

GENERAL RESEARCH

Molecular Design of Solvents: An Efficient Search Algorithm for Branched Molecules

M. Cismondi and E. A. Brignole*

PLAPIQUI, Planta Piloto de Ingeniería Química, UNS-CONICET, Casilla 717, 8000 Bahía Blanca, Argentina

A fast generation and test algorithm is applied within a new approach for designing branched molecules with physical properties and molecular constraints. Computer-aided product design requires the generation of feasible compounds and the prediction of mixture and pure-component properties by group contribution methods. A desirable property of the molecular design procedure is the chemical stability/feasibility of the generated chemical structures. On the basis of the electronegativity of the group attachments, a new characterization of group combination properties and the corresponding feasibility criteria for computer-aided generation of branched structures are presented. A synthesis procedure and a strategy for reducing the size of the combinatorial synthesis problem are discussed. Finally, the corresponding step-by-step algorithm is described.

1. Introduction

There are a great number of engineering problems, which consist of finding the best performance case for a determined function, or a set of good performance cases, from a large universe of discrete alternatives or situations. In most cases, these alternatives can be generated by some specific combinatorial-type procedure, then contrasted with a set of restrictions, and finally (those feasible alternatives) evaluated in terms of a given performance.

It is clear that these kinds of problems can be explicitly formulated as optimization problems and then solved by a specific MINLP solver. However, in some cases, a tailored and efficient procedure that performs an exhaustive search with strategies to reduce the combinatorial problem can be implemented.

In a typical algorithm of this kind, one can distinguish three different stages (Figure 1):

1. Generation of alternatives through a tailored combinatorial procedure.
2. Screening of alternatives according to different restrictions. This stage can rather be a chain of screening stages when there is more than one restriction.
3. Performance evaluation of the remaining feasible alternatives, through calculation of a set of properties or a performance index using predictive models.

In some cases, one property that can be exploited to improve the efficiency of the algorithm is the fact that different solutions belong to the same “meta”-alternative; i.e., they are equivalent from the point of view of a set of restrictions.^{1,2} In other words, if one of them satisfies the restriction, the rest will also meet the requirement. This makes it possible to insert the cor-

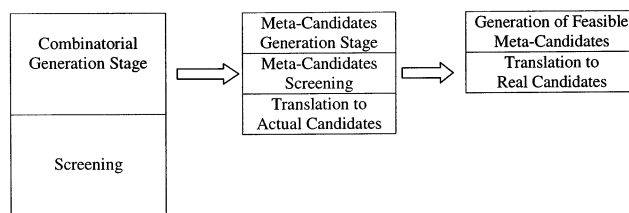


Figure 1. Conceptual steps involved in improving the efficiency of the algorithm.

responding screening in the middle of a fragmented generation stage. Meta-alternatives are generated first and then screened. In the next step, only the accepted meta-alternatives are used to generate, in a combinatorial way, the feasible real solutions that are going to be evaluated.

In addition, if the required restrictions are formulated in such a way that can be coupled with the generation stage, then two stages merge into one, and only feasible alternatives are being generated. As a consequence of these two facts, we simplify the combinatorial problem, and therefore the computational effort is considerably reduced.

A class of problems that meet these requirements is the Computer Aided Molecular Design of Solvents.^{3–5} In this paper we will discuss the application of the above ideas in the development of a procedure for the computer-aided product design of branched molecules.

2. Molecular Design of Solvents

The application of group contribution methods for the prediction of pure-component and mixture properties opened the way for efficient computer-aided evaluation of chemical compounds. The molecular design of solvents has its origin in the backward solvent selection problem, formulated as follows: “giving a set of property constraints and certain performance indexes, generate

* To whom correspondence should be addressed. Tel.: 54 291 4861700. Fax: 54 291 4861600. E-mail: ebrignole@plapiqui.edu.ar.

chemical structures with the desired set of physico-chemical properties".³

Several methods or strategies have been proposed in the last 2 decades for Molecular Design of Solvents and they differ in (i) the particular application, (ii) the evaluation of solvent properties (mainly the thermodynamic model used), and (iii) the approach implemented to find the optimum solvent structures.

For the computer-aided generation and evaluation of molecular structures a combinatorial partition strategy was proposed by Brignole et al.⁶ Several molecular design procedures based on explicit optimization algorithms have also been proposed.⁷⁻⁹ Applications have been reported for the design of polymers,⁸ refrigerants,^{2,9} and solvents.^{3,4,10,11} A compilation of different approaches to computer-aided molecular design and future developments has been presented by Achenie et al.,¹² and recently Sahinidis et al.¹³ presented a state-of-the-art review in optimization-based algorithms for molecular design. Even though the generated molecules satisfy the neutrality condition, the chemical stability of the components is not guaranteed in many cases.^{7,8,13} This is partly due to the way groups are defined in different group contribution methods and/or the lack of proper combination rules for the groups.

