

# A METHODOLOGY FOR OPTIMAL SOLVENT DESIGN FOR REACTIONS

Milica Folić, Claire S. Adjiman<sup>\*</sup>, and Efstratios N. Pistikopoulos  
Centre for Process Systems Engineering, Department of Chemical Engineering  
Imperial College London, South Kensington Campus, SW7 2AZ London, UK

## *Abstract*

We present the foundations of a computer-aided system for the design of solvents for reactions. We propose the use of the multi-parameter solvatochromic equation (Abraham et al., 1987), which correlates a small number of solvent properties (empirical solvatochromic parameters and cohesive energy density parameter) with the logarithm of the reaction rate constant. For a given reaction, the proposed methodology consists of three steps. The first step is concerned with the development of a model of solvent effects on the reaction. It involves gathering the rate data for a small number of predetermined solvents and generating the solvatochromic equation by obtaining its coefficients from linear regression. The second step involves the formulation and solution of a mixed integer optimization problem, with the objective to identify a solvent in which the reaction rate under given conditions is maximized. Group contribution methods for the estimation of solvent properties have been developed for this purpose. The third step is a verification of the results obtained by experimentation or computational chemistry. This methodology is applied to t-Butyl Chloride solvolysis reaction.

## *Keywords*

Solvent design, Optimization, Solvatochromic equation, Group contribution methods, Reaction rate

## **Introduction**

In the chemical industry, solvents play an important role as transport agents, separation agents or as reaction media. Because of the high cost and time requirements of experimental testing, computer-aided systems for identifying promising solvent candidates are an attractive approach to the solvent design and selection problem.

Computer-Aided Molecular Design (CAMD) is a synthesis activity, which aims to identify a list of candidate molecules that perform a task most effectively (Gani and Brignole, 1983). Molecular Design methods are based on the fact that, from a small set of structural groups, a large number of molecules can be generated and evaluated with respect to a certain performance index.

In the past decade, a lot of work has been done in the development of CAMD approaches (Achenie et al., 2003). However, most of the work has focused on design of solvents to be used in separation processes and currently there is hardly any tool available for the design of solvents

to be used as reaction media even though the effect of solvent on reaction rate can be great and the reaction rate constant can vary by several orders of magnitude from one solvent to the other. Modi et al. (1996) have developed a predictive tool to estimate the effects of reaction rates in solution called the 'reaction fingerprint'. However, this approach applies only to predetermined solvent molecules and therefore would be most suitable for the verification of predicted optimal solvents. Due to the lack of systematic techniques, solvents are usually chosen based on intuition, previous knowledge, databases, and possibly some use of heuristic guidance, potentially leading to sub-optimal decisions with detrimental effects on performance and economy.

In this paper, a methodology is presented for design of solvents for reactions based on their effect on reaction rate constants. Currently, the methodology can be applied

---

<sup>\*</sup> To whom all correspondence should be addressed: c.adjiman@imperial.ac.uk

to one-step reactions. We consider reaction systems where solvents are inert i.e. their sole role is solvation.

The structure of this paper is as follows: first the methodology is introduced, second, a case study is presented and the conclusions are given in the last section.

## Methodology for solvent design

### Prediction of Reaction Rates

In order to be able to quantify solvent effects on reaction, the key issue is to relate solvent properties with reaction rate data. For this purpose we use the solvatochromic equation (Abraham et al., 1987) that correlates the logarithm of the reaction rate constant with solvent solvatochromic parameters and solvent cohesive energy density parameter. The solvatochromic equation is:

$$\log k = \log k_0 + s(\pi^* + d\delta) + a\alpha + b\beta + h\delta_H^2/100 \quad (1)$$

where:

$\log k$  is the logarithm of the rate constant;

$\pi^*$  is a solvatochromic parameter that scales with solvent dipolarity/polarisability;

$\alpha$  is a solvatochromic parameter that scales with solvent HBD (hydrogen-bond donor) acidity;

$\beta$  is a solvatochromic parameter that scales with solvent HBA (hydrogen-bond acceptor) basicity;

$\delta$  is a 'polarizability correction term' equal to 0.0 for nonchlorinated aliphatic solvents, 0.5 for polychlorinated aliphatics and 1.0 for aromatic solvents;

$\delta_H^2$  is the solvent cohesive energy density parameter in cal/cc, a measure of the solvent/solvent interactions that need to be overcome to create a cavity for the solute;

$s, d, a, b, h$  are coefficients that measure the relative susceptibility of  $\log k$  to the indicated solvent properties;

$\log k$  is a constant.

