basis of a group contribution approach can also be used through user-supplied external routes. Once these pieces of information are specified, the optimization algorithm is called. We used the sequential quadratic programming algorithm[38] in which subproblem 4 or subproblem 5^M is solved as a series of NLP problems. If subproblem 4 or 5^M needs to be posed as an MINLP problem (in cases where the number of feasible solutions sent to the final subproblem is very large), then an MINLP solver can be called instead. Data flow diagrams are shown in Figures 3 and 4.

7. Case Studies

7.1. Case Study 1. (a) Description of the Problem. Recovering acetic acid from an acetic acid/water mixture is desired. The feed contains 8 wt % acetic acid, and the feed flow rate is 13 500 kg/h. The removal of acetic acid is to be accomplished by a single-stage liquid−liquid extraction at 25 °C. The flow rate of the solvent is fixed at 16 300 kg/h. The goal is to design the optimum extractant such that the flow rate of acetic acid in the raffinate is minimal. This problem taken from Seader and Henley[39] is used here mainly as a proof of concept.

A liquid−liquid extraction is a process in which two liquid components are separated by the addition of a third liquid component called the extractant. This task is accomplished by selecting a suitable extractant, which is either partially miscible or immiscible with one of the components (initial solvent). It should also be highly soluble in the solute. Here, the selective nature of the solvent in the separation of the components is based on the difference in solubilities rather than the differences in volatilities (as in the case of distillation). In a single-stage extraction unit, the extractant and the solution are mixed together and then allowed to separate into two phases, the extract phase containing the solute in the extractant and the raffinate phase containing the initial solvent. The extractant needs to strongly favor partition of the solute away from the initial solvent so that the quantity of extractant required will be minimal. Also, the solubility of the extractant in the feed solvent should be minimal to prevent high extractant loss.[17]

A schematic of a typical extraction unit is shown in Figure 5.

Pretel et al.[16] have described a strategy for the design of solvents for extraction processes. They generate optimal molecules by using solvent properties such as the distribution coefficient and solvent loss as constraints and solving the resulting property satisfaction optimization problem. They do not explicitly consider the process model, and consequently, they do not have to evaluate the exact compositions of the different components in the two phases. Since the process models

Table 3. UNIFAC Subgroups Considered in the Basis Set

subgroup	subgroup number	main group number	subgroup	subgroup number	main group number
CH_3	1	1	CH_3COO	22	11
CH_2	2	1	CH_2COO	23	11
CH	3	1	HCOO	24	12
C	4	1	CH_3O	25	13
OH	15	5	CH_2O	26	13
CH_3OH	19	9	CHO	27	13
CH_2OH	20	9	COOH	43	20
CHO	21	10	COO	77	41

are not considered, the nonlinearity associated with these models is not present and consequently the solution is much simpler. On the other hand, Naser and Fournier[17] have used a strategy in which the process model alone is stated explicitly as constraints and the optimization problem is solved for optimal extractant. One advantage of solving the problem explicitly in terms of the process model is that it also quantitatively validates the designed extractant, i.e., the compositions of all the components in the two phases are evaluated. In our approach, the solvent properties and the process model are considered together. In the proposed decomposition-based approach, the advantage is that only those molecules that satisfy the property constraints are solved for the process model.

(b) Detailed Formulation of the Problem. Objective function

$$f_{\text{obj}} = \min(x_A^R R) \tag{19}$$

Constraints

$$\sum_i \sum_j u_{ij}(2 - v_j) = 2 \tag{20}$$

$$\sum_i \sum_j u_{ij} = N_{\max} \tag{21}$$

$$\sum_j u_{ij} = 1 \tag{22}$$

$$T_{\text{m}} = 102.425 \sum_i N_i T_{\text{m}i} + \sum_j M_j T_{\text{m}j} \leq 270 \tag{23}$$

$$T_{\text{b}} = 204.359 \sum_i N_i T_{\text{b}i} + \sum_j M_j T_{\text{b}j} \geq 340 \tag{24}$$

$$m = \frac{\gamma_{A,B}^{\infty}}{\gamma_{A,S}^{\infty}} \frac{MW_B}{MW_S} \geq 0.49 \tag{25}$$