Some of the component properties of interest from the point of view of environmental considerations depend on mixture or solution properties. The first solvent design studies were based on solution properties derived from the UNIFAC group contribution method for computing activity coefficients.¹⁴ Several revisions and extensions to electrolytes, polymers, and equations of state of the original UNIFAC predictive package have been presented;¹⁵ a group contribution equation of state based on similar but more detailed group definitions has been extended to new groups and gases.¹⁶⁻¹⁸

Different group definitions have been proposed¹⁹ for the prediction of pure-compound properties, such as heat capacities, solubility parameters, formation energies, etc. In general, specific criteria⁹ to ensure the component valence neutrality and the fulfilment of the octet rule are required. These criteria are combined with optimization algorithms for the synthesis of molecules. However, the generation of chemically unstable compounds cannot always be avoided.

3. The MOLDES Approach

The program MOLDES for Molecular Design of Solvents has been developed in PLAPIQUI, UNS-CONICET, and is based in the works of Gani and Brignole,³ Pretel et al.,⁴ and Cismondi and Brignole.⁵

MOLDES performs an exhaustive solvent search for liquid-liquid extraction or extractive distillation building solvent structures made up with functional groups of the UNIFAC group contribution model. The evaluation is made through performance indexes estimated by UNIFAC and other group contribution methods.^{20,21}

Table 1 shows the features that can be identified in the MOLDES approach from the point of view of a solvent search optimization problem. The analysis of the chemical feasibility constraints is discussed next.

4. Chemical Feasibility Criteria

The chemical feasibility constraints have been proposed to eliminate not only the unstable or reactive

Table 1. Solvent Search Optimization in the MOLDES Context

general concept	expression in MOLDES
engineering problem alternatives	suitable solvent for a given separation
main restriction	molecular structures
performance index	chemical feasibility
	specific functions in terms of distribution coefficient, relative volatility, etc.
model optimization formulation	UNIFAC or group contribution based knowledge-based

molecular structures but also the ones that could lead to unreliable UNIFAC predictions due to proximity effects.

Therefore, the development of chemical feasibility criteria can be established on the basis of the electronegativity of the group bonds.^{3,4} Three basic types of group attachments (valences), with increasing degree of electronegativity, were identified⁴ in aliphatic compounds: J, L, and K. The methyl group ($-\text{CH}_3$), even though it is a J-type valence group, is identified as type M because of the different role it plays in the feasibility criteria analysis, with respect to the other J valence groups. The strong electronegative valences are considered to be of type K, for example, in the $-\text{OH}$, $\text{CH}_3\text{-COO-}$, and $\text{CH}_3\text{CO-}$ of the hydroxyl, ester, and keto groups; an L-type attachment (valence) is the case of methylene groups that are attached to a strong electronegative group fragment; for instance, $-\text{CH}_2\text{COO-}$ is a divalent group with type (L,1) and (K,1) valences, and $-\text{CH}_2\text{Cl}$ has a type (L,1) valence.

Pretel et al.⁴ proposed the following combination rules for aliphatic structures:

(R1) Type K attachments can only be combined with M or J attachments.

(R2) Type L attachments can be combined with L, M, or J attachments.

(R3) Type J attachments change to L after their combination with a type K attachment, in accordance with the definition of an L attachment.

These combination rules justified the formulation of the following criteria for the generation of linear structures:⁴

"The set of groups that makes up a molecular structure should have a number of K attachments less than or equal to the number of unrestricted carbon atoms (the sum of groups J and M)."

A synthesis procedure based on intermediate and terminal groups for the generation of linear (not branched) molecules was implemented. In a first stage, intermediate structures (i.e., those formed by the divalent groups) are generated. Then final structures are obtained through a combination of intermediate structures with terminal groups.

An extension of that synthesis procedure for aliphatic-branched molecules is developed in the present work.

It should be pointed out that the majority of group contribution predictive models do not distinguish among isomers formed by the same set of groups. Therefore, the solution of a Molecular Design Program to a specific problem does not consist of completely defined chemical structures but rather of sets of groups, where each one can lead to different isomer structures.

Consequently, the algorithm should guarantee that at least one of the possible isomers in each case satisfies the chemical feasibility constraints.

Table 2. Redefinition of Group Combination Properties in Terms of J and K Bond Status

UNIFAC group example	group valence	previous characterization	decomposed in subgroups	new combination properties
CH ₂ Cl	1	(L,1)	$J_2 + K_1$	(K,1), (J,2)
CHCl	2	(L,2)	$J_3 + K_1$	(K,1), (J,3)
CCl	3	(L,3)	$J_4 + K_1$	(K,1), (J,4)
CH ₂ CO	2	(K,1), (L,1)	$J_2 + K_2$	(K,2), (J,2)
CHNH	3	(K,1), (L,2)	$J_3 + K_2$	(K,2), (J,3)
CH ₂ N	3	(K,2), (L,1)	$J_2 + K_3$	(K,3), (J,2)

