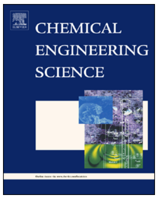
ARTICLE IN PRESS

Chemical Engineering Science ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces



# Computer-aided screening of solvents for optimal reaction rates

Danan S. Wicaksono, Adel Mhamdi, Wolfgang Marquardt*

*AVT – Process Systems Engineering, RWTH Aachen University, Turmstr. 46, Aachen 52056, Germany*

## HIGHLIGHTS

- A hybrid model-based and data-driven screening of reaction solvents is proposed.
- The proposed screening is robust with respect to uncertainty in experimental data.
- The strengths of the proposed screening are shown using real experimental data.

## ARTICLE INFO

*Article history:*
Received 24 July 2013
Received in revised form
22 November 2013
Accepted 5 December 2013

*Keywords:*
Solvatochromic equation
$S_N1$ reactions
Optimal design of experiments
Tikhonov regularization
Uncertainty propagation
Optimization under uncertainty

## ABSTRACT

A hybrid model-based and data-driven framework for the screening of promising solvents is suggested to optimize reaction rates. The framework comprises a sequence of two connected problems: (i) identification of model to predict solvent effects on reaction rate constants from experimental data and (ii) computer-aided screening exploring a databank of solvents. The resulting problems are formulated as systems of linear equations which can be solved by standard numerical linear algebra packages. In light of the uncertainty inherently presents in experimental data, a combination of Tikhonov regularization and optimal design of experiments (or data selection) is proposed to remedy uncertainty amplification from the data to the solution and circumvent unreliable screening. The results obtained using the proposed strategy are compared with the benchmark solvent selection procedures. They are shown to be in good agreement with experimental data, in relative as well as in absolute terms, for the investigated case study, i.e., the solvolysis of tert-butyl chloride, which belongs to the class of $S_N1$ reactions.

## 1. Introduction

Solvents are widely used in the chemical and process industries (Wypych, 2001) for a wide range of applications. These include dissolution of solid materials, dilution of liquids, separation, such as azeotropic distillation and liquid–liquid extraction, crystallization as well as moderating or facilitating chemical reactions in liquid phase. Watson (2012) claimed that the proper selection of solvents is one of the most important decisions regarding the implementation of green chemistry in an industrial setting. The focus of this work is on the selection of solvents used as reaction media.

Menschutkin already discovered in the late 19th century that the reaction rate of the quaternization of triethylamine with iodoethane varies for all of the 23 different investigated solvents (Menschutkin, 1890). The reaction rate of alkyl halide solvolysis[1] has also been shown to depend on the selected solvent (Grunwald and Winstein, 1948). These early works marked the beginning of many studies on the effect of solvents on the rate of chemical reactions including the more recent works of McManus et al. (2004), Fajt et al. (2008), Ngaosuwan et al. (2010), Harifi-Mood et al. (2011), Vanoye et al. (2011) and Akpa et al. (2012). A comprehensive review of the subject is provided by Reichardt and Welton (2011) in their recent monograph. These authors present many cases where the reaction rate constant varies significantly for different solvents. For example, the reaction rate constant of the solvolysis of 2-chloro-2-methylpropane (or tert-butyl chloride) is more than $10^5$ times higher in water than in ethanol at 25 °C and atmospheric pressure (Reichardt and Welton, 2011).

Hence, the selection of the solvent for a particular chemical reaction can either make or break a chemical process with respect to technical or economic viability. Consequently, there is a need for predictive methods which support the identification of the most promising solvent for a particular chemical reaction (Gani et al., 2005, 2008; Folic et al., 2007, 2008a, 2008b). Two strategies to solve such problems can be distinguished: either they rely on the screening of extensive databanks compiling candidate molecules and the physico-chemical properties of interest or they build

---

* Corresponding author. Tel.: +49 241 80 94668; fax: +49 241 80 92326.
*E-mail address:* wolfgang.marquardt@avt.rwth-aachen.de (W. Marquardt).

[1] The terminology “hydrolysis“ can be used instead if the solvent is water.

0009-2509/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ces.2013.12.006

Please cite this article as: Wicaksono, D.S., et al., Computer-aided screening of solvents for optimal reaction rates. Chem. Eng. Sci. (2014), http://dx.doi.org/10.1016/j.ces.2013.12.006

candidate molecules from molecular building blocks and use a model to predict the properties of interest from molecular structure. The latter is a special kind of a computer-aided molecular design (CAMD) problem (Achenie et al., 2003).

Gani et al. (2005) suggested a databank screening approach to perform a rule-based selection of suitable solvents for a class of organic reactions. The heuristic rules applied were derived from industrial practice. In a later paper, Gani et al. (2008) extended their previous work to multi-step reaction systems and to solvent substitution problems for common reaction steps in an existing chemical process. As in their previous work, the selection criteria were based on physical properties like solubility of the reactants in the solvent or boiling point of the solvent, but did not take the effect of the solvent on reaction kinetics into account.