$$\text{SL} = \frac{1}{\gamma_{S,B}^{\infty}} \leq 0.0038 \tag{26}$$

$$\beta = \frac{\gamma_{B,S}^{\infty}}{\gamma_{A,S}^{\infty}} \geq 11 \tag{27}$$

$$\text{SP} = \frac{1}{\gamma_{A,S}^{\infty}} \geq 0.778 \tag{28}$$

$$\gamma_A^E \times x_A^E - \gamma_A^R \times x_A^R = 0 \tag{29}$$

$$\gamma_B^E \times x_B^E - \gamma_B^R \times x_B^R = 0 \tag{30}$$

$$\gamma_S^E \times x_S^E - \gamma_S^R \times x_S^R = 0 \tag{31}$$

$$x_A^E \times E - x_A^R \times R = F x_A^F \tag{32}$$

$$x_B^E \times E - x_B^R \times R = F x_B^F \tag{33}$$

$$x_S^E \times E - x_S^R \times R = F x_S^F \tag{34}$$

$$x_A^E + x_B^E + x_S^E = 1 \tag{35}$$

$$x_A^R + x_B^R + x_S^R = 1 \tag{36}$$

u_{ij} is the binary variable indicating whether the ith position in a molecule has structural group j. v_j represents the valence of group j. $N_{\max}$ represents the maximum number of positions in a molecule. Subscripts A, B, and S represent the solute, initial solvent, and extractant, respectively. E and R represent the flow rate of the extract and raffinate phase, respectively. N_i is the number of times the first-order group i is present in the molecule. M_j is number of times the second-order group j is present in the molecule. $T_{\text{m}i}$ and $T_{\text{b}i}$ are the contributions of first-order groups to the melting point and boiling point, respectively. $T_{\text{m}j}$ and $T_{\text{b}j}$ are the contributions of second-order groups to the melting point and boiling point, respectively.

Feed compositions x_A^F, x_B^F, and x_S^F are 0.0002, 0.2878, and 0.6970, respectively, while feed flow rate F = 2334.77 kg mol^{-1} h^{-1}. The compositions of the three components in the extract (x_A^E,x_B^E,x_S^E) and raffinate (x_A^R,x_B^R,x_S^R) phases as well as the flow rate of the raffinate and extract phases are unknown. In the problem, the feed flow rates and the compositions are given in kilograms per hour and weight fraction, respectively. These are converted into kilogram-mole per hour and mole fraction, respectively, at appropriate places.

(c) Decomposed Problem. (i) Subproblem 1. Structural Constraints. Equations 20−22, which represent structural constraints, are considered in subproblem 1. These constraints are imposed to (a) ensure that the number of bonds attached to a group equals the valence of the group, (b) limit the number of groups in the molecule, and (c) ensure that only one group is present in a position.

The design of only acyclic compounds is desired. Further, the classes of compounds that we intend to generate are alcohols, ketones, aldehydes, acids, esters, and ethers. The UNIFAC groups that are required to generate these classes of compounds are considered in the basis set, and they are listed in Table 3 along with their subgroup and main group numbers.

(ii) Subproblem 2. Pure Component Property Constraints. Equations 23 and 24, which represent pure component property constraints (normal boiling point and normal melting point, respectively), are considered in subproblem 2. The extractant should be in the liquid state near the operating temperature and

pressure (298 K and 1 atm, respectively). Limits on the normal boiling point and normal melting point of the extractant need to be imposed. Since we will usually separate the solute from the extractant by passing the extract through a distillation unit, we need to ensure high relative volatility for the solute−extractant pair by imposing a minimum boiling point difference of 30−40 °C between the solute and the extractant. The boiling point and melting point are estimated using the Constantinou and Gani group contribution method.[4]

(iii) Subproblem 3. Mixture Property Constraints. Equations 25−28, which represent the various mixture properties, are considered in subproblem 3.

Pretel et al.[16] have described in detail solvent selection of these classes of problems. They state that the dominant properties of this family of separation problems are the solute distribution coefficient, solvent loss, solvent power, and solvent selectivity.

(a) Distribution Coefficient (*m*). The solvent capacity determines the flow rate of the circulating solvent and consequently the design of the extractor. The distribution coefficient of the solute between the extract and the raffinate phases is a measure of the solvent capacity. The size of the extractor is inversely proportional to the value of the distribution coefficient.

(b) Solvent Loss (SL). The solvent solubility in the raffinate phase should be very low.

Low solvent solubility in the raffinate phase determines raffinate−extract immiscibility.

(c) Selectivity (*β*). High selectivity is required to reduce the cost of recovery of the solute from the extract.

(d) Solvent Power (SP). The solvent power should be high.

All the above properties are functions of infinite dilution activity coefficients $\gamma_{A,B}^{\infty}$, $\gamma_{A,S}^{\infty}$, $\gamma_{S,B}^{\infty}$, and $\gamma_{B,S}^{\infty}$ which are predicted using the UNIFAC group contribution method.[6] These solvent properties are defined in the Appendix.