For example, if the final ranking for a given separation shows the structure defined by the set of groups CH₃, CH₂, CH, OH, and COOH, there are three possible isomers: (a) CH₃-CH (COOH)-CH₂-OH; (b) CH₃-CH (OH)-CH₂-COOH; (c) COOH-CH (OH)-CH₂-CH₃. Isomers a and b are considered feasible but not the isomer c because of a proximity effect of two strong electronegative groups that will not follow a first-order group contribution approach in the prediction of solution properties by UNIFAC

5. Synthesis of Branched Molecules and Feasibility Criteria

The extension of Pretel et al.'s⁴ feasibility criteria to branched structures would be

$$\sum_i iK_i \leq M + J \text{ (final structure)} \\ \text{with } J = J_2 + J_3 + J_4 \quad (1)$$

where K_i or J_i are the number of K or J groups with i attachments in the structure and M is the number of methyl groups. However, in many cases this criterion leads to unfeasible structures like HCOO-CH-CH₃-OH for a molecule with a tertiary carbon (CH).

Because of the third combination rule (R3), a type K attachment has a greater impact on the structure when it is combined with a trivalent or tetravalent group than when combined with a divalent one. For this reason, eq 1 fails when applied to branched structures.

5.1. New Group Combination Property Characterization. It can be seen that the formulation of robust feasibility criteria for the synthesis of branched structures requires a new approach.

We can identify many UNIFAC groups, having a functional group plus a CH₂, CH, or C as "combined" groups. A few examples are CH₂NH₂, CH-O, CH₂CO, CHNH, CCl, etc.

These groups were originally defined in this way to improve the predictions of the model. However, from the point of view of the chemical feasibility analysis and taking into account R3, the carbon-free attachments in these groups are of type L (moderate electronegativity). We can consider these "combined" groups as a combination of a K group and a J group. This would be equivalent to characterizing not only the group free attachments but also the internal bonds, in the case of groups having L attachments. When the internal and free bonds are taken into account, only two bond statuses, K (electronegative) and J (neutral), are required to define the combination properties.

It can be seen that groups with L attachments are formed by a combination of "pure" K and J groups, as shown in Table 2. In the discussion that follows, we will use the word "group" referring either to a UNIFAC

Table 3. Aliphatic Metagroups in the New Characterization of Combination Properties

number of attachments	metagroups, attachment characterization
tetravalence (X4)	K4, J4
trivalence (X3)	K3, K3J2, K2J3, K1J4, J3
dual valence (X2)	K2, K2J2, K1J3, J2
single valence (X1)	K1, K1J2, M

Table 4. Revision of the Combination Properties of UNIFAC Groups

combination properties	groups			
(M,1)	CH ₃			
(J,2)	CH ₂			
(J,3)	CH			
(J,4)	C			
(J,2), (K,1)	CH ₂ Cl	CH ₂ NH ₂	CH ₂ CN	CH ₂ SH
(J,3), (K,1)	CHCl	CHNH ₂	HCON(CH ₂) ₂	
(J,4), (K,1)	CCl	CF		
(J,2), (K,2)	CH ₂ CO	CH ₂ COO	CH ₂ O	CH ₂ NH
	CONHCH ₂	CONCH ₂ CH ₂	FCH ₂ O	CH ₂ S
	C ₂ H ₄ O ₂			
(J,3), (K,2)	CH-O	CHNH	CON(CH ₂) ₂	CHS
(J,2), (K,3)	CH ₂ N			
(K,1)	CH ₂ =CH	OH	CH ₃ CO	CHO
	CH ₃ COO	HCOO	CH ₃ O	CH ₃ NH
	C ₅ H ₄ N	COOH	CHCl ₂	CCl ₃
	CH ₂ NO ₂	I	Br	CH≡C
	Cl(C≡C)	SiH ₃	CCl ₂ F	HCClF
	CClF ₂	C ₂ H ₅ O ₂	CH ₃ S	C ₄ H ₃ S
	CONH ₂	CONHCH ₃	CON(CH ₃) ₂	CF ₃
(K,2)	CH=CH	CH ₂ =C	CH ₃ N	C ₅ H ₃ N
	CCl ₂	CHNO ₂	C≡C	COO
	SiH ₂	SiH ₂ O	C ₄ H ₂ S	CF ₂
(K,3)	CH=C	SiH	SiHO	
(K,4)	C=C	Si	SiO	
(I,1)	ACH	ACF		
(H,1)	ACCH ₃	ACOH	ACNH ₂	ACCl
	ACNO ₂			
(K,1), (H,1)	AC	ACCOO		
(J,2), (K,1), (H,1)	ACCH ₂			
(J,3), (K,1), (H,1)	ACCH			

group or to a K or J group, where the former can contain more than one group of these types.

According to the above discussion, a revision of the UNIFAC group combination properties was made. Groups taken from three different UNIFAC tables²²⁻²⁴ have been classified in 19 metagroups, 14 aliphatic and 5 aromatic, which are shown in Table 3. The complete classification of the actual groups is presented in Table 4. With the new group characterization, it is possible to formulate more general feasibility criteria.