Rather than relying on experimental data compiled in a databank, predictive models can be used to quantify the set of physico-chemical properties of interest from molecular structure. Group contribution methods, like UNIFAC (Fredenslund et al., 1975), or quantitative structure-property relationships relying on molecular descriptors (Todeschini and Consonni, 2009) have been widely used in CAMD. These prediction methods can either be used in a generate-and-test mode or in a synthesis mode. In the generate-and-test mode, a list of molecules is constructed from molecular building blocks and their performance is subsequently tested by predicting physico-chemical properties. This is done with a group contribution method as originally introduced by Gani and Brignole (1983) for solvent selection in liquid–liquid extraction. In the synthesis mode, an optimization-based method is used to solve the CAMD problem. Macchietto et al. (1990) reformulated the generate-and-test method as a constrained optimization problem to avoid explicit enumeration. Later, Odele and Macchietto (1993) introduced a mixed-integer nonlinear programming (MINLP) problem formulation to decide on a favorable molecular structure by combining suitable members from a set of molecular building blocks. In particular, an objective function is maximized or minimized while satisfying a set of constraints to obey structural feasibility, such as the octet rule. Folic et al. (2007, 2008a) were the first to use group contribution methods to predict physico-chemical properties from molecular structure as part of an optimization-based approach to find existing or novel reaction solvents resulting in optimal reaction rates in single and multiple reactions. Struebing et al. (2010, 2011) further developed this method by including more refined constraints on the structure of the target molecule. Weis and Visco (2010) demonstrated the alternative use of molecular descriptors in CAMD.

The property prediction problem in CAMD can also be addressed by more sophisticated quantum chemistry methods. For example, Peters et al. (2008) and Spiess et al. (2008) employed COSMO-RS (Eckert and Klamt, 2002) to predict reaction equilibria and to screen solvents for increasing conversion in biocatalytic two-phase reaction systems. Since these methods are often not of sufficient accuracy, Lapkin et al. (2010) suggested to improve the predictive capabilities of COSMO-RS by calibrating the model with some experimental data. More rigorous quantum chemical calculations are under investigation. Stanescu and Achenie (2006), for example, reported some density functional theory (DFT) results on reaction rate enhancement of Kolbe–Schmitt reactions by appropriately chosen solvents. However, these *ab initio* methods are not yet sufficiently predictive to apply them in the context of solvent selection to optimize the reaction rate constant.

This contribution aims at a practical and sound approach to the rational selection of solvent candidates from a databank which is expected to result in maximum reaction rate constant. Nevertheless, the approach can be easily adapted for cases of minimizing as well as targeting the desired reaction rate constant. It relies on databank screening combined with a well-established prediction model calibrated by means of experimental data to quantify the reaction rate enhancement for a given solvent.

The proposed method is based on two distinct steps. In the first step, a suitable model for the prediction of the solvent effects is identified from data and prior knowledge. Following a similar direction as Folic et al. (2007), the solvatochromic equation is employed to predict the effect of a solvent on the reaction rate constant. This model structure has to be validated and the unknown parameters have to be estimated from experimental data. In the second step, the resulting model is employed for the computer-aided (or virtual) screening of a set of reaction solvent candidates compiled in a databank (Walters et al., 1998). The process of exploring the databank is guided by a so-called scoring function. The scoring function, e.g. the reaction rate constant predicted by the solvatochromic equation in this work, assigns a score to each solvent to manifest its potency, e.g., the higher is the score, the more potent is the solvent. Since the method relies on experimental data, the effect of the uncertainty in the data on solvent screening needs special attention.

This contribution focuses on the interplay between the quality of the experimental data and the solution of the model identification and the computer-aided screening problems. The uncertainty in the data is propagated to the solution of both of these problems, and small uncertainty in the experimental data may result in significantly larger uncertainty in the screening solution. In the worst case, the most favorable solvent can be buried by many far inferior solvents as a result of uncertainty propagation.

The paper is organized as follows. The mathematical formulation of the problem is presented in Section 2. The explicit solution of the problem is derived and treatments for handling uncertainty are proposed in Section 3. Section 4 presents an illustrative case study and Section 5 provides the results and their discussions. Finally, Section 6 concludes this contribution.

## 2. Problem formulation

The solvatochromic equation has been chosen to model solvent effects on reactions because of its simplicity and predictive capabilities: it allows an efficient screening of a massive databank since it is merely a simple linear algebraic equation. It has been shown to properly predict a variety of solvent effects, including the effects on the reaction rate constant, based upon observations of only a few solvents (Reichardt and Welton, 2011). There is a wide range of solvents which are covered by the solvatochromic equation. They include traditional organic solvents as well as more sophisticated solvents, such as ionic liquids and switchable solvents (Jessop et al., 2012).

The solvatochromic equation, also called Kamlet–Taft equation or linear solvation energy relationship (LSER), deciphers multiple solvent effects on the reaction rate constant $k$ through linear combinations of the solvent properties, i.e., the hydrogen-bond donor acidity $\alpha$, the hydrogen-bond acceptor basicity $\beta$ and dipolarity/polarizability $\pi^*$ (Kamlet et al., 1977), which are commonly called the Kamlet–Taft parameters. The originally proposed equation has been generalized later (Kamlet et al., 1983). Kamlet et al. (1987) discussed the solvatochromic equation from an empirical and a theoretical perspective. Harris et al. (1988) argued that the solvatochromic equation provides insight into the reactivity of a particular reaction. From a mathematical point of view, the solvent properties are essentially the factors of a multiple linear regression model. These factors can also be regarded as molecular descriptors (Todeschini and Consonni, 2009) elucidating selected solvent effects.

Please cite this article as: Wicaksono, D.S., et al., Computer-aided screening of solvents for optimal reaction rates. Chem. Eng. Sci. (2014), http://dx.doi.org/10.1016/j.ces.2013.12.006