$\pi^*, \alpha, \beta, \delta$  and  $\delta_H^2$  are referred to as solvent parameters and  $s, d, a, b, h$  and  $\log k_0$  are referred to as reaction parameters.

This simple equation takes into account all the interactions in the system considered (in the case a solvent is inert and its only role is solvation).  $\pi^*$  is a descriptor of solvent-solute interactions by non-specific, long-range forces, while  $\alpha$  and  $\beta$  are descriptors of solvent-solute interactions by specific, short-range intermolecular forces. The cohesive energy density parameter  $\delta_H^2$  is a descriptor of solvent-solvent interactions.

Depending on the mechanism of the reaction and the structure of the transition state, some interactions can be more pronounced than others from one reaction to another. This can be judged by the statistical significance of the coefficients in the equation and the less significant terms

can be left out without affecting the quality of the predictions.

For a given reaction, it is necessary to know values of the rate constant and solvent parameters for a number of solvents in order to be able to obtain the reaction parameters in equation (1) by linear regression. Rate constant data are obtained from experiments, while solvent parameter values can be predicted by group contribution (GC) methods.

### Prediction of Bulk Solvent Properties

GC methods are based on the principles of transferability and additivity and are widely used for property prediction in CAMD. In order to make integration with existing CAMD approaches easy, we use the UNIFAC groups as building blocks for solvents.

All the properties are predicted directly with the exception of  $\delta_H^2$ . Its prediction relies on the prediction of the molar volume of the solvent,  $V_m$  and its enthalpy of vaporization,  $H_v$ , and it is defined as:

$$\delta_H^2 = 0.201 \left[ \frac{\Delta H_v - RT}{V_m} \right] \quad (2)$$

Correlations for predicting  $H_v$  and  $V_m$  were taken from the literature (Constantinou and Gani, 1994 and Constantinou et al., 1995).

Group contributions for  $\alpha$  and  $\beta$  were obtained by performing a regression on 215 solvents described by 41 first-order UNIFAC groups (Sheldon et al., 2004). The regression was performed using a statistical add-in in MS Excel spreadsheet program NAG (Numerical Algorithms Group) Ltd.

The predicted values for parameters  $\alpha$  and  $\beta$  are obtained from Eq. (3):

$$P = \begin{cases} \sum_i n_i P_i + P_0 & \text{if } \sum_i n_i P_i + P_0 > m \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where:

$P$  is the value of the property;

$P_i$  are the values of the contributions for group  $i$  used in the calculation of the value of the property;

$n_i$  is the number of groups of type  $i$  in the molecule;

$P_0$  and  $m$  are constants defined for the property of interest.

Group contributions for  $\pi^*$  were derived for the present work from 195 solvents described by 32 UNIFAC groups using  $m=0$  in Eq. (3).

### Optimization Formulation

The objective we consider is to maximize  $\log k$  as calculated from Eq. (1). The constraints involve GC methods for solvent parameters prediction and chemical feasibility constraints.

Chemical feasibility constraints include standard constraints on the maximum and minimum number of groups in the molecule, constraints on the maximum number of main and functional groups as well as constraints that forbid or limit occurrence of some groups together. The octet rule (Odele and Macchietto, 1993) and the bonding rule (as modified by Buxton et al., 1999) are included as well.

This formulation is a Mixed-Integer Nonlinear Programming (MINLP) problem. It is solved using the outer approximation algorithm (Viswanathan and Grossmann, 1990).

Integer cuts are included to allow the generation of successive 'best' solutions, giving a ranked list of candidate solvents.

The formulation is given as:

$$\begin{aligned}
 & \max \log k \\
 & \text{s.t. } \log k = \log k_0 + s(\pi^* + d\delta) + a\alpha + b\beta + h\delta_H^2 / 100 \\
 & \pi^* = \sum_i n_i \pi_i^* + 0.204 \\
 & \alpha = \begin{cases} \sum_i n_i \alpha_i & \text{if } \sum_i n_i \alpha_i > 0.15 \\ 0 & \text{otherwise} \end{cases} \\
 & \beta = \begin{cases} \sum_i n_i \beta_i + 0.1 & \text{if } \sum_i n_i \beta_i + 0.1 > 0.07 \\ 0 & \text{otherwise} \end{cases} \\
 & \delta_H^2 = \left[ \frac{\Delta H_V - RT}{V_m} \right] \\
 & V_m = \sum_i n_i V_{m,i} + 0.1211 \\
 & H_V = \sum_i n_i H_{V,i} + 6.335 \\
 & \sum_i (2 - v_i) n_i - 2m = 0 \\
 & n_j (v_j - 1) + 2m - \sum_i n_i \leq 0
 \end{aligned} \tag{4}$$

