8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.1, June 6-8, 2007, Cancún, Mexico

MOLECULAR DESIGN OF ALTERNATIVE REFRIGERANTS USING GENETIC ALGORITHMS

Ekrem Oguz Ulker, Mehmet Yuceer and Ridvan Berber

Department of Chemical Engineering, Faculty of Engineering, Ankara University
Tandogan 06100, Ankara - Turkey

Abstract: Designing new molecules in desired properties involves time-consuming and expensive steps. Among the reported computer-aided techniques, knowledge based approaches suffer from acquisition of appropriate heuristics whereas mixed integer nonlinear programming formulations are not trivial to generate. Present study describes a new computer-aided refrigerant design using genetic algorithm. Genetic algorithms perform a guided stochastic search and allow improved solutions by sampling areas of the space having higher probability for good solutions. Results, very close to those previously reported by global optimization algorithms using MINLP formulation, illustrate the effectiveness of genetic algorithms to solve such computationally intensive problems with ease of us.

Keywords: Computer-aided design, genetic algorithms

1. INTRODUCTION

Due to the rapid technological changes and hard industrial competition, chemical processing industries need more efficient, cheaper and greener substances. The fact that in some areas we still use the substances with very low efficiency dictates the necessity of designing novel molecules with desired properties to be the alternatives for old ones. Eliminating side effects of drugs, producing more efficient catalysts or using more efficient coolants are just some examples. Conceptually, the problem of prescreening potential new molecules is very difficult due to its combinatorial nature. Combination numbers of elements or groups to design a novel molecule increase exponentially with the group or element number. Computer-aided techniques have been developed in order to supplement experimental techniques by prescreening large numbers of possible molecules to select fewer to generate and test.

One other difficulty associated with computer aided molecular design is due to the property estimation techniques. Although several formulae are available for estimating thermal, mechanical, and other physical properties of compounds (Reid *et al*. 1987; Horvath 1992), the number of functional groups whose properties can be estimated is still limited. Property prediction methods are typically nonlinear functions of atomic or group compositions. These nonlinearities lead to multiple local optima that may trap conventional search techniques in suboptimal parts of the search space.

A variety of methods including molecular models, group contribution methods, empirical models, correlations etc. have been developed to solve the computer aided molecular design (CAMD) problems. A point of focus in this are has been the commonly known problem of refrigerant design to replace the CFCs due to their relatively high ozone depletion potential. The refrigerant design problem was first introduced by Jobak and Stephanopoulos (1989, 1995), later tackled by Gani and Fredenslund (1993) by using knowledge based systems. Later Marculaki and Kokossis (1998) used stochastic optimization in the form of simulated annealing algorithm for designing refrigerants to replace Freon-12. The mixed integer nonlinear programming approaches have also been suggested for the formulation and solution of the problem. First were due to Duvedi and Achenie (1996, 1997) who created mixed integer

nonlinear programming models for single component refrigerant design. The locally optimal augmented penalty outer approximation algorithm was employed to solve the MINLP problem, with initial relaxation of the MINLP and subsequent solution of NLP and MILP problems. Most recently Sahinidis *et al.* (2003) presented a mixed integer nonlinear programming formulation employing implicit enumeration of a single branch-and-reduce tree.

Although these methodologies have advantages, they also suffer from some serious drawbacks. For complex and industrially relevant molecules, these approaches have limitations due to combinatorial complexity, nonlinear search spaces with local minimal traps, difficulties in knowledge acquisition, difficulties in dealing with the nonlinear structure property correlations, and problems in incorporating higher level chemical knowledge and reasoning strategies. MINLP models, although guarantied to be globally optimum pose immense difficulties in terms of model generation (Yuceer and Berber, 2005). The rigorous mathematical formulations create MILP models that may become unmanageable for large size problems. Considering this difficulty, a MATLAB tool was generated by Yuceer *et al.* (2006) for automatically creating the MILP formulation to tackle single stage multi-product scheduling problems in plants with parallel production units. Otherwise, manual creation of such formulation may become a difficult task for large size problems, and be prone to errors.

The field of molecular design first witnessed the first developments from the perspective of knowledge based systems, like "generate and test" algorithms (Joback and Stephanopoulos 1990, Gani *et al*. 1991), which were later replaced by optimization methods. However, optimization using mathematical programming (in NLP or MINLP form) could only be applied to relatively small size problems. On the other hand, stochastic optimization (genetic algorithms, simulated annealing) suffer from high computational times particularly with increased problem size. Furthermore, deterministic optimization usually involve fairly large number of discrete variables, are sensitive to initialization and formulation of constraints is not a trivial task. Thus, there exists a critical need to explore alternate strategies for molecular design that can circumvent these problems.