(iv) Subproblem 4. The optimization problem to be solved in subproblem 4 is shown below.

Objective function

$$f_{\text{obj}} = \min(x_A^R R)$$

Constraints

$$\gamma_A^E \times x_A^E - \gamma_A^R \times x_A^R = 0$$

$$\gamma_B^E \times x_B^E - \gamma_B^R \times x_B^R = 0$$

$$\gamma_S^E \times x_S^E - \gamma_S^R \times x_S^R = 0$$

$$x_A^E \times E - x_A^R \times R = F x_A^F$$

$$x_B^E \times E - x_B^R \times R = F x_B^F$$

$$x_S^E \times E - x_S^R \times R = F x_S^F$$

$$x_A^E + x_B^E + x_S^E = 1$$

$$x_A^R + x_B^R + x_S^R = 1$$

The objective is to minimize the quantity of acetic acid in the raffinate phase.

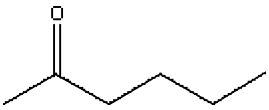

Figure 6. Designed extractant, 2-hexanone.

Process Model Constraints. A generalized thermodynamic description of the problem of designing an optimum extractant[17] can be formulated mathematically as follows.

Phase equilibrium

$$\gamma_{i,1} x_{i,1} - \gamma_{i,2} x_{i,2} = 0 \tag{37}$$

$$\sum_i x_i^E = 1 \tag{38}$$

$$\sum_i x_i^R = 1 \tag{39}$$

Mass balances

$$x_i^E E - x_i^R R - x_i^F F = 0 \tag{40}$$

$i = 1, 2, ..., N$ (number of components).

(d) Results and Discussion. Part 1 of the methodology was used since this problem is a pure component solvent design problem; 4429 feasible molecular structures were generated in subproblem 1. Of the 4429 molecules, 2796 satisfied the two pure component property constraints in subproblem 2. These molecules were considered in subproblem 3, where they were solved for the mixture property constraints. Only two of the 2796 molecules satisfied the mixture property constraints. In subproblem 4, the final optimization problem with process model constraints was solved for the two compounds. Two NLP problems corresponding to the two compounds were solved by the sequential quadratic programming (SQP) algorithm. Since only two molecules were passed into the final subproblem, the optimization part became much simpler. The optimal extractant was found out to be 2-hexanone. The results of each subproblem are shown in Table 4. The design results of the optimal compound, 2-hexanone, are given in Table 5.The structure of the optimal compound is shown in Figure 6.

The percentage recovery of acetic acid was 88.63%. The computational time for ProCAMD was 1.77 s.

For this particular problem, methyl ethyl ketone has been suggested as a solvent.[39] Here, it is interesting to note that the other compound that was passed into the fourth subproblem along with 2-hexanone was methyl ethyl ketone. This establishes that our problem formulation gives good results and is able to identify good candidates. It also implies that the method was able to identify a solvent better than the existing ones. However, final experimental verification is needed to confirm this result.

7.2. Case Study 2. (a) Description of the Problem. This case study is a real industrial problem for which this CAMD methodology was successfully employed. We cannot disclose the identity of the drug because of a confidentiality agreement with the company. This drug has very low water solubility, and the company has

Table 4. Subproblem Results

subproblem 1: no. of molecules generated, 4429
subproblem 2: no. of molecules satisfying pure component properties, 2796 of 4429
subproblem 3: no. of molecules satisfying mixture properties, 2 of 2796
subproblem 4: optimal extractant, 2-hexanone

Table 5. Design Results of the Optimal Compound (1CH$_3$,3CH$_2$,1CH$_3$CO)

T_m	T_b	m	SL	β	SP
214.45	404.09	0.491	0.0038	11.00	0.778
x_A^E	x_B^E	x_S^E	x_A^R	x_B^R	x_S^R
0.0084	0.2117	0.7799	0.0017	0.9941	0.0042
E		R		f_{obj}	
2089		249.2		0.4281	

problems in identifying a suitable formulation for the drug. The objective is to design a solvent formulation for the drug. Specifically, we need to identify an organic solvent which when added to water in a small proportion increases the solubility of the solute in the water/solvent mixture. The designed organic solvent should be completely miscible with water, and the major constituent in the final formulation should be water, because of toxicity reasons.

The water solubility of the drug estimated using the Marrero−Gani group contribution method[40] is 0.0001 mole fraction and by the UNIFAC method[6] is 0.001 mole fraction. Since experimental data are not available, these numbers are used as possible upper and lower bounds, respectively (with the observation that the value from the Marrero−Gani method of prediction is likely to be more accurate than the UNIFAC calculated value).