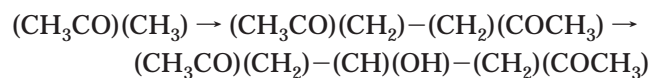
When Table 3 is examined, it should be noted that the metagroups K4M, K3M, and K2M have not been defined because all of their external attachments are K. Therefore, these groups are considered as K3, K2, and K1, respectively. Examples of these groups in UNIFAC are CH₃N and CH₃COO for the last two cases. There is no example of a K4M group in the three cited UNIFAC tables, but a possible one could be CH₃Si. If we had distinguished these groups from the strictly K ones in our classification, the results obtained with the proposed feasibility analysis and the corresponding algorithm would have been the same but with unnecessary additional definitions and computer effort. The same reasoning leads to the classification of the group CON(CH₃)₂, which would be a K3(M)₂, such as K1.

Also, when a UNIFAC group contains two or three K1 groups attached to a J3 or J4 group, we consider it to be a pure K group. Examples are CHCl₂, CClF₂, CCl₂, etc. Note that, according to the aforementioned feasibility rules, these combinations would not be permitted if

they were built starting with separate UNIFAC groups. Other compound groups considered to be of the pure K type, even though there is only one K1 group, are those containing the NO₂ group: CH₂NO₂ and CH₂NO₂. This is explained by the strong electronegative character of the group NO₂, which generates an effect over the J group comparable to that produced by two or three less electronegative K groups.

Finally, a special case involves the groups HCON-(CH₂)₂ and CON(CH₂)₂. The conformation or structure of these groups are K2(J2)₂ and K3(J2)₂. However, as unique entities, these configurations behave like K1J3 and K2J3, respectively.

5.2. General Feasibility Criteria for the Synthesis of Linear or Branched Structures. Considering that there are only pure "K" and "J" groups, the new synthesis concept that summarizes the above combination rules is as follows: *each J group cannot be attached to more than one K group.* This concept leads to the fact that the building of feasible molecules requires the existence of a J–J-type bond for each K group incorporated into the molecule (after the first one, for noncyclic structures). This is clear in the following example of a sequence of feasible structures, the ketone homologues, in which only the J–J bonds are shown with en dashes:



Now a first formulation of the feasibility criteria can be written as follows:

$$K \leq \text{NJJ} \quad (\text{cyclic}) \quad (2)$$

$$K - 1 \leq \text{NJJ} \quad (\text{noncyclic}) \quad (3)$$

where NJJ = number of J–J bonds. These conditions are valid for both the intermediate and final structures. It should be pointed out that M groups are not considered for compensating the K attachments because, being monovalent, they do not allow for the addition of another K group. A total number of J attachments balances could be obtained as follows (see the appendix for the derivation):

$$\sum_i iJ_i = 2\text{NJJ} + \text{NJF} \quad \text{when } K \leq \text{NJF} \text{ with } i = 1, 4 \quad (4)$$

or

$$\sum_i iJ_i = 2\text{NJJ} + \text{NJF} + 2(K - \text{NJF}) \quad \text{when } K > \text{NJF} \quad (5)$$

where the number of J free attachments is given by

$$\text{NJF} = J_3 + 2J_4 + 2 \quad (\text{noncyclic and } J \geq 1) \quad (6)$$

or

$$\text{NJF} = J_3 + 2J_4 \quad (\text{cyclic}) \quad (7)$$

Considering the attachment definitions and eqs 2–7, the general feasibility criteria obtained are presented in Table 5. See the appendix for the derivation.

When $J = 0$, the J attachments balance is irrelevant and the only way to comply with the feasibility rules is not to allow more than one K group in the structure.

Table 5. Feasibility Criteria for Linear and Cyclic Branched Structures

	$K \leq \text{NJF}$	$K > \text{NJF}$
noncyclic structures	$K \leq J$	$2K \leq J + \text{NJF}$
cyclic structures	$K \leq J$	$2K \leq J + \text{NJF}$
$J = 0$		$K \leq 1$

Examples of these small molecules, taking the UNIFAC group definitions as reference, are acetone, acetic acid, 2-butene, and chloroethane.

6. Computer-Aided Molecular Synthesis

Using the feasibility criteria of Table 5, an efficient combinatorial synthesis of branched molecules is implemented on the basis of metagroups, i.e., groups with the same combination properties (first column of Table 4). In the synthesis of linear molecules, the intermediate structures have two free attachments. However, the number of free attachments (NFA) in branched intermediate structures is larger:

$$\text{NFA} = 2 + \text{NV}_3 + 2\text{NV}_4 \quad (\text{noncyclic}) \quad (8)$$

or

$$\text{NFA} = \text{NV}_3 + 2\text{NV}_4 \quad (\text{cyclic}) \quad (9)$$

where NV₃ is the number of groups of valence three and NV₄ is that of valence four. The larger number of free attachments of the intermediate structures greatly increases the size of the synthesis problem in comparison to the synthesis of linear structures. To reduce the problem size and according to the previous discussion (which is applicable beyond the case of solvent design), the MOLDES computer program executes the following stages where a maximum number of 12 groups in final molecular structures (FMS) is allowed:

1. Selection of the set of intermediate and terminal groups in an interactive way.

2. Synthesis only of *feasible* meta intermediate molecular structures (IMS) with NFAs from 2 to 8 metagroups that correspond to the selected intermediate groups. Then, each meta-IMS is replaced by all of the different possible combinations of the selected groups to form "true" IMSs.