In this paper, we used genetic algorithm to solve computer aided refrigerant design problem with constraints. In the rest of the paper, the technique employed is described and the results are presented together with some comparison to earlier reports.

2. GENETIC ALGORITHMS

Evolutionary algorithms (EA) are based on Darwin's natural selection principle. Traditional deterministic search methods find the next sampling point using the problem features (such as gradients, Hessians, linearity and continuity). However, stochastic search approaches do not use these features, but randomly sample the search space. The evolutionary algorithm is one of the stochastic search methods. The state variables scope is scanned by means of evolutionary resembling operations (selection, crossover and mutation), which are applied on individuals in a population. The general idea behind genetic algorithms is the evolutionary creation of a new population of entities from an earlier generation through crossover and mutation processes and by passing on the fittest off springs to the next generation. In general, this approach is ultimately expected to lead to generations that become more and more fit through evolution thus achieving the desired design objective. The main operators are mutation and crossover. The mutation changes an individual to create a new one, while the crossover mixes two individuals and creates two new ones. An approach called *elitism* chooses best results before all genetic operations and puts it into the next offspring directly. Although genetic algorithms (GA) generally use binary representation, different notations can also be utilized. The population size is a parameter to control the optimization progress of GAs. There are no fixed rules for the population size when using a GA. Generally it must not be too low (lower than 10-20), because too much information will be lost in every generation and as a consequence the algorithm may converge too fast and end. Two general ways of dealing with constraints are penalizing infeasible solutions and rejection of infeasible individuals. There is a large body of theoretical and empirical evidence showing that, even for very large and complex search spaces, genetic algorithms can rapidly locate structures with high fitness ratings (Goldberg 1989; Davis 1991; Balku and Berber 2006).

3. DESIGNING ALTERNATIVE REFRIGERANT

The problem addressed in this work calls for the design of molecules with desired properties. Specifically, we are interested in the design of alternative refrigerants. Our goal is to develop a methodology that can serve as an aid to the design of replacement refrigerants by identifying candidates with satisfactory physical properties. Specifically, we are interested in generating more efficient refrigerant molecules than Freon (with commonly known name of R12), widely used in automotive or home refrigeration systems such as shown in Figure 1. Thus novel molecules that we are seeking through the suggested optimization technique must have a $\Delta H_{ve}/C_{pla}$ ratio bigger than that of R12. Our objective was to maximize this ratio. We assume that all novel molecules are ideal gas and there is no interaction between the groups in a molecule. The complete algorithm is shown as a flowchart in Figure 2. The search is an iterative procedure such that new molecules are generated and evaluated by the objective function. The suggested methodology is illustrated by a refrigerant design example.

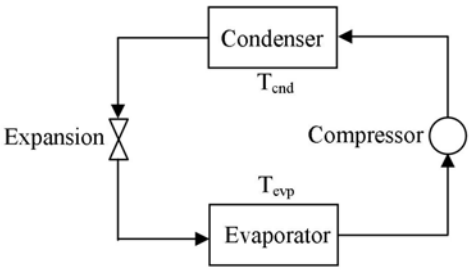

Fig. 1. A refrigeration system

3.1 Constraints

Novel refrigerant molecules must have a vaporization enthalpy bigger than or equal to that ΔH_{ve} of R12, which is 18.4 kJ/mol.K. By the same token, C_{pla} of new molecules must have heat capacity equal to or smaller than 32.2 cal/g mol.K. As has been mostly employed, we had also considered the system pressures (P_{vp}) at the evaporating temperature and at the condensing temperature were bound by the following constraints

$P_{vp}(T_{evp}) \geq 1.4$ bar, $P_{vp}(T_{cnd}) \leq 14$ bar

The sought refrigerant should have an evaporation temperature over 272.0 K a condensing temperature below 316.6 K. It is expected to have higher heat of vaporization and lower liquid heat capacity so that the amount of refrigerant required for a certain refrigeration job will be minimized. Therefore, the ratio of heat of vaporization to the liquid heat capacity is to be maximized.

4. METHODOLOGY

The genetic algorithm for searching optimum alternative refrigerant molecules was coded in MATLAB™ (The Mathworks, 1984). Functional groups, which were used to produce seed population, were selected from the list of 30 chemical groups shown in Table 1. Real number representation, which was thought to be more efficient than the binary digits representation, was used in coding. The number of functional groups to form a new molecule was random between lower and upper limits.