The designed solvent should have the following characteristics. (1) It should be a liquid at operating conditions. (2) It should have high solubility for the drug. (3) It should be completely miscible with water. (4) It should have low toxicity.

(b) Detailed Formulation of the Problem. Objective function

$$f_{obj} = \min(x_2\tau) \tag{41}$$

Constraints

$$\sum_i \sum_j u_{ij}(2 - v_j) = 2 \tag{42}$$

$$\sum_i \sum_j u_{ij} = N_{max} \tag{43}$$

$$\sum_j u_{ij} = 1 \tag{44}$$

$$T_m = 102.425\sum_i N_i T_{mi} + \sum_j M_j T_{mj} \le 270 \tag{45}$$

$$T_b = 204.359\sum_i N_i T_{bi} + \sum_j M_j T_{bj} \ge 340 \tag{46}$$

$$17 \le \delta \le 19 \tag{47}$$

$$\frac{1}{x_2} + \frac{\partial \ln \gamma_2}{\partial x_2} \ge 0 \tag{48}$$

$$\ln x_i^{sat} - \frac{\Delta_{fus}H}{T_m}\left(1 - \frac{T_m}{T}\right) + \ln \gamma_1^{sat} = 0 \tag{49}$$

$$\exp\left(\frac{\Delta_{fus}H}{T_m}\left(1 - \frac{T_m}{T}\right) - \ln \gamma_1^{sat}\right) \ge 0.05 \tag{50}$$

$$x_2 \le 0.2 \tag{51}$$

$$x_1 + x_2 + x_3 = 0 \tag{52}$$

The solubility of a nonionic solute can be expressed with eq 49. τ is the toxicity of the organic solvent. x_i^{sat} is the mole fraction of the solute dissolved in the solvent at saturation. γ_i^{sat} is the activity coefficient for the solute in the solution. T is the temperature. $\Delta_{fus}H$ is the solids heat of fusion. T_m is the melting point of the solid. $\Delta_{fus}H$ and T_m are pure component properties of solute independent of the solvent. The activity coefficient is a mixture property of the solute and solvent. The activity coefficient itself is a function of compositions and hence the nonlinearity. The UNIFAC method has been used to estimate the activity coefficient. The estimated properties of the solute are as follows: $\Delta_{fus}H = 32.96$ kg/mol and $T_m = 444.64$ K.

(c) Decomposed Problem. (i) Subproblem 1^M. Structural Constraints. The three structural constraints represented by eqs 42−44 that constituted subproblem 1 in the first case study are used in subproblem 1^M again. Design of only acyclic compounds is desired. The same UNIFAC groups listed in Table 1 are considered in the basis set.

(ii) Subproblem 2^M. Pure Component Property Constraints. The three pure component property constraints represented by eqs 45−47 are considered in subproblem 2^M. As in the first case study, to ensure that the solvent is a liquid at operating conditions, an upper limit of 270 K and a lower limit of 340 K are imposed on the melting point and boiling point, respectively. The same models described in subproblem 2 of case study 1 are used in evaluating the normal melting point and normal boiling point. The solubility parameter (δ) was used as a constraint to design solvents having high solubility for the solute. It is widely believed that "like dissolves like", implying if the solvent and the solute have similar δ values then the solubility of the solute in that solvent is high. The estimated solubility parameter of the drug is 18.02 MPa$^{1/2}$. Hence, a lower limit of 17 MPa$^{1/2}$ and an upper limit of 19 MPa$^{1/2}$ are placed on the solubility parameter. This is only a qualitative way of identifying good solvents.

(iii) Subproblem 3^M. There are no mixture property constraints related to pure component solvent design. Hence, we do not have subproblem 3^M of the methodology for mixture design.

(iv) Subproblem 4^M. Miscibility Constraint. The miscibility constraint represented by eq 48 is considered in subproblem 4^M. Since the solvent should be completely miscible with water, the solubility parameter concept, which also applies to liquid miscibility, can be used. Two substances with similar δ values are usually miscible with each other, with some exceptions. The

Figure 7. Optimal solvent, 1,2-dimethoxyethane.

solubility parameter of water is 47.8 $MPa^{1/2}$. This means that solvents that have a δ value close to 18.02 $MPa^{1/2}$ will most likely not be miscible with water. Effectively, we are trying to find a solvent that has a solubility parameter close to 18 $MPa^{1/2}$ and at the same time is completely miscible with water, which could be very rare. Since we need to find solvents that are exceptions to the like dissolves like theory, the stability function method is used instead of the solubility parameter method to evaluate the miscibility of the solvent and water. The stability function method[41] gives the necessary and sufficient condition for the phase stability of a binary pair in terms of mole fraction and activity coefficients.

$$\frac{1}{x_2} + \frac{\partial \ln \gamma_2}{\partial x_2} \geq 0$$

The activity coefficient is calculated using the UNIFAC method.[7] If the above condition is satisfied for the entire composition range, the pair is completely miscible. Only acyclic solvents are designed, as they are usually less toxic than aromatics.