### Overall Methodology

1. First, we obtain the reaction coefficients in the solvatochromic equation for the reaction studied. In order to do this, rate data ( $\log k$  values) should be gathered. As there are six parameters (including the constant  $\log k_0$ ), the minimum number of solvents that must be used is six. Statistically, the results based on such a small set of solvents are not very reliable and the accuracy of predictions is low. The choice of the number of solvents to be used for linear regression is based on a trade-off between the cost of experiments to acquire the data and the statistical quality of the regression results. The chosen solvents should represent various classes of chemicals (e.g. an aromatic, a nitrate, an amide, an alcohol, a carboxylic acid, a halosubstituted compound) and be of different polarity (e.g. a nonpolar solvent, a polar aprotic solvent, a protic

solvent). To classify solvents based on their polarity we use the  $E_T^N$  solvent polarity scale (Reichardt, 1988). Here, we present two comparative cases with the coefficients in solvatochromic equation obtained based on linear regression in 8 solvents (Case 1) and 15 solvents (Case 2).

2. After obtaining the solvatochromic equation for the studied reaction we formulate and solve the CAMD optimization problem.
3. The final step provides a way to verify the solutions obtained and it results in a final ranking of solvents which can be used as reaction media for the reaction studied. Verification can be done by performing experiments to test the best solvents generated. It can also involve more detailed calculations (e.g. computational chemistry methods) for solvent parameter predictions.

### Case Study

The methodology presented is tested on t-Butyl Chloride solvolysis reaction. Solvolysis is a reaction that is induced by the solvent and the products are derived from it. The mechanism of t-Butyl Chloride solvolysis is thought to be a unimolecular nucleophilic substitution,  $S_N1$ .

There have been many kinetic studies of this reaction (Abraham, 1972, Abraham et al., 1981, Abraham et al., 1987, Gonçalves et al., 1992, Dvorko et al., 2002) and kinetic data were gathered for 38 solvents that can be built from the available UNIFAC groups.

Based on the regression performed for Case 1 (glycerol, propane-1,3-diol, diethylene glycol, acetic acid, dimethylacetamide, chlorobenzene, benzene and pentane), the following equation was obtained:

$$\log k = -14.9 + 7.72(\pi^* - 0.014\delta) + 8.03\alpha - 2.08\beta - 0.86\delta_H^2 / 100 \tag{5}$$

Based on the regression performed for Case 2 (glycerol, phenol, propane-1,3-diol, diethylene glycol, triethylene glycol, aniline, acetic acid, N-methylacetamide, ethanol, n-butanol, dimethylacetamide, 1,2-dichloroethane, tetrahydrofuran, benzene and pentane), the following equation was obtained:

$$\log k = -13.02 + 4.67(\pi^* + 0.025\delta) + 7.46\alpha - 2.37\beta + 0.13\delta_H^2 / 100 \tag{6}$$

Comparison of the statistical quality of the two regressions is given in Table 1, as well as the ranking of the first 13 solvents for both cases (Table 2).

Table 1. Statistics of Linear Regression

Case 1 (Eq. (5))	Case 2 (Eq. (6))
$R^2 = 0.96$	$R^2 = 0.81$
Adjusted $R^2 = 0.86$	Adjusted $R^2 = 0.7$
Standard Error = 1.54	Standard Error = 1.78

Table 2. Experimental vs. Predicted Solvent Ranking

Experimental ranking	Case 1 predicted ranking	Case 2 predicted ranking
1	1	1
2	4	3
3	12	5
4	11	7
5	3	4
6	8	6
7	2	2
8	10	9
9	13	13
10	9	10
11	6	8
12	5	11
13	7	12

Solving the optimisation problem results in a list of optimal solvents, the first five of which for each case are presented in Table 3.