Table 1 Functional groups considered

1) —CH_3	9) CH≡	17) >CO	25) =N–
2) –CH_2–	10) –C≡	18) –CHO	26) –CN
3) >CH–	11) –F	19) –COOH	27) –NO_2
4) =CH–	12) –Cl	20) –COO–	28) –SH
5) >C<	13) –Br	21) =O	29) –S–
6) =CH_2	14) –I	22) –NH_2	30) –SH
7) =CH–	15) –OH	23) >NH	
8) =C=	16) –O–	24) >N–	

Decision was made by the algorithm before formation. Maximum number of groups in a molecule was limited by 20, and the minimum was 2. This number changes during the cross over process. The points where the groups will be detached and exchanged between molecules were also decided randomly. Provided that they remain between the upper and lower bounds, the number of groups may increase or decrease.

Seed population was obtained by selecting groups randomly to design a feasible molecule. During formation first group was selected from the pool of 30 groups under consideration. The second group was combined with the first one in compliance with the associated chemical bonds. The following groups were attached by the same manner consecutively. The only feasibility criteria considered here was to satisfy the chemical bonds. Thus, the procedure employed never allows an unfeasible chemical molecule in this respect, and a penalizing criteria was not needed.

Maximum number of iterations was set to be 10 000. A run took about 2 minutes of CPU time in a 3.2 GHz PC with 512 MB of RAM. In the course of calculations, the fitness of the generated molecules to the objective function was calculated by group contribution method (Joback and Reid, 1987).

4.1 Modeling of physical properties

The properties to be considered in the calculations are the critical temperatures, critical pressures and the normal boiling point temperatures. Equations 1 through 4 reflect a group contribution method due to Joback and Reid (1987) to calculate the boiling temperature T_b, critical temperature T_c, critical pressure P_c and ideal gas heat capacity at average temperature C_{p0a} respectively, of the newly synthesized compound. These thermal properties are used to derive other thermal properties in the sequel.

$$T_b = 198.2 + \sum_{i=1}^{N} n_i T_{bi} \quad (1)$$

$$T_c = \frac{T_b}{0.584 + 0.965\sum_{i=1}^{N} n_i T_{ci} - \left(\sum_{i=1}^{N} n_i T_{ci}\right)^2} \quad (2)$$

$$P_c = \frac{1}{0.113 + 0.0032\sum_{i=1}^{N} n_i a_i - \left(\sum_{i=1}^{N} n_i P_{ci}\right)^2} \quad (3)$$

Recall that T is a given constant. All other parameters in these equations are constants whose values wee taken from Joback and Reid (1987).

Using the above properties and the given operating temperatures of the cycle, Eqs. 5, 6, 7, and 8 were used to calculate the reduced boiling temperature T_{br}, the reduced average temperature T_{avgr}, the reduced condensing temperature T_{cndr}, and the reduced evaporating temperature T_{evpr}, respectively.

