Systematic methods for solvent design: Towards better reactive processes

Eirini Siougkrou

A thesis submitted for the Doctor of Philosophy degree of Imperial College London and for the Diploma of Imperial College

The Department of Chemical Engineering
South Kensington Campus
Imperial College London
London SW7 2AZ
United Kingdom

13 February 2014

Abstract

The focus of this thesis is the development of novel methodologies for systematic identification of optimal solvents for chemical reactions. Two aspects are considered: the integrated solvent and process design using a mixed solvent, and the design of an optimal solvent using *ab initio* methods that do not rely on experimental data.

A methodology is developed for the integrated design of a CO_2-expanded solvent in a reaction process. Posing as objective function the cost of the process, for a defined production rate, an optimisation problem is formulated, with decision variables that include the organic co-solvent, the composition and the mass of the mixed solvent. Emphasis is placed on the prediction of the reaction rate, for which the solvatochromic equation combined with a preferential solvation model are used, and on solid-vapour-liquid phase equilibrium, for which the group-contribution volume translated Peng-Robinson equation of state is used. The proposed methodology is applied to the Diels-Alder reaction of anthracene and 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD), and three CO_2-expanded solvents are considered (acetone, acetonitrile and methanol). Acetonitrile and acetone are found to offer good performance over a range of CO_2 concentrations. The importance of taking into account multiple process performance indicators, when designing gas-expanded liquids, is highlighted.

As a further step toward systematic solvent design approaches that are not limited by the availability of experimental data and consider a large number of candidate solvents, an *ab initio* methodology is developed for the design of optimal solvents for reactions. The developed method combines quantum mechanical calculations with a computer-aided molecular design formulation. In order to limit the number of QM calculations but also retain accuracy and ensure convergence, the Kriging approach

is used. Kriging is a response surface approach, which has recently attracted a lot of attention because it is an exact extrapolator with a statistical interpretation which makes it stand out from other methods. The proposed approach is used successfully to identify promising solvents for the Menschutkin reaction of phenacyl bromide and pyridine and the Cope elimination of methylamine oxide. The use of Kriging as the surrogate model is found to lead to improved solvents when compared to the simpler solvatochromic equation used in previous work.

I, Eirini Siougkrou, declare that the work in this thesis is my own and that work from others has been referenced appropriately.

Acknowledgements

I am really grateful to my supervisors Claire Adjiman and Amparo Galindo for their support, help and guidance throughout these years. I would like to thank Terrence for helping me fight (and win!) the computer demons. I feel privileged to be a member of the MSE group and I would like to thank everyone in the group for all the good moments we had together and especially Apostolis, Olga, Thomas, Vassilis, Esther, Christianna, Katerina, Zara, Panos, Carlos, Niall, Erik. Thank you to my dear κουμπάρους Xenia and George for believing in me and supporting me. Thank you also to Vaso and Olga for always standing by me, understanding me and being happy for me. Nothing would have been possible without my parents' love and support and I want to thank them for always worrying about me and being there for me. Finally, the truth is that this work would not have been completed if it wasn't for the invaluable contribution of my partner Manolis; thank you for putting up with me!