Table 3. Five best ranked generated solvents

Case 1	Case 2
1,7-binitroheptane	1,2-binitroethane
2-methyl-1,6-binitrohexane	1,3-binitropropane
2,5-bimethyl-1,7-binitroheptane	1,4-binitrobutane
1,6-binitrohexane	1,5-binitropentane
2-methyl-1,7-binitroheptane	1,6-binitrohexane

It can be seen that for both cases the designed molecules belong to the same group of chemical compounds (binitro alkane derivatives). The statistics are better for the Case 1, however the solvent ranking is slightly worse. Still, based on our experience with this and other studies the number of solvents to be used for linear regression does not need to be greater than 8.

## Conclusions

An optimization framework to identify the optimal solvent molecule for a reaction is presented. We use a mathematically simple empirical equation that accounts well for all the interactions in the system. The algorithm described provides a guide for experimentation and can

very easily be integrated with other CAMD algorithms. We plan to complete the methodology by implementation of a verification step and then proceed to the integration of this solvent design approach in plant-wide design strategy.

## Acknowledgments

The authors gratefully acknowledge the ORS scheme for financial support.

## References

- Abraham, M. H. (1972). Substitution on Saturated Carbon. Part XIV. Solvent Effects on the Free Energies of Ions, Ion-pairs, Non-electrolytes, and Transition States in Some  $S_N$  and  $S_E$  Reactions. *J. Chem. Soc. – Perkin Trans. 2*, 1343.
- Abraham, M. H., Taft, R. W., Kamlet, M. J. (1981). Linear Solvation Energy Relationships. 15. Heterolytic Decomposition of the tert-Butyl Halides. *J. Org. Chem.*, 46, 3053.
- Abraham, M. H., Doherty, R. M., Kamlet, M. J., Harris, J. M., Taft, R. W. (1987). Linear Solvation Energy Relationships. Part 37. An Analysis of Contributions of Dipolarity-Polarisability, Nucleophilic Assistance, Electrophilic Assistance, and Cavity Terms to Solvent Effects on t-Butyl Halide Solvolysis Rates. *J. Chem. Soc. – Perkin Trans. 2*, 913.
- Achenie, L. E. K., Gani, R., Venkatasubramanian, V., Eds. (2003). Computer Aided Molecular Design: Theory and Practice. *Elsevier*. Amsterdam.
- Buxton, A., Livingston, A. G., Pistikopoulos, E. N. (1999). Optimal Design of Solvent Blends for Environmental Impact Minimization. *AIChE J.*, 45, 817.
- Constantinou, L., Gani, R. (1994). New Group Contribution Method for Estimating Properties of Pure Compounds. *AIChE J.*, 40, 1697.
- Constantinou, L., Gani, R., O’Connell, J. P. (1995). Estimation of the Acentric Factor and the Liquid Molar Volume at 298 K Using a New Group Contribution Method. *Fluid Phase Equilib.*, 103, 11.
- Dvorko, G. F., Zaliznyi, V. V., Ponomarev, N. E. (2002). Kinetics and Mechanism of Monomolecular Heterolysis of Commercial Organohalogen Compounds: XXX. Correlation Analysis of Solvation Effects in Heterolysis of tert-Butyl Chloride. *Russian Journal of General Chemistry*, 72, 1549.
- Gani, R., Brignole, E. (1983). Molecular Design of Solvents for Liquid Extraction Based on UNIFAC. *Fluid Phase Equilib.*, 13, 331.
- Gonçalves, R. M. C., Simões, A. N. M., Leitão, R. A. S. E., Albuquerque, L. M. P. C. (1992). Correlation of Rate Constants for the Solvolysis of tert-Butyl Halides. Effect of Temperature. *J. Chem. Research (S)*, 330.
- Modi, A., Aumond, J. P., Mavrovouniotis, M., Stephanopoulos, G. (1996). Rapid Plant-Wide Screening of Solvents for Batch Processes. *Comput. Chem. Eng.*, 20, S375.
- Nag Statistical Add-Ins for Excel. [www.nag.com](http://www.nag.com).
- Reichardt, C. (1988). Solvents and Solvent Effects in Organic Chemistry. *VCM Publishers*. UK.
- Sheldon, T., Adjiman, C. S. and Cordiner, J. L. (2004). Pure Component Properties from Group Contribution. *in preparation*.
- Viswanathan, J., Grossmann, I. E. (1990). A Combined Penalty Function and Outer Approximation Method for MINLP Optimisation. *Comput. Chem. Eng.*, 24, 769.