3. In the same way, obtaining pre-FMSs by adding NFA – 2 terminal groups to each IMS.

4. Screening of the pre-FMSs according to the physical property constraints.

5. Termination of FMSs by adding to each accepted pre-FMS different combinations of two terminal groups that preserve their molecular feasibility.

6. Screening of the synthesized FMSs according to the physical constraints for the desired product.

7. Ranking the remaining FMSs in accordance with molecular complexity and performance index, or specific user-defined criteria, indicating the predicted product properties.

We have used "product" instead of "solvent" to remark the generality of the stages in this procedure. In the case of solvent design, it should be noted that between stages 2–3, 3–4, and 5–6 the synthesis procedure eliminates all intermediate and final structures with unknown binary interaction parameters, reducing in this way the size of the combinatorial problem and the computing time. In any case, the synthesis program runs in a very short time on a standard PC.

Table 6. A Case of Solvent Design: Separation of Benzene from Hexane

Problem Definition	
component to be recovered	(ACH) ₆
principal component in the raffinate	(CH ₃) ₂ (CH ₂) ₄
separation type	liquid-liquid extraction
operation temperature	320 K
intermediate groups selected	C, CH, CH ₂ , DMF-2
terminal groups selected	CH ₃ , CH ₂ =CH, OH, CHO, CH ₃ COO, HCOO, CH ₂ CN, CH ₂ SH
Problem Constraints	
minimum selectivity for pre-FMS (wt)	1.00
maximum solvent loss for pre-FMS (wt %)	8.00
minimum selectivity (wt)	6.00
minimum solvent power (wt %)	20.00
maximum solvent loss (wt %)	4.00
minimum distribution coefficient (wt)	0.60
maximum molecular weight	260.0

Table 7. Results of the Synthesis Procedure

number of meta-intermediate structures generated	255
number of intermediate structures generated	255
number of meta prefinal solvents	1664
number of prefinal solvents	25595
prefinal solvents rejected by MW restriction	16247
prefinal solvents rejected by lack of binary parameters	4156
prefinal solvents evaluated	5192
number of accepted prefinal solvents	4622
number of meta final solvents generated	4638
number of final solvents generated	42436
solvents rejected by MW restriction	21887
solvents rejected by lack of binary parameters	10846
solvents evaluated	9703
number of final solvents that satisfy all physical constraints	742

The algorithm contemplates all possible cases for the set of groups chosen. For example, the user is not required to select at least one group per each metagroup. There is no need for him/her to be aware of the metagroups.

The metastructures are generated with an increasing K level, until the feasibility restriction is violated and the corresponding cycle is bypassed. It is clear that in this way the algorithm improves the "generate and test" paradigm,^{3,4} achieving the required efficiency to deal with bigger problems such as in the case of branched molecules.

Because of the partition of the combinatorial procedure for completing the IMSs with terminal groups, i.e., the formation of pre-FMS for a first evaluation, the same FMS can be obtained from different pre-FMSs. To solve this problem, a specific subroutine eliminates at the generation stage the synthesis of repeated structures.

6.1. MOLDES Application to Liquid Extraction.

The results of the synthesis procedure are illustrated with an example of solvent design for liquid extraction for the typical problem of separating aromatics from paraffins: We have considered the classic hexane/benzene system (see Table 6). In this example, 12 groups are selected for the synthesis procedure: intermediate (4) and terminal groups (8). We ran this problem with a fixed maximum molecular weight of 260. The solvent property constraints⁴ were highly restrictive (see Table 6), with the only exception of solvent loss due to the nature of the separation problem.

A summary of the results of the synthesis procedure is given in Table 7. In Table 8, the best solvents, one isomer from each family of compounds, are shown. The computer time required for running this example is less than 1 min on a standard PC.

Among the best solvents, we find some dialdehydes, triesters, and different combinations of these functional groups in the same structure. The thiol group and the olefinic group are also present in some cases. It should be noted that, in some cases, compounds in which only one functional group is present, repeated or not, are easier to obtain, and therefore less expensive, than those presenting different functional groups.

Figure 2 shows how solvent property constraints limit the search for solvent structures. It illustrates the evolution of the boiling points and selectivity, distribution coefficient, and solvent loss values through the homologues of different solvent families. It is interesting to note the opposite role that molecular weight plays in selectivity and the distribution coefficient. While selectivity decreases, distribution coefficients increase in all cases with the exception of dialdehydes. Precisely this feature, together with their having low solvent loss values, makes dialdehydes of low molecular weight the most promising structures. However, the chemical reactivity of aldehydes will prevent these compounds from being used in practice.