(v) Subproblem 5^M. The optimization problem to be solved in subproblem 5^M is shown below

$$f_{obj} = \min(x_2\tau)$$

Subject to the constraints

$$\ln x_i^{sat} - \frac{\Delta_{fus}H}{T_m}\left(1 - \frac{T_m}{T}\right) + \ln \gamma_1^{sat} = 0$$

$$\exp\left(\frac{\Delta_{fus}H}{T_m}\left(1 - \frac{T_m}{T}\right) - \ln \gamma_1^{sat}\right) \geq 0.05$$

$$x_2 \leq 0.2$$

$$x_1 + x_2 + x_3 = 1$$

The objective is to minimize the toxicity (τ) of the final formulation. The toxicity depends on the composition of the organic solvent in the formulation and hence is evaluated as the product of composition of the organic solvent in the final formulation times the toxicity of the solvent. In this study, $-\log(LC_{50})$ is considered a quantitative measure of the toxicity. LC_{50} represents the aqueous concentration causing 50% mortality in fathead minnow after 96 h. The higher the value of $-\log(LC_{50})$, the more toxic the compound. $-\mathrm{Log}(LC_{50})$ values are evaluated using a group contribution method.[42] The first constraint represented by eq 49 is the solid−liquid equilibrium constraint. The second constraint represented by eq 50 makes sure that the solubility of the drug in the water/solvent mixture is greater than 5%. The third constraint represented by eq 51 makes sure that the composition of the solvent in the final formulation is less than 20%. The fourth constraint represented by eq 52 is for mole fraction satisfaction.

(d) Results and Discussion. This problem is a mixture design problem; 3498 feasible molecular structures were generated in subproblem 1^M. Of the 3498 molecules, 645 molecules satisfied the two pure component property constraints in subproblem 2^M. Subproblem 3^M of the methodology was ignored because we did not have any mixture properties related to pure component solvent design in this problem. The 645 molecules from subproblem 2^M were considered in subproblem 4^M, where they were solved for the miscibility constraint. Only five of the 645 molecules satisfied the miscibility constraint. In subproblem 5^M, the optimal formulation is identified by solving the optimization problem for the five compounds. The results of each subproblem are given in Table 6. The properties of the optimal pure component solvent 1,2-dimethoxyethane are given in Table 7. The structure of the optimal compound is shown in Figure 7.The design results of the optimal formulation are given in Table 8. The objective function value of the optimal formulation is $f_{obj} = \min(x_2\tau) = 0.2709$. The computational time for ProCAMD was 4.58 s.

Table 6. Subproblem Results

subproblem 1^M: no. of molecules generated, 3498
subproblem 2^M: no. of molecules satisfying pure component properties, 645 of 3498
subproblem 3^M: no. of molecules satisfying mixture properties, 5 of 645
subproblem 4^M: optimal solvent, 1,2-dimethoxyethane

Table 7. Design Results of the Optimal Single Compound ($2CH_2$,$2CH_3O$)

T_b	T_m	δ	miscibility
377.82	209.06	17.93	completely miscible

Table 8. Design Results of the Formulation

constituent	composition
solute	0.05
1,2-dimethoxyethane	0.1642
water	0.7858

Table 9. Mole Percent Solubility of the Solute in Different Solvents

x_{solv}	x_w	water − 1,2-dimethoxyethane	water − ethanol	water − ethylene glycol
0.05	0.95	1.59	0.62	0.66
0.1	0.9	3.48	2.27	1.99
0.15	0.85	4.64	3.97	3.23
0.2	0.8	5.34	5.10	4.07

(e) Analysis. It is known that ethyl alcohol and ethylene glycol are presently being employed as cosolvents with water to improve the water solubility of the solute. Hence, in this section, we compare 1,2-dimethoxyethane with the above-mentioned solvents. The mole percent solubility of the solute at various compositions (ranging from 0.05 to 0.2) of the water/solvent mixture at 298 K is evaluated using the UNIFAC method.[6] The results are given in Table 9. SLE plots of temperature versus composition of solute on a solvent free basis with the solvent composition fixed at 0.1 and 0.2 are shown in Figures 8 and 9, respectively.

