A Group Contribution Approach to Computer-Aided Molecular Design

Rafiqul Gani, Bjarne Nielsen, and Aage Fredenslund
Engineering Research Centre IVC-SEP, Chemical Engineering Dept.,
Technical University of Denmark, 2800 Lyngby, Denmark

A flexible and structured methodology for the computer-aided molecular design (CAMD) by the group contribution approach is presented. The proposed CAMD algorithm takes the following four steps: 1. preselect groups and target properties; 2. generate only a feasible set of compound structures in an optimal fashion; 3. predict properties for the screening of the set of feasible compound structures; and finally 4. select/design the compound. Since the success of any CAMD algorithm depends to a large extent on its ability to predict/compute the needed properties, the importance of the choice and applicability of these methods is considered along with the computational aspects related to the development of a computer program based on the proposed methodology. Finally, the scope of CAMD technology is highlighted using several practical examples.

Introduction

In the computer-aided molecular design by group contribution approaches, compounds or a mixture of compounds are represented as a collection of functional groups such that the resulting compound (or a mixture of compounds) has a set of properties (target properties) within specified ranges. In predicting properties by the group contribution approach, the property values are computed having the identities of the functional groups *SG* and the number of times they are present *NT* in a compound or a mixture. In the CAMD, the reverse problem is solved. Given the values for the target properties, such variables as *SG*, *NT* and *NG*, where *NG* is the dimension of the arrays of *SG* and *NT*, are determined so that the compounds or mixtures of compounds formed by them have the target properties. While the computation of property values in most cases yields a unique value, the reverse problem for the CAMD in most cases produces more than one solution. Thus, a screening is needed since only one of the alternatives must be chosen for the specified problem.

In principle, the scope of CAMD is significant since it can be applied to various types of problems, including innovation of separation processes, development of drugs, choice of refrigerants, finding solvents for polymers, and the design of solvent mixtures for the paint industries. In practice, however, the scope of CAMD is limited by the availability, reliability and accuracy of the models employed to predict the target properties. These properties are both pure-component properties and mixture properties.

In the last decade, significant advances have been made in the use of group contribution methods for the prediction of pure-component properties and mixture properties. See, for example, the published data by Reid et al. (1987), Joback and Reid (1987), Fredenslund et al. (1977), Larsen et al. (1986), Sander et al. (1986), and Dahl et al. (1990).

Contributions of CAMD have been reported by Gani and Brignole (1983), Brignole et al. (1986), Joback and Stephanopoulos (1989), and Macchietto et al. (1990). Gani and Brignole (1983) and Brignole et al. (1986) presented methods of solvent design related to liquid-liquid extraction. Joback and Stephanopoulos (1989) presented methods of solvent design and refrigerant design based on pure-component properties and ideal mixing rules for mixture properties. Macchietto et al. (1990) have coupled an optimization algorithm to CAMD and applied this to designing solvents for liquid-liquid extraction and gas absorption.

The formation of compounds by joining groups is usually carried out considering all the possible combinations of a preselected number of groups and then testing for their chemical feasibility. While both Brignole et al. (1986) and Joback and Stephanopoulos (1989) provided procedures to reduce the potential size of the combinatorial problem, this problem can

still become significantly large for any realistic application of CAMD.

Recently, Nielsen et al. (1990) have shown how one may overcome the difficulties related to the size of the combinatorial problem through a strategy that forms compounds in such a manner that only combinations considered are those resulting in the generation of chemically-feasible structures. This article expands and updates this strategy that can now be used in the automatic mode (generating all the possible feasible structures for a specified problem) or interactive mode (simultaneously adding groups to an intermediate "start" group and directing the final structure toward the feasible region with respect to the constraints). It also proposes the coupling of an optimization algorithm to the CAMD algorithm in a manner different from Macchietto et al.'s (1990). While Macchietto et al. (1990) employ optimization to form compounds that minimizes/maximizes a predefined objective function, we have decoupled the problem into a generation of feasible chemical compounds and preliminary screening part, and a final design/selection part. Only in the final design/selection part, do we propose to employ an optimization algorithm. The objective of this article is to present the group contribution approach to the CAMD technology and a CAMD methodology that can be used as the basis of developing more general and versatile methodologies. In addition, the CAMD technology is applied to several problems of practical interest to show its scope.

Development of the CAMD Methodology

Problem formulation

Mathematically, the reverse problem of properties prediction can be expressed as:

Determine

$$NG_k,\ SG(i)_k,\ NT(i)_k, \quad i=1,\ldots,NG \quad k=1,\ldots,NC \tag{1}$$

subject to

$$P_l(j)<P(j)<P_u(j) \qquad j=1,\ldots,NP \tag{2}$$

and

$$C1(m)=s \qquad m=1,\ldots,MC1 \tag{3}$$

$$C2(n)=s \qquad n=1,\ldots,MC2. \tag{4}$$

where $SG(i)_k$ is the group identification number for group i present in component k and belongs to the preselected set of groups $SGT(j)$, $j=1,\ldots,NGT$. $NT(i)_k$ is the number of times group $SG(i)_k$ appears in the structure of compound k. NG_k is the number of different types of groups present in the structure of compound k.

Equation 2 represents the constraints with respect to the target properties $P(j)$, $j=1,\ldots,NP$. NP is the total number of target properties. Functions $C1(m)$, $m=1,\ldots,MC1$, represent the rules of the formation of chemically-feasible compounds. Functions $C2(n)$ represent the rules of the preliminary screening of compounds based on the ability to predict the target properties and the final selection of compounds. $MC1$ and $MC2$ are the sizes of the vectors $C1$ and $C2$, respectively. In Eqs. 3 and 4, s signifies a specified value. NC is the total number of designed compounds that satisfy the constraints (Eqs. 2-4). This mathematical formulation is explained with the following simple example. Let us design compounds that are miscible with water, have a boiling point of 351.5 K and a molecular weight smaller than 100, and the compound must not be a chloride.

Let us consider the following five functional groups, CH3, CH2, OH, COOH, and CH2Cl, with group identification numbers, 1, 2, 10, 19, and 30, respectively. The values for the different variables for this problem are: SGT(array) = 1, 2, 10, 19, 30; $NGT=5$; $P(1)=351.5$ K; $P(2)=100$; $C1(1)=0$ (total number of free attachments in the final structure); $C1(2)=0$ (number of times group 30 is allowed to appear in a compound); and $C2(1)=0$ (the number of unavailable group parameters in a compound). One of the compounds that satisfies all the constraints is ethanol, representing the groups CH3, CH2, and OH, for which $NG=3$; NT(array) = 1, 1, 1; SG(array) = 1, 2, 10.

The process of designing a set of chemically-feasible compounds that satisfies the property constraints takes the following four steps:

Stage 1. Preselect the array elements for SGT, P, $C1$, and $C2$, as well as the values for P_l, P_u, NGT, NP, $MC1$, and $MC2$. A knowledge-based system may be employed.

Stage 2. Form chemically-feasible compounds ensuring that for the compounds designed, the relevant group parameters for the methods of prediction (pure-component and/or mixture properties) are available. At this stage, NG_k, $NC1$, and all the elements of the arrays SG_k and NG_k are determined based on the constraints of Eq. 3. $NC1$ is the total number of chemically-feasible compounds before the compounds are tested for their property values. A knowledge-based system may be employed, for example, to reject the structures whose properties cannot be presently predicted.

Stage 3. Predict the target properties by appropriate prediction methods and then retain only those compounds that satisfy the property constraints of Eq. 2. Accordingly, the number of compounds is reduced from $NC1$ to $NC2$.

Stage 4. Perform process simulation to evaluate the performance of the designed compounds. The final design/selection can be made either by ordering the compounds from stage 3, in terms of predefined performance indices, after a final screening of compounds based on constraints of Eq. 4 or by using an optimization algorithm to minimize/maximize a predefined objective function.

Stage 1: classification and preselection of target properties and groups

Different problems have different sets of properties as constraints. Thus, while for solvent design such properties as boiling point, molecular weight, density, solubility, and selectivity are important, for refrigerant design such pure-component properties as vapor pressure and heat of vaporization are important.