Although the three remaining families show comparably good values of distribution coefficient, diesters suffer from very high solvent loss and low selectivity. Because of the highest selectivity and lowest solvent loss values, triesters as well as dicyanides tend to show their best performances at low molecular weights, even when the distribution coefficients are not so good.

In general, the results of solvent screening by computer-aided molecular design should be taken as a preliminary guide in solvent selection. The reliability of UNIFAC as a predictive model for molecular design was discussed for different separation problems by Pretel et al.⁴ by comparison of computer solvent selection with actual experimental data. For the present case study, in our laboratory, we have confirmed the excel-

Table 8. Ranking of Solvents for the Separation of Benzene from Hexane by Liquid Extraction

solvent	selectivity weight basis	solvent power (wt %)	solvent loss (wt %)	distribution coefficient weight basis
(CH ₃) ₁ (CH ₂) ₁ (CH) ₁ (CHO) ₂	10.04	71.96	3.10	1.21
(CH ₃) ₂ (CH ₂) ₃ (CH) ₃ (CH ₃ COO) ₁ (HCOO) ₂	7.68	64.87	2.10	1.10
(CH ₃) ₄ (CH) ₅ (CHO) ₂ (HCOO) ₁	6.29	63.60	2.59	1.07
(CH ₃) ₁ (CH ₂) ₂ (CH) ₁ (CHO) ₁ (HCOO) ₁	6.85	62.54	3.17	1.06
(CH ₃) ₃ (CH ₂) ₁ (CH) ₄ (CHO) ₁ (CH ₃ COO) ₁ (HCOO) ₁	6.85	61.92	3.46	1.05
(CH ₂) ₈ (CH) ₁ (CHO) ₁ (HCOO) ₂	6.46	60.90	1.63	1.03
(CH ₃) ₂ (CH ₂) ₁ (CH) ₃ (CH ₃ COO) ₂ (HCOO) ₁	8.45	60.83	3.37	1.03
(CH ₃) ₁ (CH ₂) ₄ (C) ₁ (HCOO) ₂ (CH ₂ SH) ₁	6.03	60.60	3.14	1.02
(CH ₃) ₁ (CH ₂) ₁ (CH) ₂ (CH ₂ =CH) ₁ (CHO) ₁ (HCOO) ₁	6.51	58.04	3.34	0.98