Στους γονείς μου,
Γιολάντα και Κώστα

Contents

List of Tables

List of Figures

Chapter 1

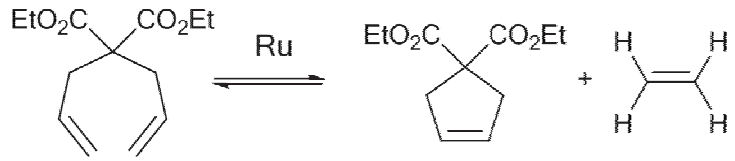
Introduction

1.1 Solvent effects on chemical reactions

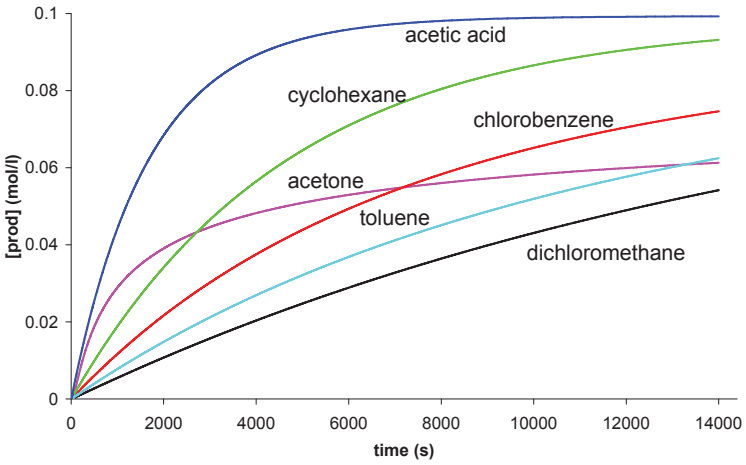
A solution is a homogeneous liquid phase that consists of more than one component, and the component that is in excess is called the solvent [Reichardt, 2007]. Solvents have attracted a lot of interest for thousands of years. Water for example was the first substance to be considered as a solvent in ancient times by the Greek philosophers and is still the most popular solvent [Buncel et al., 2003, Reichardt and Welton, 2011]. Alchemists and chemists from the 15^{th} to the 18^{th} century were looking for a universal solvent, which would dissolve all substances, the so-called "Alkahest", which was never found. Nowadays many solvents are used in a large number of industrially important processes, such as absorption, crystallisation, CO_2 capture, chemical reactions, etc.

Chemical reactions can take place in the gas, liquid, or solid state, but, the majority take place in solution. The choice of solvent can have significant effects. For example, solvents can affect the reaction rate and selectivity, help to control the reactor temperature, transport the reactants/products in a process or separate the products [Chipperfield, 1999]. The effect of solvents on chemical reactions was noted for the first time by Berthelot and de Saint-Gilles [1862], while studying the esterification of acetic acid with ethanol. A few years later, Menschutkin [1890a,b] studied the reactions between trialkylamines and haloalkanes and noted that solvents *can greatly influence the course of reactions* [Bruylants, 1977]. Solvent effects on reactions and

other processes have been widely investigated, mainly experimentally, throughout the years and a good understanding of the underlying chemistry and physics has been attained [Reichardt and Welton, 2011]. A characteristic recent example is the work



(a)



(b)

Figure 1.1: (a) Grubbs II (Ru) catalysed ring-closing metathesis reaction of diethyl diallylmalonate (diene) to form the corresponding cyclopentene (prod) and ethylene. (b) Simulated concentration of the product of the Grubbs II catalysed ring-closing metathesis reaction, based on ^{1}H NMR spectroscopy data at 298 K, with 0.1 M diene and 0.00042 M catalyst. Figure adapted from Adjiman et al. [2008].

by Adjiman et al. [2008] on a ring-closing metathesis reaction catalysed by the 1,3-dimesityl-4,5-dihydroimidazol-2-ylidene ruthenium complex (Ru) (figure 1.1a). The reaction rate varies considerably according to the solvent, as can be seen in figure 1.1b. The effects of the solvent here are multiple; the rate constant, the solubility of the catalyst and the deactivation of the catalyst are all altered by the nature of the solvent. For example, although the reaction proceeds quite fast in acetone, the catalyst deactivates rapidly so that the overall rate tapers off quickly and complete conver-

sion cannot be reached. On the other hand, dichloromethane leads to low reaction rate, but no catalyst deactivation occurs so that complete conversion can eventually be reached. In cyclohexane, where the rate is very high and no deactivation of the catalyst is observed, the solubility of the catalyst is very low. This effect cannot be observed on figure 1.1b, but would lead to large reactor volumes for a given production rate. Therefore, selecting a good solvent for a given reaction in order to achieve good performance is both important and challenging. In addition, when considering that there are between 250 and 300 [Marcus, 2002] commonly available solvents in academia and industry, the solvent selection problem is by no means trivial.