Specification of the problem type identifies the corresponding target properties. A knowledge-based system can, therefore, determine a set of target properties to define the constraints of Eq. 2 for various problem types belonging to the allowable set of problems that we can presently solve through CAMD. Since all the target properties may not be

Table 1. Classification of Groups*

Class	Category 1	2	3	4	5
0	CH3OH CH3SH (CH2OH)2 NMP Diethyl Glycol 2-propanol CCL2F2	CH3NO2 CH3CN CH2CL2 CH3NH2 CCL3F C4H4S	H2O Furfural CHCL3 TCE Pyridine CHCL2F Morpholine	CH3NH2 HCOOH ACRY MFA 1-propanol CHCLF2	CF3 DMSO DMF TMS CS2 CCLF3
1	CH3	CH2CN CH2NO2 CH2NH2	CH3CO CONH2 CONHCH3 CON(CH3)2	OH CHO COOH CH2CL I, Br F, CL CH3COO CH3O C2H5O2 CH3S	CCL2F CH2SH CH3NH CHCL2 C4H3S SH C≡CH COO CCL3 SiH3 CH2NH2 CCL2F CHCLF
2	CH2	CHNO2	CH2CO CH2COO CH2O CONCH3CH2	CHNH2 CH2NH CHCL CONHCH2 C2H4O2 CH2S	CH=CH CH2=C CH4H2S CH3N C≡C SiH2
3	CH		CON(CH2)2	CHNH CH2N CCL CH−O CHS	CH=C CCL2 SiH SIH2O SiHO
4	C				C=C SiO Si
5	ACH		ACCH2 ACCH AC	ACCH3	ACOH ACNH2 ACCL ACNO2

*For class zero groups, the categories do not apply.

computable, it is convenient to classify them as explicit target properties and implicit target properties. For explicit properties prediction methods are available. For implicit properties prediction methods are not presently available and thus a combination of experience, information from the open literature and experiments are needed to determine their values.

Tables 1 and 2 show functional groups and properties for which prediction of molecular properties can be made. The groups in Table 1 are based on the UNIFAC (Fredenslund et al., 1977) functional group definition. Groups and target properties are preselected according to these tables. The groups in Table 1 are divided into classes and categories. The class of a group with the exception of class 5 signifies the number of free attachments available to the group. The category of a class of groups signifies the degree of restriction they have with respect to their joining with other groups.

Groups of class 0 without any free attachments are molecules and thus the category is to be neglected. Groups of category 1 of classes 1 to 4 have no restrictions except those imposed by the total number of groups allowed in a molecular structure. For example, structures of four groups cannot have any groups from class 4. Groups of categories 2, 3, 4 and 5 have varying degrees of restrictions. Table 3 shows the restrictions imposed on groups of categories 3, 4, and 5 of classes 1 to 5. These restrictions are related to the chemical feasibility, stability and limitations of the available methods for the prediction of properties. Groups of category 2 of classes 1 to 4 are considered as special groups and can appear more than once, but cannot be connected to each other or with another group from the same or higher category. More than three groups from categories 2 through 5 cannot join with the same class 4 group. The class 5 groups belong to those functional groups, which when joined form aromatic compounds. Category 3 groups of class 5 provide the possibilities of joining with the groups from other classes; for example, ethyl benzene is formed by 5ACH, ACCH2 and CH3 groups. It is beyond the scope of this article to explain in detail how Table 1 was prepared. The above information, however, describes the CAMD methodology suf-

Table 2. List of Properties*

Property	Property Type	Function	Thermodynamic Model
Mixture Properties			
Solvent Power	secondary 1	$f(x,\gamma)$	*GC*
Selectivity, S_s	secondary 1	$f(\gamma^\infty)$	*GC*
Selectivity, S_f	secondary 1	$f(x^f, \gamma^\infty)$	*GC*
Solubility (Polymer)	secondary 1	$f(x, \gamma)$	*GC*
Solubility (Electrolytes)	secondary 1	$f(x, \gamma)$	*GC*
Separation Factor	secondary 1	$f(x, \theta, \pi, \underline{\gamma}, \underline{\varphi})$	*GC*
Miscibility	secondary 1	$f(x, \theta, \pi)$	*GC*
Activity Coefficient	primary 1	$f(x, \theta, \pi, G)$	*GC*
Fugacity Coefficient	primary 1	$f(x, \theta, \pi, G)$	*GC*
Boiling Point	secondary 1	$f(x, \theta, \pi, \underline{\gamma}, \underline{\varphi})$	*GC*
Flash Point	secondary 1	$f(x, \theta, \pi, \underline{\gamma})$	*GC*
Viscosity	primary 1	$f(x, \theta, \underline{\gamma}, G)$	*GC*
Surface Tension	primary 1	$f(x, \theta, \underline{\gamma}, G)$	*GC*
Density	primary 1	$f(x, \theta, \pi, G)$	*GC*, correlation
Heat of Vaporization	secondary 1	$f(x, \theta, \pi)$	*GC*, correlation
Toxicity Measure	secondary 2	$f(x, \theta, \gamma)$	*GC*
Pure Component Properties			
Critical Properties	primary 1	$f(G)$	correlation
Molar Volume	primary 1	$f(G)$	correlation
	secondary 1	$f(P_{crit}, p_s, \omega)$	correlation
Viscosity	primary 1	$f(\theta, G)$	correlation
Surface Tension	primary 1	$f(\theta, G)$	correlation
Vapor Pressure	primary 1	$f(G)$	correlation
	secondary 1	$f(T_c, \omega)$	correlation
Boiling Point	primary 1	$f(G)$	correlation
Solubility Parameter	primary 1	$f(G)$	correlation
	secondary 1	$f(\theta, h_{vap})$	correlation
Heat of Vaporization	secondary 1	$f(t_b)$	correlation
Heat Capacity	primary 1	$f(G)$	correlation
Heat of Formation	primary 1	$f(G)$	correlation
Corrosivity	secondary 2		experiment
Odor	secondary 2		experiment

*GC = group contribution based methods
G = group parameters
1 = explicit property
2 = implicit property

ficiently. In general, lower category signifies less restrictions for groups of categories 3, 4 and 5.

Since the functional groups of Table 1 are based on the UNIFAC (Fredenslund et al., 1977) definition of functional groups, they can be used directly by the UNIFAC-based methods for computation of mixture properties. For computation of pure-component properties, new parameter tables for the groups of Table 1 have been developed using the information given in the various group contribution methods and their corresponding parameter tables (Joback and Reid, 1987; Reid et al., 1987). The alternative approach of using the groups defined by a pure-component method to form the compounds requires estimation of UNIFAC group interaction parameters for computation of the mixture properties.

Preselection of groups for a specified problem helps reduce the size of the problem in stage 2. There are several ways of reducing or preselecting the array elements of the vector of functional groups *SGT*. The following two selection criteria are used in this work:

- Search for certain types or classes of compounds, and therefore the corresponding groups is excluded. For example, excluding halides from the search means eliminating all halide groups from *SGT*.
- It is designed in several steps, designing in each step only certain types of molecular structures. For example, first design all straight-chained molecules and repeat for all aromatic compounds. For each design step, a relatively smaller number of groups is selected. The design problem can thus be divided into several smaller subproblems.

Identification of the implicit target properties often helps eliminate some groups. For example, according to chemists (Rosendahl-Jensen, 1990), for refrigerant design the compounds should not have double or a higher number of bonds, since it is highly possible for them to polymerize in a relatively short period of time. Solvent design for separation of food products should avoid aromatic compounds for reasons of health.

Stage 2: formation of feasible chemical compounds

An efficient, structured algorithm for joining of groups to form feasible chemical compounds has been developed by considering the classification of groups from stage 1. The algorithm generates for a preselected set of groups all the chemically feasible compounds for which the target properties can be predicted.

For any specified problem, we start with a preselection of groups represented by the array *SGT*. The molecular structures to be formed must satisfy the following primary conditions:

- The final molecular structure cannot have any free attachments.

Table 3. Rules Related to the Primary Conditions

Total No. of Groups	Largest Class of Group	No. of Groups from Largest Class	Max. No. of Groups Allowed from Sum of Categories				
			3	4	5	3+4+5	4+5
For Nonaromatic Compounds							
2	1	2	2	1	1	2	1
3	2	1	2	1	1	2	1
4	3	1	2	1	1	2	1
4	2	2	2	2	1	2	2
5	4	1	2	1	1	2	1
5	3	1	2	2	1	2	2
5	2	3	3	2	1	3	2
6	4	1	3	2	1	3	2
6	3	2	3	2	1	3	2
6	3	1	3	2	1	3	2
6	2	4	3	3	1	3	3
7	4	1	3	2	1	3	2
7	3	2	3	3	1	3	3
7	3	1	3	3	1	3	3
7	2	5	4	3	1	4	3
8	4	2	3	2	1	3	2
8	4	1	3	3	1	3	3
8	3	3	3	2	1	3	2
8	3	2	3	3	1	3	3
8	3	1	3	3	1	3	3
8	2	6	4	3	1	4	3
For Aromatic Compounds							
6	5	6	0	3	1	3	3
7	5	5	1(1)	2	1	3	2
8	5	5	2(1)	2	1	3	2
8	5	5	1(2)	2	1	3	2
9	5	5	3(1)	0	0	3	0
9	5	5	1(3)	2	1	3	2
9	5	5	1(1)+1(2)	1	1	3	1

*Numbers in parenthesis indicate the number of free attachments of the group.