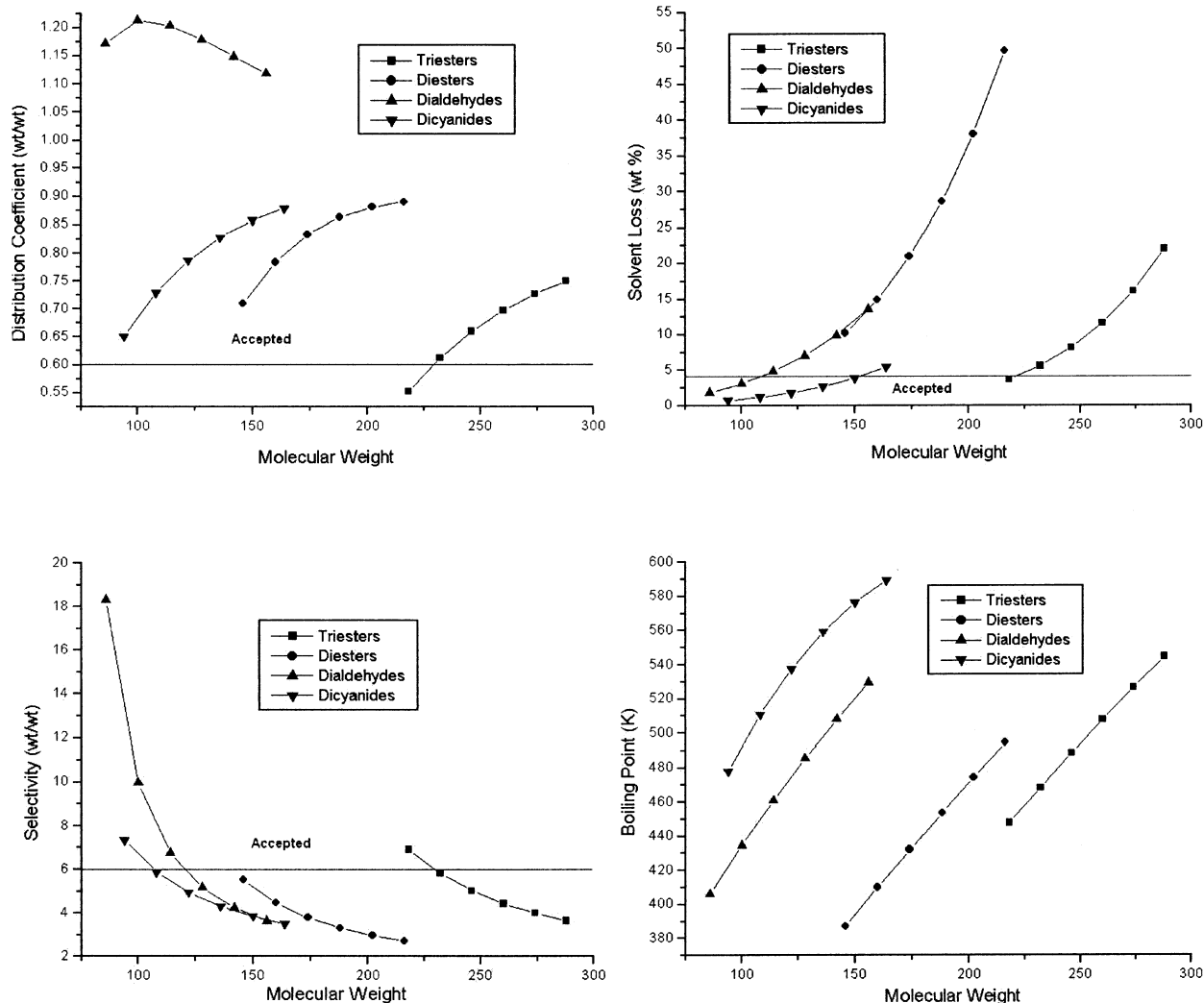


Figure 2. Performance of different solvent families with a single functional group for the benzene/hexane separation problem.

lent properties of the smallest triester (triacetin) for this application.

The UNIFAC thermodynamic model for the calculations of the activity coefficient is not a truly predictive method based on pure group parameters but requires information on the group binary interaction parameters. The availability of binary interaction coefficients in the model parameter tables^{22–24} is a limiting factor in the number and type of groups that can participate in a given synthesis process. Therefore, for the application of the program to the synthesis of solvents for nonideal mixtures (typical of liquid extraction or extractive distillation) using UNIFAC or other similar group contribution methods, the program screens out the groups for which binary information is not available in connection with the components participating in a particular separation problem and does the same at any stage of the synthesis procedure. This process leads to a considerable reduction of the size of the solvent molecular design problems. This size limitation is not present when a truly predictive method, based on pure group properties, is used. This has been the case in many of the computer-aided molecular design applications reported in the literature,^{2,8,9,13,25} which can lead to very large molecular design combinatorial problems. Also, the use of more fundamental ways of computing the molecular properties, like molecular simulations or “ab initio” models, will require different and more

advanced optimization procedures for computer-aided molecular design. The above example, in which 12 different groups were selected for the synthesis of molecular structures, shows that, even for a rather large problem, the synthesis of feasible molecules by group contribution using the present search algorithm is a fast, easy to implement, and reliable procedure.

Acknowledgment

The authors gratefully acknowledge financial support from the Argentine Research Council (CONICET) and the Universidad Nacional del Sur.

Appendix: Feasibility Criteria

J Attachments Balance. Let us consider a given set of K and J groups, which may also include tri- or tetravalent groups and can give different isomer structures. If we build a first intermediate structure only with the J groups, there will be two possibilities for each J attachment: to participate in a J – J bond or to be a J free attachment. Therefore, the balance would be

$$\sum_i j_i J_i = 2N_{JJ} + N_{JF} \quad \text{when } K \leq N_{JF} \quad (4)$$

The above intermediate J structure can only lead to a feasible final structure when there are no more K

groups than NJF in the original set. Otherwise, the remaining K groups must be inserted in the intermediate J structure. Therefore, there will be a third possibility for the J attachments: to be involved in a J–K bond. In this case, the balance becomes

$$\sum_i iJ_i = 2NJJ + NJF + 2(K - NJF) \quad \text{when } K > NJF \quad (5)$$

The decision of putting first the J groups together was completely arbitrary. However, this is just the less restrictive way for the structure feasibility analysis. Let us remember that computer-aided molecular design must provide structures that form at least one feasible isomer. Therefore, when evaluating feasibility, one should take into account the most promising structures among all possible combinations from a given set of groups.

Feasibility Criteria Derivation. Here the derivation for the expressions from Table 5 is presented.

For noncyclic structures, we start from the inequality (3).

$$K - 1 \leq NJJ \quad (\text{noncyclic}) \quad (3)$$

Then, multiplying by 2, taking 2NJJ from eq 4, when $K \leq NJF$, and replacing NJF by eq 6, we have after rearranging:

$$K \leq J_2 + J_3 + J_4$$

or by definition

$$\mathbf{K} \leq \mathbf{J}$$

When $K > NJF$, eq 5 is used instead of eq 4 and the corresponding expression is obtained:

$$2\mathbf{K} \leq \mathbf{J} + \mathbf{NJF}$$

Similarly, we obtain the same general criteria for cyclic structures from the inequality (2) and eq 7 instead of eq 6.