From Table 9 and Figures 8 and 9, it is clear that 1,2-dimethoxyethane will be able to perform better than ethanol and ethylene glycol. At a lower composition of solvent (0.1), 1,2-dimethoxyethane performs significantly better than ethanol, particularly at lower temperatures. At a higher composition of solvent (0.2), 1,2-dimethoxyethane still performs better than ethanol, but the difference is small. Since the requirements of the formulation are a low composition of solvent and a low temperature (298 K), we can safely say that 1,2-dimethoxyethane performs better than both ethanol and

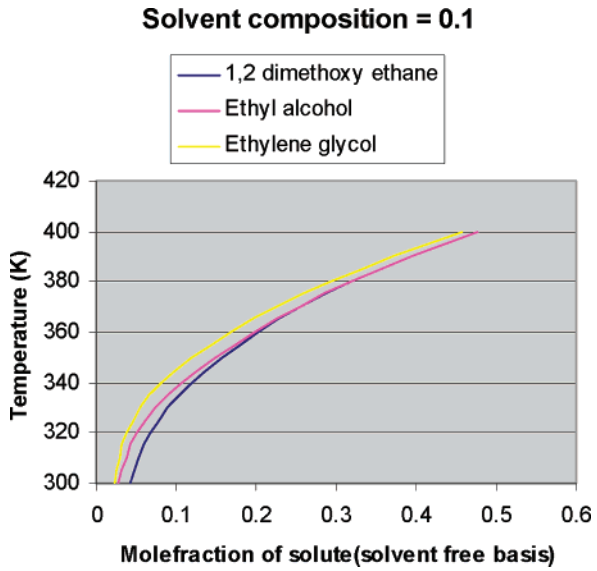

Figure 8. SLE diagram 1.

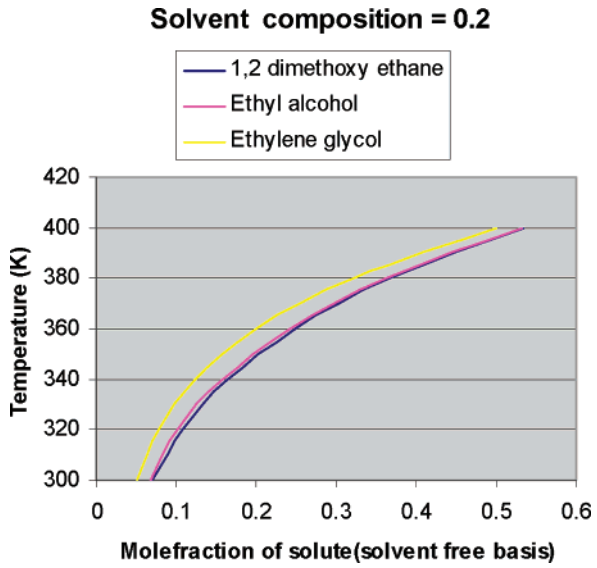

Figure 9. SLE diagram 2.

ethylene glycol. However, final experimental verification is needed to confirm these results.

8. Other Case Studies

The decomposition solution strategy for CAMD has been used to solve a number of molecule and mixture design problems. A solvent design problem for a drug, ibuprofen, has been described by Karunanithi et al.,[35] and a solvent mixture design problem solved using the interval analysis technique of Achenie et al.[33] has also been solved using the decomposition-based methodology.

9. Conclusions

This paper presents a new decomposition-based solution strategy for solving computer-aided molecular/mixture design problems. The proposed methodology involves formulating the molecular/mixture design problem as an MINLP model and solving the problem as a series of solvable subproblems. For mixture design, the pure component molecular design problem is solved to identify candidates and then the mixture design is carried out. The structural constraints, pure component property constraints, mixture property constraints, miscibility constraints, and objective function along with process model constraints are handled in separate subproblems. This approach systematically reduces the search space, thereby making the solution of the final optimization problem easier. Also, some of the property models are handled in the initial subproblems, and hence, they do not bring the complexities associated with them into the final optimization problem. The methodology has been implemented in computer-aided tools, namely, ProCAMD and OPT-CAMD$^+$, which are linked together. The former is used for molecular structure generation and property predictions, while the latter is used to solve the final optimization problem. The application of this methodology was illustrated with the help of two case studies. In the first case study, an optimal extractant was designed for separating acetic acid from water by liquid−liquid extraction. In the second case study, the problem of optimal formulation of a pharmaceutical compound was addressed. The results from these case studies indicate the effectiveness of the proposed methodology. Finally, although the application of the decomposition-based methodology has been highlighted through solvent/solvent−mixture design problems, in principle, it is applicable to a vast range of molecular and mixture design problems. For example, given a set of candidate molecules (found through molecular design), find the optimal mixture for a specific function (delivery of drug or pesticide, provide heating or cooling, increase the rate of uptake of drug or pesticide, control evaporation rate, etc.); these are problems found in typical chemical product design. Current and future work will provide case studies highlighting the application of the decomposition-based methodology to various chemical product design problems.