- Groups of categories 2 and higher have restrictions on their joining with other groups that must be fulfilled.

In addition, a molecular structure must satisfy the following secondary conditions:

- All groups appearing in a molecular structure must have the necessary parameters for the methods to be employed for the prediction of target properties.
- Structures with too many branches are to be avoided.
- Structures that may exhibit proximity effects will be rejected.

From the above conditions, it can be observed that the primary conditions will always be applied and ensure chemical feasibility, while the secondary conditions are temporary and problem-dependent and are needed because of the present limitations of the different prediction methods.

Let us consider the following simple example, considering seven groups: CH3, CH2, CH, CH2NO2, CH3CO, OH, and CH2=CH. We would like to form the compounds of only four groups. Without any restrictions on how these groups can join, 210 structures are generated. Adding only the restriction that the total number of free attachments must be zero in the final structure, only 40 structures are obtained. Note that some of these 40 structures may not be chemically-feasible or stable. Adding restrictions on the categories of groups (see Table 3) reduces the number of structures to nine.

The minimum and maximum numbers of groups that can appear in the molecular structures for a specified problem are governed by the type of structure, limits of molecular weight and physical insight. For example, for only straight-chained structures, the minimum number of groups is two. For branched structures, the minimum number of groups is four, and for aromatic compounds the minimum number of groups is six. The upper limit of the maximum number of groups may be governed by the maximum allowable molecular weight.

The following step-by-step algorithm for the joining of groups to form feasible chemical compounds was made having a list of: preselected groups together with their classification information; target properties; and prediction methods to be employed.

Step 1. Select a type of molecular structure and determine the minimum (MA) and maximum number (MB) of groups than can appear in any compound and set $MC=1$.

Step 2. For nonaromatic compounds, choose a group from category MC and class equal to $M2=MB-1$. If $M2>4$, then choose any group from class 4 and category MC or 5, and set $M2=4$. For aromatic compounds, choose any group from class 5 and category MC or 5. For either case, if more than one group exists for the specified class and category, determine a separate structure for each case.

Step 3. Determine $M3=MB-M2$. If $M3=1$, terminate the molecular structures by choosing (considering the restrictions given in Table 3) $M2$ groups of class 1 and any category that

satisfy the primary and secondary conditions. Repeat until all structures formed in step 2 are considered. On termination, go to step 6. If $M3 > 1$, go to step 4.

Step 4. Choose (considering the restrictions given in Table 3) any group from class $M4 = MB - M2$ and any category that satisfies the primary and secondary conditions. If $M4 \geq 4$, then $M4 = 4$. Repeat this step for all allowable categories within class $= M4$. Repeat until all structures from step 3 have been considered.

Step 5. Determine $M5 = MB - M2 + 1$. If $M5 = 1$, go to step 3. Otherwise, for each structure from step 4, choose (considering the restrictions given in Table 3) any group from class $M5 = MB - M2 + 1$, set $M2 = M2 + 1$, and repeat step 4.

Step 6. Set $MC = MC + 1$. If $MC \geq 5$, go to step 7. Otherwise, go to step 2.

Step 7. Set $MB = MB - 1$ and repeat step 2 until $MB = MA$. If $MB < MA$, select a new type molecular structure or compound type and repeat step 1.

An example of the application of this algorithm is demonstrated in the Appendix.

This algorithm has the following important aspects.

- We bypass totally the "generate and test paradigm" (that is, first generate structures through all possible combinations of the preselected groups and then test for their chemical feasibility).
- In step 7, the type of structures includes aromatic, straight-chained compounds, and the class of compounds includes hydrocarbons and alcohols.
- By searching for a particular class of compounds of each design subproblem, the size of the problem is kept small. A manageable number of compounds is generated by solving the design subproblem several times for different types of structures or classes of compounds.

It, however, has its limitations. At present, the algorithm does not form cyclic compounds or isomers. Also, some structures are rejected because of doubtful stability or because group parameters are not available. Some structures are also rejected because the application range of the methods is limited for prediction of the target properties. Using Table 3, for nonaromatic compounds, structures with a minimum of two groups and a maximum of eight groups can be formed. For aromatic compounds, structures with a minimum of six groups and a maximum of nine groups can be formed.

Stage 3: prediction of target properties and screening of compounds

Once the chemically-feasible molecular structures are formed, they can be tested for the constraints on the target properties. At this stage, therefore, the properties prediction methods are needed, and only the explicit target properties are considered. The explicit target properties are of two types: pure-component properties and mixture properties. Each type has both primary and secondary properties. Primary properties can be determined directly from a group contribution method. Secondary properties are not determined by group contribution methods, but are determined by methods that use some of the primary properties as specified variable. For example, properties like the critical temperature and critical pressure are classified as primary pure-component properties. Properties like the vapor pressures p_s, heat of vaporization h_{vap}, and boiling points t_b are classified as secondary properties. Joback and Stephanopoulos (1989) suggested the use of a two-parameter group contribution method for secondary pure-component properties like p_s, h_{vap}, and t_b. For accurate prediction, we decided this method is not appropriate. For mixtures, properties such as activity coefficients γ, fugacity coefficients φ, and compressibility factors z are classified as primary mixture properties, while properties like solubility S_p, selectivity S_s, partial miscibility M_p, and boiling point t_b are classified as secondary mixture properties. The list of properties given in Table 2 also gives information on the property type and classification, and the methods employed in this article for their prediction. It can be observed that these properties cover a wide range of potential CAMD problems.

For the screening of the chemically-feasible compounds from stage 2, the explicit target properties are ordered: Primary pure-component properties, secondary pure-component properties, primary mixture properties, and secondary mixture properties. This ordering of target properties follows the principle of computing first, the properties that are easy to calculate. Also, some of the secondary properties cannot be determined before the evaluation of some other properties. According to this ordering of prediction of properties, molecular weight m_w is usually the first property to be determined. The compounds are screened after each property has been determined. Thus, for the compounds for which constraints on the pure-component properties are not satisfied, the mixture properties related to them need not be computed.

The application range and success of any CAMD methodology depends very much on the accuracy and applicability of the prediction methods of the explicit target properties. The applicability of almost all the methods listed in Table 2 is limited by variables such as temperature θ, pressure π, and/or groups SG. Extrapolation to compute properties outside the range of application of some of these methods can, in many instances, lead to significant errors (see computational aspects section). Therefore, proper selection of methods is important, especially when extrapolations from the recommended application range of the prediction method are needed.

Computation of some of the secondary mixture properties can be simplified (or linearized) by employing ideal mixing rules such as

$$P_i = \sum x_k p_k \tag{5}$$

where P_i is any secondary mixture property i, x_k is the mole fraction of component k in the mixture, and p_k is the corresponding pure-component property of component k. These mixing rules should be used with caution especially in problems related to solvent design. Examples of the deviation from ideality for properties such as viscosity and boiling point are shown in Figures 1a and 1b. Visser (1988) and Wu and Klein (1990) have also shown the deviation from ideality for properties such as flash points, surface tension, viscosity, and evaporation rate. Such equations as Eq. 5 should be used when no other alternatives are available. For prediction of some of the secondary properties needed for solvent design, we propose the following simple rules.

Determination of binary azeotropic points.

$$\gamma_1^\infty < p_{v1}/p_{v2} < 1/\gamma_2^\infty \tag{6}$$

$$\gamma_1^\infty > p_{v1}/p_{v2} > 1/\gamma_2^\infty \tag{7}$$

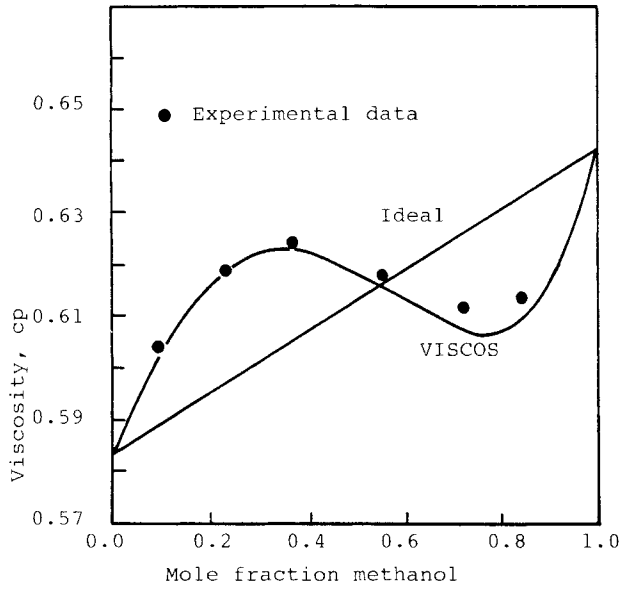

Figure 1a. Comparison of calculated viscosities for methanol-benzene at 298 K by VISCOS (Sørensen et al., 1990a) and ideal mixing rule with experimental data (Helme, 1989).