The importance of research on solvent effects is further heightened because millions of tons of solvents are used in industrial processes annually and their impact on the environment and on energy consumption cannot be neglected. For example, in the pharmaceutical industry, 20 million tons of volatile organic compounds (VOCs) are released per year and solvent use is responsible for the 50% of green-house gas emissions from typical pharmaceutical processes [Jiménez-González et al., 2005a]. Furthermore, solvents have been found to be responsible for 60% of the energy used in the production of an active pharmaceutical ingredient [Jiménez-González et al., 2005a]. Thus, the need to not only to minimise the amount of solvent used in industrial processes, but also to search for alternative, more environmentally friendly solvents, is pressing. According to Jiménez-González et al. [2011], solvent selection or optimisation is one of the major green engineering research areas for sustainable manufacturing, as there is much room for improvement.

In general, the choice of solvent for a particular purpose is made based on experience and insight [Gani et al., 2005]. It is interesting that the most widely used solvents for the metathesis reaction in figure 1.1a are dichlroromethane and toluene [Adjiman et al., 2008], but it can be seen in figure 1.1b that in these solvents the reaction rate is the lowest, compared to the other solvents tested. The ability to predict the performance of a reaction in a solvent and to choose the most appropriate solvent for a reaction are very important tasks and, at the same time, complex challenges, as testified by the large number of researchers who focus their research on the effects

of solvents on chemical reactions [Buncel et al., 2003]. However, solvent selection for improved performance has been studied mainly experimentally and only in the 1990s did computational design approaches emerge, when the concept of *computer-aided molecular design* (CAMD) appeared. In CAMD methods, the objective is to design a molecule that best matches a number of given requirements, such as desired properties, and maximises performance measures. There has been significant progress on CAMD [Achenie et al., 2003] and there are many notable works on CAMD of solvents for various processes, such as liquid-liquid extraction [Karunanithi et al., 2005, Lek-Utaiwan et al., 2008], cleaning in lithographic processes [Sinha et al., 1999, 2003b], and crystallisation [Samudra and Sahinidis, 2013]. A few attempts have been made also to tackle the problem of solvent design for chemical reactions [Gani et al., 2005, Folić et al., 2007, 2008a, Struebing et al., 2013], while the problem of designing solvent mixtures for reactions has not been considered yet. Gani et al. [2005] considered a number of solvent performance criteria (e.g. solubility, selectivity, environmental, health and safety properties, kinetics, etc.) based on databases for the reactions. There has been significant progress in including predictive models for reaction kinetics in solvent design. Folić et al. [2007] and Struebing et al. [2013] considered kinetics in their CAMD approaches, using a linear empirical model to connect the reaction rate constant with the solvent properties, based either to experimental data [Folić et al., 2007], or on *ab initio* predictions [Struebing et al., 2013] for the rate constant. The use of a linear model proved to be successful in a few case studies, but in general solvent effects are not linear [Buncel et al., 2003, Reichardt and Welton, 2011]. It is evident that it is necessary to go beyond the linear model in order to increase the reliability of the predictions.

1.2 Scope

The development of systematic, novel methodologies for the design of the optimal solvent that maximises the performance of a given chemical reaction is the subject of this thesis. In order to achieve this, a number of requirements have to be fulfilled.

The formulation of the problem must take into account a large number of solvent candidates, as well as a large number of performance metrics and constraints. It would be also ideal not to rely on the availability of experimental data, to enable the general use of the methodology in the early stages of process design. Thus, the use of an accurate kinetic model is necessary to achieve reliable predictions. Last but not least, computational efficiency is always a target, especially in solvent design where the number of possible solvents is very large.

Two scenarios, where CAMD is used to select the best solvent, are considered in this thesis: (a) the design of a solvent mixture, where a methodology for the design of a gas-expanded liquid for a reaction is developed, which takes also into account process criteria, and (b) the design of the optimal solvent for a reaction, where a methodology is developed in which no experimental data are required. The kinetics in this novel methodology are calculated at the quantum mechanical level of accuracy, while maintaining computational efficiency.