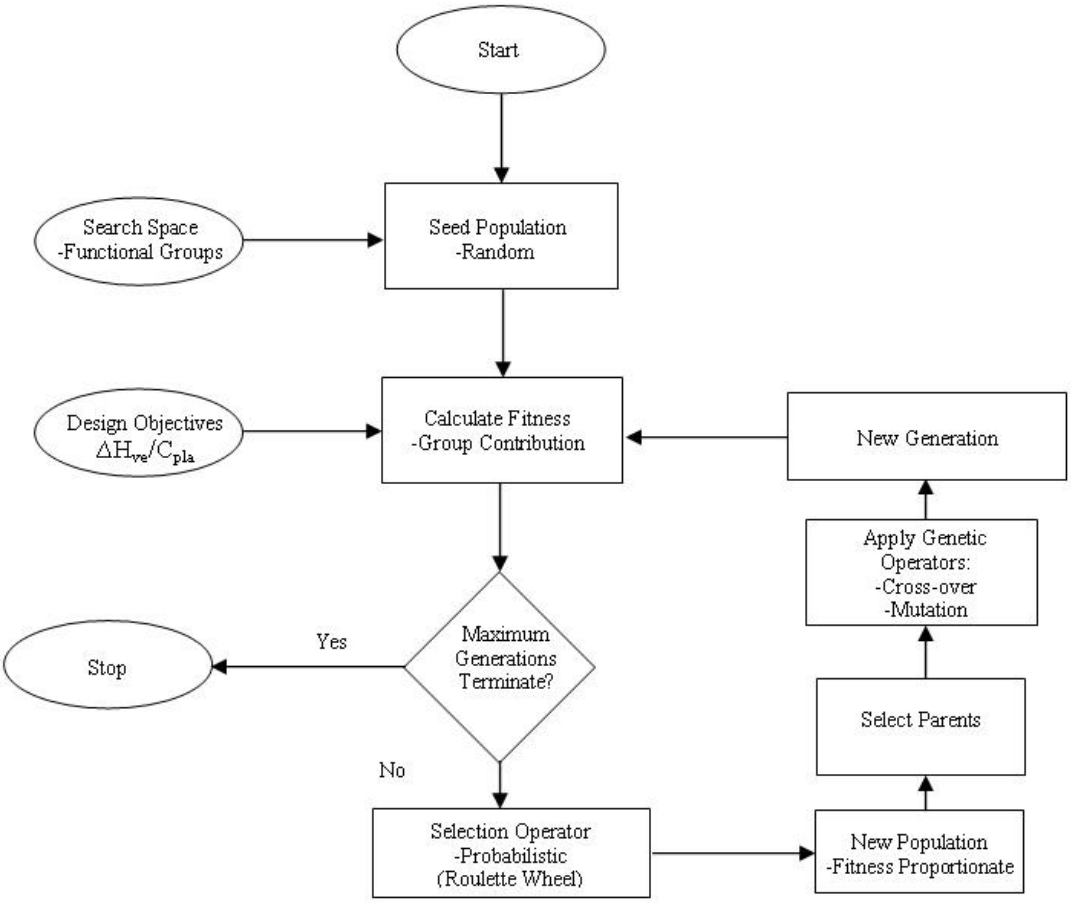

Fig. 2. Computational flowchart for genetic algorithm in refrigerant design problem

$$C_{p0a} = \sum_{i=1}^{N} n_i C_{p0ai} - 37.93 + \left(\sum_{i=1}^{N} n_i C_{p0bi} + 0.21 \right) T_{avg} + \left(\sum_{i=1}^{N} n_i C_{p0ci} - 3.91 \times 10^{-4} \right) T_{avg}^2 + \left(\sum_{i=1}^{N} n_i C_{p0di} + 2.06 \times 10^{-7} \right) T_{avg}^3 \quad (4)$$

$$T_{br} = \frac{T_b}{T_c} \quad (5) \qquad T_{avgr} = \frac{T_{avg}}{T_c} \quad (6)$$

$$T_{cndr} = \frac{T_{cnd}}{T_c} \quad (7) \qquad T_{evpr} = \frac{T_{evp}}{T_c} \quad (8)$$

After these steps acentric factor were calculated as follows (Sahinidis *et al.* 2003):

$$\alpha = -5.97214 - \ln\left(\frac{P_c}{1.013}\right) + \frac{6.09648}{T_{br}} + 1.28862 \ln(T_{br}) - 0.169347 T_{br}^6 \quad (9)$$

$$\beta = 15.2518 - \frac{15.6875}{T_{br}} - 13.4721 \ln(T_{br}) + 0.43577 T_{br}^6 \quad (10)$$

$$\omega = \frac{\alpha}{\beta} \quad (11)$$

Using the acentric factor and previously calculated values of average temperature, the liquid heat capacity C_p (Eq. 12) is derived from the ideal gas heat capacity using Rowlinson's modification of Bondi's equation (Sahinidis *et.al.* 2003).

The group contribution method of Joback and Reid (1987) was used to estimate the enthalpy of evaporation temperature ΔH_{vb} (Eq. 13) and vaporization at boiling temperature ΔH_{ve} (Eq. 14). After all that calculations we have the ratio of $\Delta H_{ve}/C_{pla}$ for every candidate molecules. At this point we put those molecules on a circle having the area proportionately fitting the objective function. So the most fitting candidate has the biggest area and the least fitting candidate has the smallest area as shown in Figure 3.