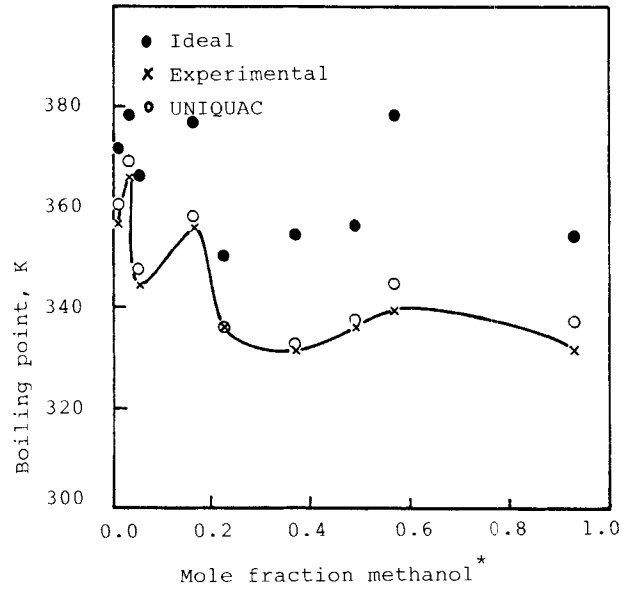

Figure 1b. Comparison of calculated boiling points at 0.1 MPa for methanol-benzene-toluene by UNIQUAC and ideal mixing rule with experimental data (Gmehling et al., 1982).

*indicates toluene free mole fractions.

If for a binary mixture Eq. 6 is satisfied, a minimum boiling azeotrope will be formed. If Eq. 7 is satisfied, a maximum boiling azeotrope will be formed. The solvent being designed should not form an azeotrope with the solute (compound to be recovered). This condition can be ignored in some cases.

Determination of partial miscibility. A binary mixture will show partial miscibility, if the following condition is not satisfied,

$$d\ln\gamma_1/dx_1 > -1/x_1. \tag{8}$$

Location of solubility limits. The existence of the upper critical solution temperatures (UCST) or the lower critical solution temperature (LCST) for a mixture of the solvent and any of the solutes can be determined by the following tests (derived from Rowlinson and Swinton, 1982): if H^E = positive and $S^E > -R/2$, UCST exists; if H^E = negative, S^E = negative, and $S^E < -R/2$, LCST exists where

$$H^E/RT^2 = -\sum x_i d\ln\gamma_i/dT \tag{9}$$

$$S^E = -T\sum x_i d\ln\gamma_i/dT - \sum x_i \ln\gamma_i \tag{10}$$

The existence and location of these temperatures are important, because they indicate the measure of solubility (or solvent power) as a function of temperature. For a symmetric mixture with LCST, the maximum solvent power is obtained when $\ln\gamma^\alpha$ is less than 7.39. The maximum solvent power for a symmetric mixture with UCST is obtained when $\ln\gamma^\alpha$ is greater than 1.6. A symmetric mixture has the same solubility at x_1 and $1-x_1$ (binary mixture).

A list of chemically-feasible compounds, which have also satisfied the constraints with respect to most of the explicit target properties, is obtained at this stage. The remaining explicit and implicit target properties are considered in stage 4.

Stage 4: final design/selection

The remaining set of explicit target properties are computed for the compounds obtained from stage 3. Usually, evaluation of these properties require process simulation. Process simulation may include simulation of an entire process flowsheet, part of a process flowsheet or a single unit of a process flowsheet. Again, these properties are ordered in terms of the complexity of the problems, the computational times and the availability of information. In this article, only the single-unit computations are considered.

For solvent design problems (involving LLE, VLE, GLE, or SLE), for example, separation factors, solvent loss, product yields, and so on are calculated by single-unit (usually flash) computations. The design variables, usually θ, π, and feed compositions x^f, are obtained through computations of phase diagrams. For problems related to paint additives (minimum cost solvent mixtures) and polymer blends, computations of miscibility ranges are needed. For the refrigerant design, the effects of expansion and compression need to be evaluated.

For the final design/selection, two methods are proposed: one which rates the compounds according to a performance index, the other which determines the best compound (or mixture of compounds) by solving an optimization problem.

Method 1. After screening the alternatives (compounds from stage 3) based on the remaining explicit target properties, the compounds are rated in terms of a performance index defined as

$$PI = f(\underline{P}) \tag{11}$$

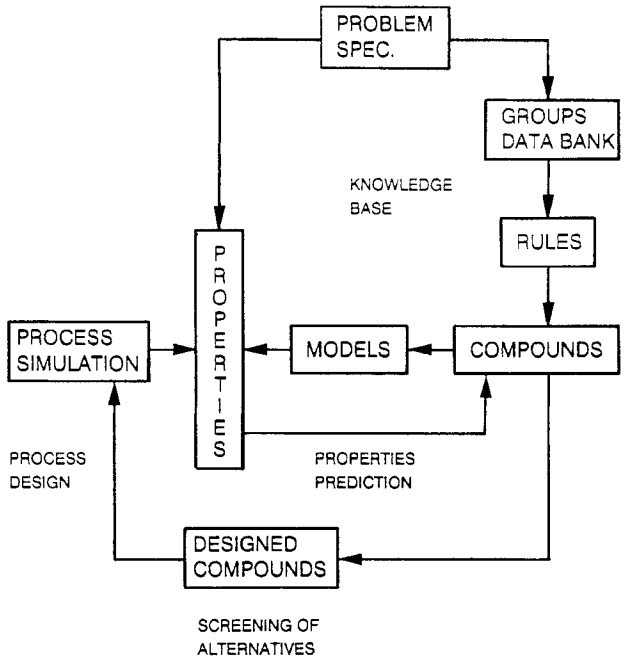

Figure 2. Flow diagram showing the structure of the CAMD methodology.

where PI is a function of the explicit target properties. This function differs for different CAMD problems. The following expression has been used for solvent design problems:

$$PI = S_f \cdot S_F / m_w. \tag{12}$$

For refrigerant design, the following expression may be used:

$$PI = h_v^s / c_p^s \tag{13}$$

where superscripts indicate scaled values.

The compounds with the best ratings can be selected for detailed future studies. These should include process design aspects as well as evaluation of the remaining implicit target properties.

Method 2. A predefined objective function (which could also be Eq. 11) is maximized or minimized such that the remaining explicit target property constraints are satisfied. The optimal solution may result in a pure compound or a mixture.

Computational Aspects

The flow diagram for the proposed group contribution CAMD methodology is shown in Figure 2. A computer program incorporating this methodology needs to interface with knowledge bases, routines for computation of the explicit target properties, routines for process simulation, and the routines for the formation of feasible chemical compounds. The knowledge bases consist of information on models, groups, properties, and different CAMD problems. Since different group contribution methods have their own group parameter tables, a model "unification" feature (Gani et al., 1987) increases the scope of the program for CAMD by adapting the parameters from one model to another. For a consistent design, the same group representations must be used in all stages for a compound. This requires the use of a group-model master table storing information on the availability of groups and their parameters for the different models.

Knowledge base

The knowledge base includes information on the classification of groups as given in Table 1, the groups related to a class of compounds or a type of molecular structures, the list of explicit and implicit target properties for different CAMD problems, the availability of group/model parameters, and the group/parameter tables for the different methods for the computation of the target properties. The knowledge base also contains the rules that govern the preselection of groups and target properties in stage 1, the formation of feasible compounds in stage 2, the ordering of properties and the transfer of the needed group/model parameters in stages 3 and 4, and the rating of the final design alternatives. Different aspects of the knowledge base are, therefore, needed during different stages of CAMD. For example, the information on group classification, target properties, and group availability are needed in stages 1 and 2, while in stages 3 and 4 parts of the group/model parameter tables together with details of the representation of compounds by groups are needed. Information on the presentation of compounds by groups are generated in stage 2 and added to the knowledge base for use in stages 3 and 4.

The availability of group parameters for the computation of the target properties is determined through the indices of the group-model master table. Identification of any group from Table 1, therefore, also determines the models that contain the corresponding parameters for this group.