1.3 Outline

The outline of this thesis is as follows. In Chapter 2, the main computer-aided molecular design approaches and several key applications mainly in the area of solvent design are presented. Emphasis is placed on computer-aided molecular design of mixtures and of solvents for reactions. In Chapter 3, a methodology for the integrated design of a solvent mixture, particularly a gas-expanded liquid for a reactive process is introduced. The design mainly focuses on kinetics and thermodynamics, as well as process criteria, and it is applied to a Diels-Alder reaction. In Chapter 4, several widely-used methods for the prediction of reaction rate constants are briefly reviewed and an expression for the reaction rate constant using conventional transition theory and a continuum solvation model is derived. The aforementioned expression is used for the prediction of the rate constant of a Menschutkin reaction in several solvents. In Chapter 5, a detailed formulation of the solvent design problem for reactions is presented, with integrated quantum mechanical calculations for the reaction rate

constant. An algorithm for the solution of the problem is developed and applied to the same Menschutkin reaction. In Chapter 6, an *ab initio* solvent design methodology for reactions is introduced, where a nonlinear surrogate model, Kriging, is used to limit the number of quantum mechanical calculations. The approach is successfully applied to the Menschutkin reaction and a Cope elimination reaction. Finally, in Chapter 7, the conclusions of the thesis are presented and future directions are discussed.

Chapter 2

Computer Aided Molecular Design

2.1 Introduction

Computer aided molecular design (CAMD) appeared as a concept in the early '80s [Gani and Brignole, 1983] and it has since been developed both in theory and practice. CAMD can be defined [Achenie et al., 2003, Giovanoglou et al., 2003] as the formulation where *"Given a set of building blocks and a specified set of target performance measures, determine the molecule or molecular structure that matches this index."* The performance measures can either focus on the molecule performance, e.g. optimise some physical properties [Maranas, 1996, Sheldon et al., 2006], or on process performance, e.g. minimise process cost [Giovanoglou et al., 2003], maximise production [Folić et al., 2008a, Cheng and Wang, 2010]. According to the above definition, CAMD is applied when the desired properties/performance characteristics of the product are given, but its molecular structure is unknown. It is, thus, the reverse problem of property prediction, where given the identity of the molecule/molecular structure, a set of properties are calculated or predicted. CAMD can be applied to systems of various levels of size and complexity, such as solvents, refrigerants, polymers, etc. The design is usually based on the generation of chemically feasible molecular structures of compounds having specific properties. The molecular structure is usually made with the use of a database or a program that contains a database, or through an optimisation procedure. In both cases, the generation is conducted by

property and structural constraints. The properties are estimated by using some kind of fragment-based methodology, where the contributions for a specific property of each fragment present in the compound are added to determine the compound property value. The structural (molecular feasibility) constraints are related to the structure of the compound that is to be designed by defining, for example, the maximum and/or minimum number of fragments forming the final product, the maximum and/or minimum number of aromatic groups, etc.

There are cases where, in order to attain all the desired properties, or to ameliorate the already existing properties for a particular procedure, a mixture or blend needs to be designed. Computer aided mixture/blend design problems are defined as [Achenie et al., 2003] *"Given a set of chemicals and a specified set of property constraints, determine the optimal mixture and/or blend."* Here, although the molecular structures of the candidate chemicals are known, which chemicals are the most appropriate to be used and in which proportions is not known. In this case, the problem is, also, described as an optimisation problem, but constraints that determine the mixture properties need to be added.

CAMD is an alternative rational approach to product design, and to the design of solvents, compared to traditional methods of solvent design, which are based on experiments. Of course, CAMD methods are not by themselves as accurate as experimental methods, but when combined with experiments (i.e., used as a guide for experiments), the overall technique can become more effective, as it is less time-consuming and less expensive. In this section, an introduction to the main CAMD approaches is presented with special emphasis on the computer-aided molecular design of mixtures and solvents for reactions. There are extensive reviews on this matter [Achenie et al., 2003, Gani, 2004] and the reader is referred to these for further details.