Notation (All of the Symbols Refer to Quantities in a Molecular Structure)

FMS = final molecular structure

IMS = intermediate molecular structure, i.e., the skeleton of the structure, without terminal groups

J_i = number of J groups with i attachments

J = number of J groups, $J = \sum_i J_i$

K_i = number of K groups with i attachments

K = number of K groups, $K = \sum_i K_i$

M = number of M groups

NFA = number of free attachments in the intermediate structure

NJF = number of J free attachments after the combination of J groups only

NJJ = number of bonds between two J groups

pre-FMS = prefinal molecular structure: the structure with two missing terminal groups

Literature Cited

(1) Stephanopoulos, G.; Townsend, D. Synthesis in Process Development. *Chem. Eng. Res. Des.* **1986**, *64*, 160.

(2) Joback, K. G.; Stephanopoulos, G. Designing molecules possessing desired physical property values. Proceedings of FOCAPD'89, Snowmass, CO, 1989.

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

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

(5) Cisondi, M.; Brignole, E. A. ECOFAC—Computer aided solvent design and evaluation in environmental problems, based on group contribution methods with association. In *Proceedings of the 11th European Symposium on Computer Aided Process Engineering*, Denmark, May 2001; Gani, R., Jorgensen, S. B., Eds.; Elsevier: New York, 2001; ISBN 0-444-50709-4, pp 375–380.

(6) 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.

(7) Odele, O.; Machietto, S. Computer aided molecular design: a novel method for optimal solvent selection. *Fluid Phase Equilib.* **1993**, *82*, 47.

(8) Venkatasubramanian, V.; Chan, K.; Caruthers, J. M. Computer aided molecular design using genetic algorithm. *Comput. Chem. Eng.* **1994**, *18*, 833.

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

(10) Gani, R.; Nielsen, B.; Fredenslund, Aa. A group contribution approach to computer-aided molecular design. *AIChE J.* **1991**, *37*, 1318.

(11) Markoulaki, E. C.; Kokossis, A. C.; Batzias, F. A. Novel chemicals for clean and efficient processes using stochastic optimization. *Comput. Chem. Eng.* **2000**, *24*, 705.

(12) Achenie, L. E. K., Gani, R., Venkatasubramanian, V., Eds. *Computer Aided Molecular Design: Theory and Practice*; Elsevier Publishers: New York, 2003.

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

(14) Fredenslund, Aa.; Gmehling, J.; Rasmussen, P. *Vapor liquid equilibria using UNIFAC*; Elsevier Scientific: Amsterdam, The Netherlands, 1977.

(15) Fredenslund, Aa.; Sorensen, J. Group Contribution Methods. In *Models for Thermodynamic and Phase Equilibria Calculations*; Sandler, S. I., Ed.; Marcel Dekker: New York, 1994; Chapter 4.

(16) Skjold-Jorgensen, S. Group contribution equation of state (GC-EOS): A predictive method for phase equilibrium computations over wide ranges of temperature and pressure up to 30 MPa. *Ind. Eng. Chem. Res.* **1988**, *27*, 110.

(17) Gros, H. P.; Bottini, S.; Brignole, E. A. A group contribution equation of state for associating mixtures. *Fluid Phase Equilib.* **1996**, *116*, 537.

(18) Espinosa, S.; Foco, G.; Bermúdez, A.; Fornari, T. Revision and extension of the group contribution equation of state to new solvent groups and higher molecular weight alkanes. *Fluid Phase Equilib.* **2000**, *172*, 129.

(19) Pöling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *The properties of gases and liquids*, 5th ed.; McGraw Hill Inc.: New York, 2000.

(20) Pretel, E.; Lopez, P.; Mengarelli, A.; Brignole, E. Prediction of physical properties for molecular design of solvents. *Lat. Am. Appl. Res.* **1992**, *22*, 187.

(21) Cisondi, M.; Díaz, M. S.; Espinosa, S.; Brignole, E. A. Molecular solvent design and near critical solvent optimization with ECOFAC. *Lat. Am. Appl. Res.* **2003**, *33*, 269.

(22) Magnussen, T.; Rasmussen, P.; Fredenslund, Aa. UNIFAC parameter table for prediction of liquid–liquid equilibria. *Ind. Eng. Chem., Process Des. Dev.* **1981**, *20*, 331.

(23) Hansen, H. K.; Rasmussen, P.; Fredenslund, Aa.; Schiller, M.; Gmehling, J. Vapor–liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352.

(24) Bastos, J. C.; Soares, M. E.; Medina, A. G. Infinite dilution activity coefficient prediction by UNIFAC group contribution. *Ind. Eng. Chem. Res.* **1988**, *27*, 1269.

(25) Maranas, C. D. Optimal computer-aided molecular design: a polymer design case study. *Ind. Eng. Chem. Res.* **1996**, *35*, 3403.

Received for review July 21, 2003

Revised manuscript received November 26, 2003

Accepted December 3, 2003