Computation of target properties and process simulation

Computer programs are available (Sørensen et al., 1990a) for most of the models in Table 2. A word of caution should be added on the accuracy of these models/methods. Table 4 shows the calculated primary and secondary pure-component properties for a selection of compounds, together with those obtained through the DIPPR tables (Daubert and Danner, 1986). Table 4 shows the possibilities of significant errors for the primary and secondary pure-component properties and the dangers related to extrapolation of the methods for computing secondary pure-component properties. In addition, for the mixture properties, one has to be aware of the need to use the appropriate mixing rules and interaction parameters. One way of avoiding these errors could be to use a databank with pure-component properties [for example, the DIPPR tables (Daubert and Danner, 1986)]. The prediction methods will be used only for those compounds that are not present in the databank.

For the computation of primary pure-component properties, usually only group parameters are needed. For the computation of the secondary pure-component properties and some primary pure-component properties, some model parameters are needed. For the computation of the primary mixture properties, pure-component properties, group parameters, and group interaction parameters are needed. For the secondary mixture properties, only the primary mixture properties are needed usually.

For the computation of explicit target properties in stage 4 and for process simulation, the IVC-SEP program package

Table 4. Computed Pure-Component Properties

Compound	Method*	Properties T_c K	P_c MPa	t_b at 0.1 MPa K	h_v at 272 K kJ/mol	p_s at 272 K MPa
Ethane	DIPPR	305.4	4.88	184.6	9.2	2.320
	Meth1	296.3	5.09	245.2	18.7	0.952
	Meth2	—	—	—	18.7	0.273
Methyl Chloride	DIPPR	416.3	6.68	248.9	20.4	0.249
	Meth1	415.9	5.55	259.7	22.3	0.161
	Meth2	—	—	—	21.6	0.161
Methyl Acetylene	DIPPR	402.4	5.62	249.9	20.9	0.245
	Meth1	417.8	5.49	258.1	21.7	0.169
	Meth2	—	—	—	20.9	0.169
Propadiene	DIPPR	393.3	5.66	238.6	18.8	0.342
	Meth1	334.0	5.31	260.5	20.85	0.217
	Meth2	—	—	—	20.85	0.155
Ethyl Fluoride	DIPPR	375.3	5.02	235.4	18.15	0.415
	Meth1	384.4	4.78	244.4	19.7	0.311
CHCLF2	DIPPR	369.3	4.97	232.3	17.69	0.477
	Meth1	369.3	5.03	232.3	17.74	0.472

*DIPPR = values calculated from the DIPPR data bank (Daubert and Danner, 1986)
Meth 1 = method used in this work
Meth 2 = values reported by Joback and Stephanopoulos (1989)

(Sørensen et al., 1990a) was used. They include programs for phase diagrams, single-unit (flash, distillation, extraction, etc.) simulation and process simulators, SEPSIM (Andersen et al., 1985), and DYNSIM (Sørensen et al., 1990b).

Structure of the CAMD program

The structure of the CAMD program is shown in Figure 3. The knowledge base is divided into several sections for the different needs of the various stages of the CAMD methodology. A central administrator routine keeps track of the current stage and corresponding needs. It also calls on the different routines for the computation of properties and formation of compounds. This structure makes the future updating of the parameter tables, addition of groups, models and parameters very simple. For example, addition of new groups means simply increasing the number of rows of the group-model master table and the corresponding group parameter tables. Addition of models means simply increasing the number of columns of the group-model master table and adding the new group parameter table(s) together with the routine for the computation of properties.

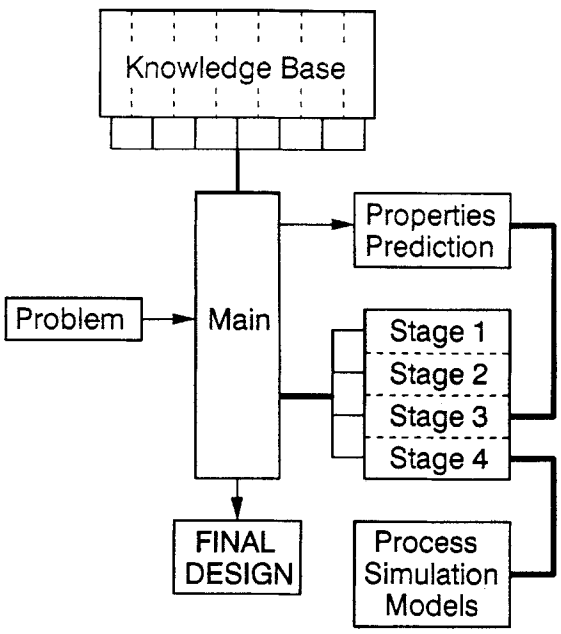

Figure 3. Structure of the CAMD computer program.

Automatic Design and Interactive Design. Automatic and interactive design modes depend on how the algorithms of stages 2 and 3 are employed. During the formation of feasible compounds in stage 2, at each step more than one group may be available for addition to the molecular structure being generated. In the automatic mode all the available options will be considered, while in the interactive mode only one option will be considered. For the selection from the available groups in each step of the interactive design mode, derivative data on the groups and target properties are employed (Macchietto et al., 1990). The proposed interactive design, however, does not need to employ optimization algorithms, since with each new addition of groups the target properties of the compound proceed toward the feasible region. In this way, the procedure is similar to that of Joback and Stephanopoulos (1989). The important difference, however, is that by employing a two-parameter correlation for the target properties, Joback and Stephanopoulos (1989) can solve the problem rapidly and graphically. The cost for this, however, is inaccurate prediction of properties and limited application, since for many problems such two-parameter representation has not proven feasible.

A computer program with the same structure as shown in Figure 3 is being developed. The current version employs Ta-

Table 5a. Target Property Constraints for the Refrigerant Design Problem

Property		Value
1. p_s	at 272.05 K	$\geq$ 1.4 MPa
2. p_s	at 316.45 K	$\leq$ 15 MPa
3. h_v	at 272.05 K	$\geq$ 18.4 kJ/mol
4. c_p	at 294.15 K	$\leq$ 134.72 kJ/(kmol·K)

Table 5b. Preselected Groups for the Refrigerant Design Problem

CH3, CH2, CH, C
CH3NH, CH2NH, CHNH
CH3N, CH2N
F, CL, SH
CH2CL, CHCL, CCL, CHCL2, CCL2, CCL2F, CHCLF, CCLF2
Molecular Groups: CH3CL, CCL3F, CHCL2F, CHCLF2, CCLF3, CCL2F2

Table 5c. Refrigerant Design Alternatives

	Properties			
Compound	h_v at 272 K kJ/mol	p_s at 272 K MPa	p_s at 316 K MPa	c_p at 294 K kJ/(kmol·K)
Ethyl Fluoride	19.7	0.311	1.185	106.4
SHF	20.7	0.148	5.512	77.9
Methyl Chloride	20.5	0.255	9.844	47.2
CCl2F2	18.5	0.305	1.080	117.0

bles 1 and 3, and allows the computation of most of the properties in Table 2. Most of the programming related to stages 2 and 3 has been completed including the algorithm for formation of feasible chemical compounds and preliminary screening of compounds based on primary and secondary properties. The first version of the knowledge base related to the availability of group parameters, group master table, group classification, target properties has also been developed. Programming related to stage 4 has been restricted to the calculation of *PI*, final design/selection of compounds based on *PI*, and incorporation of single-unit operation computations (for example, flash). Also, in the present version, it is possible to add user-specified compounds/groups and to use any or all stages of the CAMD methodology.

Application Examples

The results from the application examples should be regarded only as illustrations of the scope of the CAMD technology, rather than final solutions for these problems. The results, however, may be used as a basis for further detailed studies. The application examples are grouped into two types of problems: those for which only pure-component properties are considered and those for which pure-component and mixture properties are considered.

Problems based on pure-component properties

Refrigerant design and polymer design are typical examples of problems of this type. Joback and Stephanopoulos (1989) have reported a typical refrigerant design problem. We attempt here to solve the same problem with additional constraints. The design specifications for this problem are given in Table 5a.

All the explicit target properties for this problem are in the CAMD methodology considered as secondary pure-component properties. In addition, we have included some implicit target properties. For example, the refrigerant should not polymerize over a very long period of time; it should not be explosive nor corrosive.

According to Rosendahl Jensen (1990), low boiling compounds that polymerize over a period of use should not have double or triple bonds in their structure. Also, low-boiling compounds containing halides and nitrogen have strong tendencies to explode. Thus, with the help of our knowledge base on properties, we are able to preselect the groups from Table 1. These are given in Table 5b. Table 5c lists the designed compounds as well as the computed explicit properties. Only two of the compounds listed by Joback and Stephanopoulos (1989) are included in this table. Other compounds were rejected by our CAMD program, because of their real property values (see Table 4), because of the existence of double and triple bonds, or because of the presence of halides and nitrogen in the same molecular structure. It should be noted that better alternatives can be generated by adding more groups and/or properties. The diagnostic information from the CAMD program is given in Table 6. The reported CPU time refers to running the program on a HP9000 minicomputer.