Acknowledgment

This material is based upon work supported by National Science Foundation Grant 0109928. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. A.T.K. and L.E.K.A. also thank CAPEC, Department of Chemical Engineering, Technical University of Denmark, for providing us with the ICAS and UNIFAC software for activity coefficients.

Appendix

$$\text{distribution coefficient } (m) = \frac{\gamma^{\infty}_{A,B}}{\gamma^{\infty}_{A,S}} \frac{MW_B}{MW_S}$$

$$\text{solvent loss (SL)} = \frac{1}{\gamma^{\infty}_{S,B}}$$

$$\text{selectivity } (\beta) = \frac{\gamma^{\infty}_{B,S}}{\gamma^{\infty}_{A,S}}$$

$$\text{solvent power (SP)} = \frac{1}{\gamma^{\infty}_{A,S}}$$

Literature Cited

(1) Achenie, L. E. K.; Gani, R.; Venkatasubrmaniam, V. *Computer aided molecular design: theory and practice*; Elsevier Science: Amsterdam, 2003.

(2) Constantinou, L.; Bagherpour, K.; Gani, R.; Klein, J. A.; Wu, D. T. Computer aided product design: Problem formulations, methodology and applications. *Comput. Chem. Eng.* **1996**, *20*, 685−702.

(3) Joback, K. G.; Reid, R. C. Estimation of pure component properties from group contributions. *Chem. Eng. Commun.* **1987**, *57*, 233.

(4) Constantinou, L.; Gani, R. New group-contribution method for estimating properties of pure compounds. *AIChE J.* **1994**, *40* (10), 1697−1710.

(5) Marrero, J.; Gani, R. Group contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, *183*−*184*, 183−208.

(6) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086.

(7) Kang, J. W.; Abildskov, J.; Gani, R.; Cobas, J. Estimation of mixture properties from first and second-order group contributions with the UNIFAC model. *Ind. Eng. Chem. Res.* **2002**, *41* (13), 3260−3273.

(8) Kier, L.; Hall, L. *Molecular connectivity in chemistry and drug research*; Academic Press: New York, 1976.

(9) Camarda, K.; Maranas, C. D. Optimization in polymer design using connectivity indices. *Ind. Eng. Chem. Res.* **1999**, *38* (8), 1884−1892.

(10) Venkatasubramaniam, V.; Chan, K.; Caruthers, J. M. Computer aided molecular design using genetic algorithms. *Comput. Chem. Eng.* **1994**, *18* (9), 833−844.

(11) Vaidyanathan, R.; El-Halwagi, M. Computer aided synthesis of polymers and blends with target properties. *Ind. Eng. Chem. Res.* **1996**, *35* (2), 627−634.

(12) Maranas, C. D. Optimal computer-aided molecular design: A polymer design case study. *Ind. Eng. Chem. Res.* **1996**, *35* (5), 3403−3414.

(13) Gani, R.; Brignole, E. A. Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilib.* **1983**, *13*, 331.

(14) Brignole, E. A.; Bottini, S.; Gani, R. A strategy for the design and selection of solvents for separation processes. *Fluid Phase Equilib.* **1986**, *29*, 125−132.

(15) Macchietto, S.; Odele, O.; Omatsone, O. Design of optimal solvents for liquid−liquid extraction and gas absorption processes. *Trans. Inst. Chem. Eng.* **1990**, *68* (A), 429−433.

(16) Pretel, E. J.; Lopez, P. A.; Bottini, S. B.; Brignole, E. A. Computer-aided molecular design of solvents for separation processes. *AIChE J.* **1994**, *40*, 1349−1360

(17) Naser, S. F.; Fournier, R. L. A system for the design of an optimum liquid−liquid extractant molecule. *Comput. Chem. Eng.* **1991**, *15* (6), 397−414.

(18) Odele, O.; Machietto, S. Computer aided molecular design: A novel method for optimal solvent selection. *Fluid Phase Equilib.* **1993**, *82*, 39−46.

(19) Buxton, A.; Livingston, A. G.; Pistikopoulos, E. N. Optimal design of solvent blends for environmental impact minimization. *AIChE J.* **1999**, *45*, 817−843.