Wheel was randomly turned k times, k being the number of candidate molecules. In this process the most fitting candidate has bigger chance to be chosen whereas the least fitting candidate has the smallest.

$$C_{pla} = \frac{1}{4.1868} \left\{ C_{pa} + 8.314 \left[1.45 + \frac{0.45}{1 - T_{avgr}} + 0.25\omega \left(17.11 + 25.2 \frac{(1 - T_{avgr})^{\frac{1}{3}}}{T_{avgr}} + \frac{1.742}{1 - T_{avgr}} \right) \right] \right\} \quad (12)$$

$$\Delta H_{vb} = 15{,}3 + \sum n_i \Delta H_{vb,i} \quad (13)$$

$$\Delta H_{ve} = \Delta H_{vb} \left(\frac{1 - T_{evp} / T_c}{1 - T_b / T_c} \right)^{0{,}38} \quad (14)$$

So new set of candidates were created from the old ones. After this step genetic operators were used. First genetic operator was crossing-over. For crossing over, two molecules are chosen randomly from the new set, and two or more groups are taken from the chosen molecules and replaced. After that the second genetic operator, mutation, was performed. This mean that randomly selected group of a random molecule was changed with another randomly selected suitable group from groups considered.

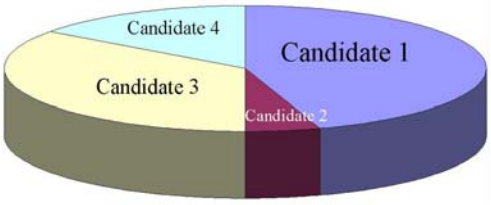

Fig. 3. A sample roulette wheel for selecting new candidates

This procedure resulted in a completely different and new set of candidate molecules. At this point we returned to the calculation step and continued on iteratively.

The algorithm ended, as reflected in Figure 2, when the number of iterations reached the pre-selected maximum.

Generated refrigerant molecules are shown in Table2. The last column in this table shows the order in which the molecules generated by the MINLP approach of Sahinidis *et al.* (2003) appeared. In both study, same group contribution method was used and almost same groups were chosen to produce novel molecules. Thus, compared to the global optimization approach, genetic algorithm performed very well. It is seen that most of the molecules seem to be suitable as candidates for refrigeration, eight out of thirteen being commercial molecules, namely formyl chloride, methyl chloride (R40), 1-propyne, methylacetaldehyde, allene, difluoramine and chlorodifluoromethane (R22). Among those, the molecules identified as number 9 and 13 fall within the group of CFCs. Their possible exclusion in the generation requires further constraints in terms of ozone depletion potential whose calculation for certain molecules is still limited.

From the optimization viewpoint, one of the main advantages of evolutionary computational techniques is that they do not have much mathematical requirements about the optimization problem; all they need is an evaluation of the objective function.

The results demonstrate that evolutionary algorithms, combined with property estimation techniques by group contribution method can be effectively used for molecular design of alternative refrigerants, a problem with known combinatorial difficulties. The algorithm presents ease of implementation and computational times considerably less than previous reports, because the evolutionary algorithms are known to be 0-order methods with almost only mathematical involvement of an objective function evaluation, they can handle nonlinear-problems with discrete, continuous or mixed-integer formulations. Thus, the algorithm can be utilized for similar molecular design problems like fuel additives, polymers, pharmaceuticals and solvents.

Table 2 Generated refrigerant molecules

#	Molecule	Structure	ΔH_{ve}	C_{pla}	P_{vpc}	P_{vpe}	$\Delta H_{ve}/C_{pl}$	Order in Sahinidis's Results
1	FNO	F – N = O	23.2849	18.0788	6.8854	1.5293	1.2880	1
2	CHClO	Cl–CH=O	26.4161	22.3792	10.1231	2.4538	1.1804	2
3	CH_3Cl	CH_3–Cl	21.5817	19.2367	6.2412	1.6078	1.1219	4
4	$CFONH_2$	O=CH–NH–F	28.4569	28.7643	7.7960	1.5430	0.9893	8
5	ClFO	Cl–O–F	20.8654	21.2431	7.1031	1.7272	0.9822	9
6	C_3H_4O	CH_3–CH=C=O	27.7733	28.8739	6.6663	1.5118	0.9619	11
7	C_3H_4	CH_3–C≡CH	21.4391	23.1079	6.3454	1.6886	0.9278	13
8	C_2F_2	F–C≡C–F	19.6080	21.2471	7.9169	2.1076	0.9229	14
9	$ClCH_2F$	Cl–CH_2–F	20.7080	22.7189	6.6507	1.6722	0.9115	15
10	C_3H_6O	CH_3–CH_2–CH=O	27.0480	31.0442	7.4916	1.7330	0.8713	21
11	C_3H_4	CH_2=C=CH_2	20.8490	24.0872	5.9674	1.5521	0.8656	24
12	NHF_2	F–NH–F	19.0570	22.5049	11.0137	2.7176	0.8468	27
13	CHF_2Cl	(F-)(F-)>CH–Cl	19.4542	23.0369	6.9970	1.7568	0.7770	32