Examples based on pure-component and mixture properties

Most CAMD problems need a combination of pure-component and mixture properties. Typical examples of problems of this type involve the design of solvents for different separation processes (for example, liquid-liquid extraction, azeotropic distillation, and gas absorption), solvent mixtures for paints, polymer blends, and many more.

Design of Solvents for Liquid-Liquid Extraction. Solvents have been designed for liquid-liquid extraction of ethanol from ethanol-water mixtures and acetic acid from acetic acid-water mixtures. Table 7 lists the target properties for these problems, and Table 8 lists the final design alternatives along with some of their computed properties. Note that the compounds in Table 8 are not the best solvents, rather a selection representing the list of compounds reported by the CAMD program. These compounds are taken from different parts of the ordered list to demonstrate the variation of *PI* with compound type. For the ethanol-water and acetic acid-water problems, interesting comparisons can be made with the reported results by other workers (Brignole et al., 1985; Joback and Stephanopoulos, 1989; Macchietto et al., 1990). While Brignole et al. and Macchietto et al. have used the UNIFAC(LLE) model (Magnussen et al., 1981) for the prediction of the needed secondary mixture properties, Joback and Stephanopoulos (1989) have employed solubility parameters (Lo et al., 1983; Barton, 1983) and ideal mixing rules.

It is noted that some of the compounds used in this work (see Table 8) have also been reported by Macchietto et al. (1990). The compounds reported by Brignole et al. (1985) and Joback and Stephanopoulos (1989), however, were not found

Table 6. Diagnostic Information from the CAMD Program

Variable	Refrigerant Design Problem	Solvent Design Problems: Ethanol-Water LLE	Ethanol-Water VLE*	Acetic Acid-Water LLE	Acetic Acid-Water VLE*	Styrene VLE
Number of Structures Rejected Because of						
m_w	—	0	0	2	0	0
t_b	—	91	21	214	25	31
S_s	—	1,303	84	3,492	0	635
S_p	—	0	0	0	0	0
S_f	—	99	0	0	0	0
S_F	—	0	4	0	14	0
M_p	—	115	453	20	1,539	1,042
p_s at 272 K	4	—	—	—	—	—
p_s at 316 K	—	—	—	—	—	—
h_v	1,389	—	—	—	—	—
c_p	1	—	—	—	—	—
Parameters**	0	0	0	1,741	0	7,751
FS	1,398	1,667	595	5,535	1,632	10,462
NC	4	59	33	33	52	43
NGT	26	16	12	18	16	20
CPU(s)	1.12	6.98	1.82	2.92	2.06	3.12

*Separation factor computed at 380 K and 0.1 MPa.
**Structures rejected because group parameters are not available.
FS = total number of chemically feasible structures formed
NC = total number of compounds satisfying all target property constraints
NGT = number of preselected groups

by our program for CAMD. The reason probably is the specification of different sets of target properties. For example, in stage 4, the liquid-liquid separation is carried out at 298.15 K and 0.1 MPa. In Table 9, the performance of the solvents for the acetic acid-water problem used in this work is compared with those reported by others (Brignole et al., 1985; Joback and Stephanopoulos, 1989; Macchietto et al., 1990). The diagnostic information from the CAMD program is given in Table 6. The CPU time for the ethanol-water example is larger than that for others because this example needed more phase split computations. In general, most of the total CPU time is required for stages 3 and 4 (nearly 80% of the total CPU).

Azeotropic and/or Extractive Distillation. The systems of ethanol-water, acetic acid-water and acetone-heptane form azeotropes and therefore are often separated by azeotropic or extractive distillation. Using the modified UNIFAC model (Larsen et al., 1987), we have tried to determine solvents for separation by azeotropic/extractive distillation. The performance of these solvents is shown in terms of ternary-phase boundary diagrams (Figures 4a, 4b, and 4c) showing the azeotropic points, the heterogeneous liquid surface (if they exist), and the corresponding vapor line. It can be seen that for the acetone-heptane example, it has been possible to find a solvent that forms a homogeneous system and does not form azeotropes with either acetone or heptane. Also, for the acetic acid-water problem, a homogeneous system has been found. The solvent (ethylbenzene) forms azeotropes with acetic acid as well as water. For the ethanol-water problem, solvents (for example, toluene) that form heterogeneous systems have been found. Although detailed design has not been considered, Figures 4a, 4b and 4c do show that it is possible to obtain pure acetone, ethanol, or acetic acid. The separation of styrene from xylene is an extremely difficult separation task. For this problem, we have tried to obtain solvents for extractive distillation.

The solvents selected by the proposed CAMD program for ethanol-water, acetic acid-water and styrene-xylene separations are in Table 8 (the specified target properties are given in Table 7). It should be noted that the choice of these solvents depend strongly on the conditions at which the separations are performed. In this work, the separation factors were determined through *PT* flash calculations with pressure at 0.1 MPa and a range of temperatures. That is, the solvents in Table 8

Table 7. Specifications for Solvent Design Problems

Target Properties	Ethanol-Water LLE	Ethanol-Water VLE	Acetic Acid-Water LLE	Acetic Acid-Water VLE	Styrene VLE
Max. m_w	300	300	300	300	300
Min. t_b, K at 0.1 MPa	360	330	400	330	340
Min. S_s	10	1	50	1.0	1.0
Min. S_p	0.05	0.05	0.40	0.05	0.05
Min. S_f	5.0	1.0	3.0	1.0	1.0
Min. S_F	2.0	1.0	3.0	1.0	1.0
Temp., K	298.2	360–380	298.2	360–380	360–410
Miscibility	partial	—	partial	—	—
Phase Split(L/V)	—	yes	—	yes	yes

Table 8. Alternatives for Solvent Design Problems

Compound	S_s	S_p	S_f	S_F	m_w	t_b K	PI
Ethanol-Water (LLE)							
5CH3,3CH	122.0	0.104	13.4	18.7	114.0	381.1	2.19
4CH3,2CH2,2CH	121.9	0.104	13.3	18.6	114.0	381.6	2.18
2CH3,6CH2	121.8	0.104	13.3	18.6	114.0	382.4	2.16
3CH3,4CH2,CH – 0	16.8	0.209	3.0	4.7	131.0	405.4	0.11
Acetic Acid-Water (LLE)							
4CH3,2CH2,CH,CH = C	398.2	0.676	27.6	37.7	127.0	410.1	8.19
4CH3,2CH2,CH,CH = C	410.0	0.650	26.2	36.2	126.0	408.9	7.52
2CH3,5CH2,CH2 = C	337.8	0.653	20.8	29.6	126.0	401.9	4.87
4CH3,CH2,2CH	691.5	0.544	40.3	55.3	100.0	358.7	3.76
3CH3,3CH2,CH – O	95.7	1.192	8.5	11.9	131.0	405.1	0.77
*Acetic Acid-Water (VLE)**							
CH3,5ACH,ACCH2	39.7	0.243	10.4	22.6	106.0	409.1	2.212
5CH3,CH2,C,CH – O	4.5	0.298	5.4	25.3	130.0	401.2	1.050
2CH3,CH2CO	7.8	0.650	3.9	19.1	72.0	344.8	1.050
CH3,CH2,CH3CO	5.8	0.705	3.9	17.2	72.0	344.8	0.937
*Ethanol-Water (VLE)**							
5ACH,ACCH3	33.3	0.131	9.1	31.9	92.0	386.2	3.165
6ACH	21.8	0.113	4.5	16.0	78.0	358.4	0.924
2CH3,CH2,CH,CH3CO	4.8	0.422	5.0	29.8	100.0	390.1	1.490
2CH3,CH2,CH2CO	5.3	0.360	4.4	17.7	86.0	367.7	0.899
*Styrene-Xylene (VLE)***							
CH3,CH2,CH2CN	1.5	0.758	1.3	1.2	69.0	393.3	0.023
CH3,CH2,CH3COO	1.4	0.820	1.1	1.2	88.0	349.1	0.014
2CH3,CH2,CH2COO	1.3	0.930	1.1	1.2	102.0	372.9	0.013
2CH3,CH2,2CH2O	1.2	0.823	1.4	1.2	104.0	358.6	0.012

*Selection of compounds depend on the separation temperature (360-380 K).
**Same separation temperature (400 K) used for all the compounds.

were not obtained from a single run of the CAMD program, but from a series of runs with the temperature for *PT* flash as a variable. The diagnostic information from the CAMD program is given in Table 6. The low *PI* values for some of the solvents reflect the degree of difficulty of separation associated with these solvents. Detailed design, however, would be needed before making any final conclusions.