(20) Sinha, M.; Achenie, L. E. K.; Ostrovsky, G. M. Environmentally benign solvent design by global optimization. *Comput. Chem. Eng.* **1999**, *23* (10), 1381−1394.

(21) Wang, Y.; Achenie, L. E. K. Computer aided solvent design for extractive fermentation. *Fluid Phase Equilib.* **2002**, *201* (1), 1−18.

(22) Duvedi, A. P.; Achenie, L. E. K. Designing environmentally safe refrigerants using mathematical programming. *Chem. Eng. Sci.* **1996**, *51* (15), 3727−3739.

(23) Churi, N.; Achenie, L. E. K. Novel mathematical programming model for computer aided molecular design. *Ind. Eng. Chem. Res.* **1996**, *35* (10), 3788−3794.

(24) Sahinidis, N. V.; Tawarmalani, M.; Yu, M. Design of alternative refrigerants via global optimization. *AIChE J.* **2003**, *49* (7), 1761−1775.

(25) Chavali, S.; Lin, B.; Miller, D. C.; Camarda, K. V. Environmentally-benign transition metal catalyst design using optimization techniques. *Comput. Chem. Eng.* **2004**, *28* (5), 605−611.

(26) Gani, R.; Neilsen, B.; Fredenslund, A. Group contribution approach to computer aided molecular design. *AIChE J.* **1991**, *37*, 1318.

(27) Joback, K. G.; Stephanopoulos, G. Designing molecules possessing desired physical property values. Proceedings FOCAPD 89, Snowmass, CO, 1989, pp 363−387.

(28) Gani, R.; Fredunslund, A. Computer aided molecular and mixture design with specified property constraints. *Fluid Phase Equilib.* **1993**, *82*, 39.

(29) Klein, J. A.; Wu, D. T.; Gani, R. Computer aided mixture design with specified property constraints. *Comput. Chem. Eng*. **1992**, *16*, S229.

(30) Marcoulaki, E. C.; Kokossis, A. C. On the development of novel chemicals using systematic optimization approach. Part I. Optimisation framework. *Chem. Eng. Sci*. **2000**, *55* (13), 2529−2546.

(31) Harper, P. M.; Gani, R.; Kolar, P.; Ishikawa, T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilib.* **1999**, *158*−*160*, 337−347.

(32) Viswanathan, J.; Grossman, I. E. A combined penalty function and outer approximation method for MINLP optimization. *Comput. Chem. Eng.* **1990**, *14* (7), 769−782.

(33) Achenie, L. E. K.; Sinha, M. Interval global optimization in solvent design. *Reliab. Comput.* **2003**, *9*, 317−338.

(34) Porn, R.; Harjunkoski, I.; Westerlund, T. Convexification of different classes of non-convex MINLP problems. *Comput. Chem. Eng.* **1999**, *23* (3), 439−448.

(35) Karunanithi, A. P. T.; Achenie, L. E. K.; Gani, R. Optimal (solvent) mixture design through a decomposition based CAMD methodology. *Proc. Eur. Symp. Comput.-Aided Process Eng.* **2004**, *14*, 217−222.

(36) ICAS documentation. Internal report, 2003. CAPEC, Department of Chemical Engineering, Technical University of Denmark, Lyngby, Denmark.

(37) Harper, P. M. A multi-phase multi level framework for computer aided molecular design. Ph.D. Thesis, Technical University of Denmark, Lyngby, Denmark, 2002.

(38) Bossen, B. S. Simulation and optimization of ammonia plants. Ph.D. Thesis, Technical University of Denmark, Lyngby, Denmark, 1995.

(39) Seader, J. D.; Henley, E. J. *Separation process principles*; Wiley: New York, 1998.

(40) Marrero, J.; Gani, R. Group contribution based estimation of octanol/water partition coefficient and aqueous solubility. *Ind. Eng. Chem. Res.* **2002**, *41* (25), 6623−6633.

(41) Bernard, G.; Hocine, R.; Lupis, C. H. P. Thermodynamic conditions for spinodal decomposition in a multicomponent system. *Trans. Metall. Soc. AIME* **1967**, *239*, 1600−1604.

(42) Martin, T. M.; Young, D. M. Prediction of the acute toxicity (96-h LC_{50}) of organic compounds to the fathead minnow (*Pimephales promelas*) using group contribution method. *Chem. Res. Toxicol*. **2001**, *14* (10), 1378−1385.

Received for review July 29, 2004
Revised manuscript received March 21, 2005
Accepted April 13, 2005

IE049328H