2.2 CAMD methods

The methods that have been applied for solving CAMD problems can be categorised in two general groups: (a)*Enumeration techniques* or *Generate and Test*

approaches, where mathematical and qualitative representations are combined, and (b)*Mathematical programming*, where numerical optimisation methods are applied.

2.2.1 General principles of "Generate and Test" methods

The Generate and Test approach, which was first developed for solvent selection and design [Gani and Brignole, 1983, Brignole et al., 1986], consists of two basic parts, the "generation" and the "testing". In the first part, the synthesis of feasible molecular structures takes place, based on specific rules, while in the second part, after the properties of the resulting structures are evaluated and a ranked set of product candidates is constructed.

In CAMD, group-contribution methods, such as UNIFAC [Fredenslund et al., 1975], are usually used for property calculations. The main concept of group-contribution methods is that the physical properties of fluids are modelled as a function of the sum of the contributions of the functional groups of the molecule. The new molecules that are designed are built from those functional groups and have to fulfil a number of property requirements.

"Generation"

In the first step of the method feasible molecules are generated through three main stages: group selection, group characterisation and molecular feasibility rules. The atom groups are classified based on their free attachments, otherwise known as their valency, and, specifically, according to the type and number of their attachments [Gani and Brignole, 1983, Brignole et al., 1986, Pretel et al., 1994]. They are divided into two main categories: the "terminal" groups, that have only one free attachment, and the "intermediate" groups, that have more than one attachment. In order to generate molecules, these groups are combined, according to combination and feasibility rules [Brignole et al., 1986, Pretel et al., 1994, Brignole and Cismondi, 2003]. The combination rules define which attachments are allowed to be combined and the feasibility rules ensure that the molecules resulting from the combination of the groups

are feasible and can exist in reality. Because of the huge number of combinations, Cismondi and Brignole [2004] introduced an algorithm in order to reduce the size of the problem. Property and feasibility constraints have been proposed by various researchers, such as Pretel et al. [1994] and Cismondi and Brignole [2004].

"Testing"

In this part, the structures resulted in the "generation" step are evaluated according to property and operating constraints. These constraints depend on the design problem, since in each process the properties of interest differ, such as the values of them, too. For example, the required properties of a solvent for maximizing the rate constant of a reaction, are different from those of a solvent for separation. The desired properties (e.g. molar volume, melting/boiling temperature, etc.) are usually calculated with group-contribution methods Constantinou and Gani [1994], Marrero and Gani [2001], Conte et al. [2008], Hukkerikar et al. [2012].

2.2.2 Applications of the "Generate and Test" concept

Generate and Test methods have most frequently been applied to solvent selection, starting with Gani and Brignole [1983] and Brignole et al. [1986], who applied the UNIFAC group contribution approach [Fredenslund et al., 1975] in CAMD and they introduced an approach called "molecular design of solvents", MOLDES. The MOLDES procedure, which was further developed later by Pretel et al. [1994], makes use of UNIFAC groups in order to synthesize molecular structures. The procedure followed can be divided into five steps:

- Definition of the problem.
- Selection of intermediate and terminal groups.
- Synthesis of intermediate molecular structures (IMSs).
- Synthesis of molecular structures (SMSs) by the combination of intermediate and final structural groups.

- Ranking of final molecular structures according to process constraints.

Although MOLDES gave satisfactory results, it was limited by the lack of UNIFAC interaction parameters. However, the number of UNIFAC interaction parameters has been significantly augmented and Brignole and Cismondi [2003] successfully applied the method to solvent selection for two separation processes, liquid extraction and extractive distillation.