Gas Absorption. For a mixture of gases containing nitrogen, carbon dioxide, and hydrogen, solvents have been designed to absorb carbon dioxide. It should be noted that separations by reaction has not been considered in this article. The solvents obtained in this work (see Table 10) can be compared with those reported by Macchietto et al. (1990) in terms of solubility (in mol %). From the compounds selected, solvent mixtures can be determined by solving the following optimization problem.

Having target properties and their values, determine the compounds and their composition in the mixture such that the mixture satisfies properties constraints and maximizes a performance index (Eq. 11). Examples of some solvent mixtures (not optimal solutions) are also shown in Table 10. The compounds forming solvent mixtures were selected by trial and error. Also, because the models for calculating gas solubilities have not yet been implemented in the CAMD program, cal-

Table 9. Performance of Solvents for Acetic Acid-Water Example Calculated at 298.15 K and 0.1 MPa

Solvent	Solute Recovered %	Raffinate Removed %	Solvent Loss %	Ref.
4CH3,3CH2,CH = C	74.1	92.2	0.005	This work
4CH3,2CH2,CH,CH = C	72.2	91.9	0.007	This work
CH3,4CH2,CH3O	complete miscibility			Joback and Stephanopoulos (1989)
CH3,4CH2,CH3CO	complete miscibility			Joback and Stephanopoulos (1989)
2CH3,CHOH + 2CH3,8CH2	43.1	95.8	17.9	Joback and Stephanopoulos (1989)
2CH3,CH(COOH) + 4CH3,C	49.6	86.2	22.5	Joback and Stephanopoulos (1989)
4CH3,CH2,2CH	69.0	94.7	0.008	This Work; Macchietto et al. (1990)
6CH3,4CH	63.2	96.4	0.004	Macchietto et al. (1990)
3CH3,6CH2,CH	65.6	96.3	0.003	Macchietto et al. (1990)
CH3,2CH2,CH3COO	complete miscibility			Brignole et al. (1985)

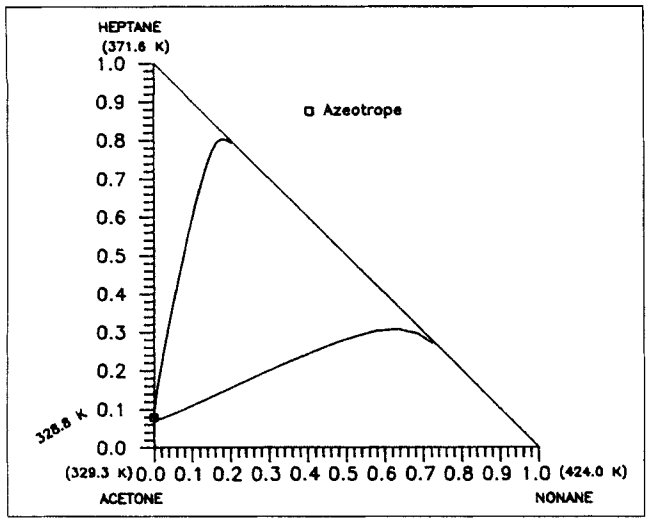

Figure 4a. Acetone-heptane-nonane (solvent) residue curves and location of binary azeotrope (homogenous system).

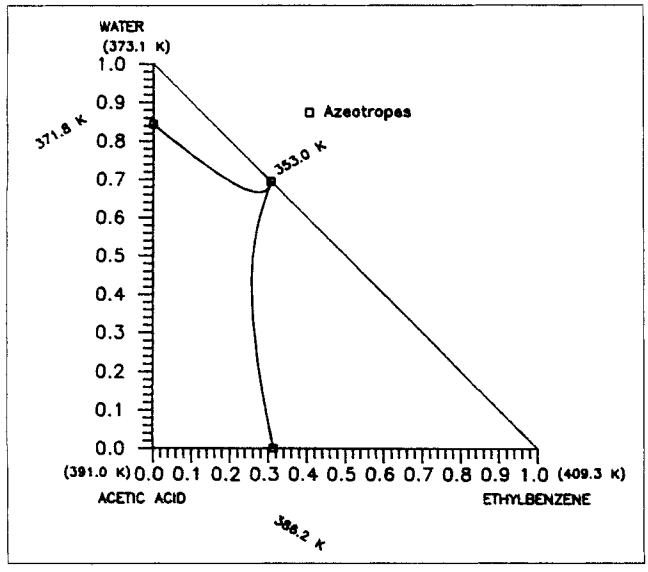

Figure 4b. Acetic acid-water-solvent (ethylbenzene) residue curves and location of azeotropic points.

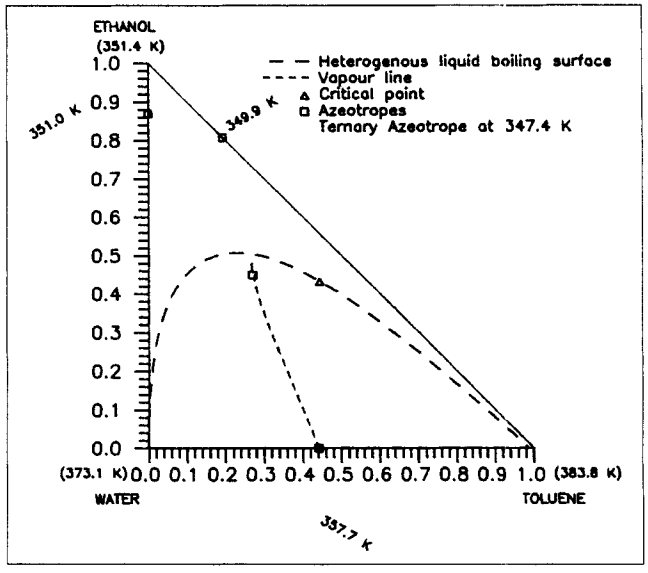

Figure 4c. Phase diagram (VLLE) for ethanol-water-solvent plus location of azeotropic points.

culations for the different stages have been performed separately. Future work in this area will deal with the optimization algorithm and new application results.

Conclusions

The scope of CAMD has been discussed using several examples. The CAMD methodology has shown to form chemically-feasible compounds in an optimal fashion. It is important to predict target properties accurately, and therefore it is dangerous to oversimplify the computation methods of target properties. Oversimplifications should be used as a last resort when no other option is feasible.

A computer program based on the CAMD methodology needs to be developed further in the areas of the interface of design/simulation and selection/verification of the final design alternatives. In addition, the computation methods of the target properties need to be improved, their group parameter tables need to be enlarged, and a system for generating default values for target properties needs to be developed. This work addresses some of these issues. Although improving computational methods or estimating new group parameters may be

Table 10. Solvents for Absorption of Carbon Dioxide

Solv. No.	Solvent Name	Equimolar Mixture of Solvent No.	Solubility (mol %) at 298.15 K		
			CO2	N2	H2
1	CH3,CH3COO	—	2.11	0.052	0.028
2	CH3,CH2,CH3CO	—	1.95	0.057	0.032
3	1-Pentane	—	1.21	0.150	0.075
4	Pentane	—	1.18	0.137	0.063
5	Heptane	—	1.19	0.170	0.063
6	2-Propanol	—	0.65	0.040	0.036
7	Ethanol	—	0.63	0.035	0.019
8	Water	—	0.62	0.001	0.002
9	—	1,3	1.64	0.102	0.054
10	—	2,3	1.60	0.102	0.054
11	—	6,7,8	0.48	0.026	0.014

beyond the scope of CAMD, application of the CAMD technology to new problems still should be studied. Research on determining minimum-cost solvents for paint formulations and solvents for polymers (and polymers blends) is in progress. Under consideration for future study is separation by reaction.

For the CAMD technology to be successful, considerable effort is needed to properly interface the computation methods of target properties with the screening of design alternatives. Accurate prediction of the performance of the designed compound(s) depends on the accuracy of the models used to compute the target properties. Acceptance of some of these models in process design is doubtful. Since CAMD is the reverse problem of computation of thermodynamic properties, its acceptance depends on the acceptance of the computation methods being used for these properties.

Acknowledgment

The authors wish to acknowledge the support and interest received from Dr. D. T. Wu (E. I. duPont de Nemours and Company, Inc.) and Dr. P. Gupte (Union Carbide Chemicals and Plastics Company, Inc.).