Nomenclature

α : intermediate variable used to calculate acentric factor, ω
β : intermediate variable used to calculate acentric factor, ω
a_i : number of atoms of group *i*
C_{p0a} : ideal gas heat capacity at average temperature, cal/g-mol-K
Cp_{0ai} : constant contribution of group *i* to ideal gas heat capacity
Cp_{0bi} : first-order contribution of group *i* to ideal gas heat capacity
Cp_{0ci} : second-order contribution of group *i* to ideal gas heat capacity
Cp_{0di} : third-order contribution of group *i* to ideal gas heat capacity
C_{pla} : liquid heat capacity at average temperature, cal/g-mol-K
n_i : number of groups of type *i* selected
P_c : critical pressure, bar
P_{ci} : contribution of group *i* to critical pressure
T_{avg} : average temperature, T_{avg}=294.26 K
T_{avgr} : reduced average temperature
T_b : boiling temperature, K
T_{bi} : contribution of group *i* b_i to boiling temperature
T_{br} : reduced boiling temperature, K
T_c : critical temperature, K
T_{ci} : contribution of group *i* to critical temperature
T_{cnd} : condensing temperature, T_{cnd}=316.48 K
T_{cndr} : reduced condensing temperature
T_{evp} : evaporating temperature, T_{evp}=272.04 K
T_{evpr} : reduced evaporating temperature
ΔH_{vb} : enthalpy of vaporization at boiling temperature, kJ/g-mol
ΔH_{vbi} : contribution of group *i* to enthalpy of vaporization at boiling temperature
ΔH_{ve} : enthalpy of vaporization at evaporating temperature, kJ/g-mol
ω : acentric factor

6. REFERENCES

Balku, S. and R. Berber "Dynamics of an Activated Sludge Process with Nitrification and Denitrification: Start-up Simulation and Optimization using Evolutionary Algorithm" *Comp. Chem. Engng.* **30**, 490 2006.

Churi, N. and L.E.K. Achenie "On the Use of a Mixed Integer Nonlinear Programming Model for Refrigerant Design" *Int. Trans. Opl Res.* **4**, 45 1997.

Constantinou, L., and R. Gani ''New Group Contribution Method for Estimating Properties of Pure Compounds'' *AIChE J.*, **40**, 1697 1994.

Duvedi, A.P. and L.E.K. Achenie "Designing Environmentally Friendly Refrigerants using Mathematical Programming" Chem. Engng Sci. **51**, 3727 1996.

Gani, R., and A. Fredenslund ''Computer-Aided Molecular and Mixture Design with Specified Constraints,'' *Fluid Phase Equilibria*,. **82**, 39 1993.

Gani, R., B. Nielsen, and A. Fredenslund ''A Group Contribution Approach to Computer-Aided Molecular Design,'' *AIChE J.* **37**, 1318 1991.

Goldberg.,D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

Horvath, A. L., "Molecular Design", Elsevier Science Publishers B. V. Amsterdam (1992)

Joback, K. G., and R. C. Reid ''Estimation of Pure-Component Properties from Group-Contributions,'' *Chem. Eng. Comm.* **57**, 233 1987.

Joback, K. G., and G. Stephanopoulos ''Designing Molecules Possessing Desired Physical Property Values, *Proc. of 1989 Foundations of Computer-Aided Process Design Conf.* Snowmass, CO, Elsevier, Amsterdam 195 1990.

Marcoulaki, E.C. and A.C. Kokossis "Molecular Design Synthesis Using Stochastic Optimization as a Tool for Scoping and Screening" *Computers Chem. Engng.* **22**, S11-S18 1998.

The MathWorks Inc. Natick- MA. (1984) http://www.mathworks.com

Rowlinson, J.S., 1969. Liquids and Liquid Mixtures. 2nd edn., Butterworth & Co., London.

Sahinidis, N.V, M. Tawarmalani and M. Yu "Design of Alternative Refrigenrants via Global Optimization" *AIChE J*. **49**(7), 1761 2003.

Yuceer, M. and Berber, R. A Heuristic Approach for solution of MINLP problem for Production Scheduling in Preemptive Mode, *Chem. Engng. & Proc.* **44**, 933 2005.

Yuceer, M., Berber, R. and Ozdemir, Z. Automatic Generation of Production Scheduling Models in Single Stage Multi-Product Batch Plants: Some Examples, *Mathematical and Computer Modelling*, In press, 2007.