Notation

c_p = heat capacity
$C1$ = rules for formation of feasible compounds
$C2$ = rules for final design/selection stage
G^E = excess Gibbs energy
H^E = excess enthalpy
m_w = molecular weight
M_p = partial miscibility
$MC1$ = total number of rules corresponding to $C1$
$MC2$ = total number of rules corresponding to $C2$
NC = total number of feasible compounds formed
NT = number of times a group appears in a compound
NP = total number of target properties
NGT = total number of preselected groups
NG = number of different types of groups present in a compound
p_s = pure component vapor pressure
P = array of target properties
s = specified value (see Eqs. 3 and 4)
S^E = excess entropy
SG = identity of group present in a compound
SGT = group identification number
S_p = solvent power, solubility
S_s = selectivity
S_f = selectivity based on composition of mixture to be separated
S_F = separation factor
x = mole fraction
z = compressibility factor

Greek letters

γ = activity coefficient
φ = fugacity coefficient
θ = temperature
π = pressure

Literature Cited

Andersen, P. M., E. A. Brignole, and Aa. Fredenslund, "A Multiple-Purpose Phase Equilibrium Package for Separation Process Simulation and Design," *I. Chem. Eng. Symp. Ser.*, No. 92, 313 (1985).

Barton, A. F. M., *CRC Handbook of Solubility Parameters and Other Cohesive Parameters*, CRC Press, Boca Raton, FL (1983).

Brignole, E. A., S. Bottini, and R. Gani, "A Strategy for Design and Selection of Solvents for Separation Processes," *Fluid Phase Equilibria*, **29**, 125 (1986).

Dahl, S., P. Rasmussen, and Aa. Fredenslund, "The MHV2 Model—A UNIFAC-Based Model for Prediction of Gas Solubility and Vapor-Liquid Equilibria at Low and High Pressures," SEP9023, Institut for Kemiteknik, Technical University of Denmark, Lyngby (1990).

Daubert, T. E., and R. Danner, *DIPPR Data Compilation*, AIChE, New York (1986).

Fredenslund, Aa., J. Gmehling, and P. Rasmussen, *Vapor-Liquid Equilibria Using UNIFAC*, Elsevier, Amsterdam (1977).

Gani, R., and E. A. Brignole, "Molecular Design of Solvents for Liquid Extraction Based on UNIFAC," *Fluid Phase Equilibria*, **13**, 331 (1983).

Gani, R., M. Castier, and Aa. Fredenslund, "Model Unification: A Means of Communication Between a Class of Activity Coefficient Models," SEP8701, Institut for Kemiteknik, Technical University of Denmark, Lyngby (1987).

Helme, A. D., "Prediction of Liquid Mixture Viscosities by UNIFAC," MSc Thesis, Institut for Kemiteknik, Technical University of Denmark, Lyngby (1989).

Joback, K. G., and R. C. Reid, "Estimation of Pure Component Properties from Group Contributions," *Chem. Eng. Commun.*, **57**, 233 (1987).

Joback, K. G., and G. Stephanopoulos, "Designing Molecules Possessing Desired Physical Property Values," *Proc. FOCAPD'89*, Snowmass, CO (1989).

Larsen, B., P. Rasmussen, and Aa. Fredenslund, "A Modified UNIFAC Group-Contribution Model for Prediction of Phase Equilibria and Heats of Mixing," *I&EC Res.*, **26**, 2274 (1987).

Lo, T. C., M. H. I. Baird, and C. Hanson, *Handbook of Solvent Extraction*, Wiley, New York (1983).

Macchietto, S., O. Odele, and O. Omatsome, "Desig.i of Optimal Solvents for Liquid-Liquid Extraction and Gas Absorption Processes," *Trans. IChemE.*, **68**, Part A, 429 (1990).

Magnussen, T., P. Rasmussen, and Aa. Fredenslund, "UNIFAC Parameter Tables for Prediction of Liquid-Liquid Equilibria," *I&EC Process Des. and Dev.*, **20**, 331 (1981).

Nielsen, B., R. Gani, and Aa. Fredenslund, "Computer-Aided Molecular Design by Group Contribution Approach," *Comp. Applic. in Chem. Eng., Process Technol. Proc.*, Vol. 9, p. 227, Elsevier, Amsterdam (1990).

Rosendahl-Jensen, S., personal communications, Institute for Organic Chemistry, Technical University of Denmark, Lyngby (1990).

Reid, R. C., J. M. Prausnitz, and B. E. Poling, *The Properties of Gases and Liquids*, McGraw-Hill, New York (1987).

Rowlinson, J. S., and F. L. Swinton, *Liquids and Liquid Mixtures*, Butterworth Scientific Publishers, 3rd ed., Great Britain (1982).

Sander, B., Aa. Fredenslund, and P. Rasmussen, "Calculation of VLE in Mixed Solvent/Salt Systems Using an Extended UNIQUAC Equation," *Chem. Eng. Sci.*, **41**, 1171 (1986).

Sørensen, J. M., R. Gani, and Aa. Fredenslund, "A Computer Program Package for Phase Equilibria," *Comp. Applic. in Chem. Eng., Process Technol. Proc.*, Vol. 9, p. 315, Elsevier, Amsterdam (1990a).

Sørensen, E. L., H. Johansen, R. Gani, and Aa. Fredenslund, "A Dynamic Simulator for Design and Analysis of Chemical Processes," *ibid*, p. 13 (1990b).

Visser, T. N., "Calculation of the Flash Point of Solvent Blends," *Symp. der Technische Academie Esslingen ueber Stand und Neuere Entwicklungen bei Organischen Loesemitteln fuer das Lack und Klebstoffgebiet*, Amsterdam (Apr., 1988).

Wu, D. T., and J. Klein, "Prediction of Mixture Properties by Group Contribution Approach," XX FATIPEC Congress, Nice, France (Sept., 1990).

Appendix

The algorithm in stage 2 (formation of chemically-feasible compounds) can be applied as follows.

Step 1. Let us specify that all the groups in Table 1 be combined to branched, nonaromatic compounds: lower limit for $MA = 4$; assume upper limit for $MB = 5$, and set $MC = 1$. Then, structures with minimum four groups and maximum five groups will be formed. Let us start with five-group structures ($MB = 5$).

Step 2. We now have to choose a group from category MC and class $M2 = MB - 1$; thus, choose a group for $MC = 1$ and $M2 = 4$. And then, choose the group, C. In step 3, we will form all five-group structures with C as one of the groups.

Step 3. We now determine $M3 = MB - M2 = 5 - 4 = 1$. Therefore, we must choose $M2 = 4$ groups from class $M3 = 1$ according to the restrictions in Table 3. From category 1, the only available group of class 1 is CH3. Thus, the first five-group structure will have four CH3 groups and 1 C group. The next set of structures will have one C group, three CH3 groups and one group from categories 2, 3, 4 or 5. Since there are 32 groups in categories 2 through 5 of class 1, 32 structures are formed. The next set of structures will have one C group, two CH3 groups and two groups from categories 2–5 such that the restrictions given in Table 3 are not violated. That is, there cannot be groups from categories 4 and 5 in the same structure. One hundred ninety-three structures with five groups are formed.

The next set of structures will have one C group, one CH3 group and three groups from categories 2–5. Note that the sum of groups of categories 3, 4 and 5 cannot be greater than 2 for a five-group structure (see Table 3). Three hundred eighty-three structures with five groups are formed. The next set of structures will have one C group, and four groups from categories 2–5. Since the sum of groups from categories 3, 4 and 5 cannot be greater than 2 in a five-group structure, there must be at least two groups from category 2. Also, any single group cannot be attached to more than three groups from categories 2–5. Thus, we cannot form any five-group structure with 1 C group and no CH3 group.

Step 4. In this example, this step is not needed yet.

Step 5. In this example, this step is not needed yet.

Step 6. Set $MC = MC + 1$.

Step 7. Set $MB = MB - 1$. Repeat from step 2 with new values of MC ($= 2$), MB ($= 4$) and $M2 = MB - 1 = 3$.

Note that in step 2 a total of only 613 chemically-feasible structures are formed (all structures having group C) out of a total of more than 20,000 possible combinations. If all structures containing halides and/or the groups: CH3S, CH2SH, C4H3S, SH, and SiH3 are removed, and then only 238 chemically feasible structures are formed. Group parameters for all the 238 compounds are available for designing solvents for separation of ethanol-water mixtures. For designing solvents for separation of acetic acid-water mixtures, 29 of the compounds have to be rejected because of unavailable group parameters.

Steps 4 and 5 are needed, when groups from class 2 and higher are considered in step 3. Steps 2 through 7 are repeated until $MB = MA$. In this example, therefore, steps 2 through 7 are repeated four times (twice for five-group structures and twice for four-group structures with classes 3 and 2 as the largest class of the groups present in the structures. For example, if we start with the group CH, then we choose one group from class 2 and three groups from class 1 to have five-group structures. For four-group structures with one CH group, we can only choose groups from class 1. For five-group structures, if we choose a group from category 4 of class 2, then according to Table 3, the remaining three class-1 groups cannot be from category 4 or 5.

Manuscript received Feb. 25, 1991, and revision received July 15, 1991.