Harper et al. [2003] introduced a *hybrid* CAMD method, which applies the "Generate and Test" approach. Their method consists of three parts:

- The pre-design part
- The design part
- The post-design part

In the pre-design part, the problem is defined; the aims of the specific CAMD problem are posed clearly and in detail. Then, the properties that need to be evaluated and the evaluation methods for each property are enumerated. Next, the methods for property evaluation as well as the constraints are selected. In order to quantify what is so far a qualitative problem formulation, the "problem formulation algorithm" is proposed, where all available information concerning property and constraint values is added to the formulation. The main objective of the design part is to generate feasible candidates that satisfy all the property constraints. These candidates are synthesised from a set of building blocks (i.e. groups) and then they are tested against the design specifications; the property constraints. This is carried out over four levels, where the input of each level is the output of the previous one. A generate and test approach is applied in each level and the extent of detail varies from level 1, which is the coarsest level, and level 4, which is the most detailed. More specifically, at the first level, group vectors are generated from the combination of groups from a basic set of groups. At the second level, the groups from the first level's vectors are combined to form new feasible molecules, including isomers. A atomistic representation

of the new molecules is derived at the third level, where full (atom-based) connectivity information is obtained and the use of property prediction methods based on a higher order of structural descriptors [Hovarth, 1992, Camarda and Maranas, 1999] is enabled. Finally, in the last level, a three-dimensional representation of the molecules is created, by assigning bond lengths, bond angles and torsion angles. The multi-level structure of the method reduces considerably the number of molecules tested with the most computationally demanding prediction techniques, a fact that makes the method significantly less computationally expensive. In the post-design part, the molecular candidates created in the design part are tested against constraints that were not included in previous parts. Those could be more general constraints, such as cost constraints, or environmental constraints. The optimal candidate is selected, based on its performance in all sections of interest. This hybrid methodology has been partially implemented as a computer program, "ProCAMD" [ICAS documentation].

Vinson [2003] applied the hybrid CAMD method of Harper et al. [2003], described above, for the case of solvent selection for complex solutes. He also proposed a procedure for the combination of experimental work with CAMD, where experimental results are considered as a guide for the setting of the CAMD problem. Cordiner [2003] applied ProCAMD to solvent selection in processes including complex molecules. In addition to ProCAMD, she applied another early evaluation tool, SMSWIN (developed by Syngenta and now part of ProCAMD). The solvent selection in SMSWIN is based on database search, where the solvents are selected according to solubility and operability constraints.

"Generate and Test" methods have been widely used in process design and optimisation and especially ProCAMD [ICAS documentation] (e.g. Hostrup et al. [2001], Gani [2004], Folić et al. [2008b], Patel et al. [2010]). In the next section, the general principles of optimisation CAMD methods are presented.

2.2.3 General principles of optimisation methods

Many researchers approach CAMD problems by formulating a corresponding optimisation programming mathematical problem. The mathematical problem that usually

arises is a mixed-integer non-linear problem, MINLP, which can take the form:

$$\min/\max\ f_{obj}(\mathbf{X},\mathbf{Y}) \tag{2.1}$$

subject to

$$g_1(\mathbf{Y}) \leq 0 \tag{2.2}$$

$$g_2(\mathbf{Y}) \leq 0 \tag{2.3}$$

$$g_3(\mathbf{X},\mathbf{Y}) \leq 0 \tag{2.4}$$

$$g_4(\mathbf{X},\mathbf{Y}) = 0 \tag{2.5}$$

where f_{obj} is the performance objective function which, depending on the process, needs to be minimized or maximized, $g_1(\mathbf{Y})$, $g_2(\mathbf{Y})$, $g_3(\mathbf{Y})$, $g_4(\mathbf{Y})$ correspond to the structural constraints, pure component property constraints, mixture property constraints and process model constraints, respectively, $\mathbf{Y}$ is a vector of binary integer variables, related to the identities of the building blocks and/or molecules, and $\mathbf{X}$ is a vector of continuous variables, related to the mixture and/or process variables.

In most applications of this approach, the constraints on the allowed combinations of groups proposed by Odele and Macchietto [1993] or Churi and Achenie [1996] are applied (e.g. Wang and Achenie [2002], Giovanoglou et al. [2003], Folić et al. [2007], Struebing et al. [2013]). These constraints can be categorised in two groups:

- Structural feasibility constraints, which ensure that two adjacent molecular groups are not linked by more that one bond and that the resulting molecule has zero valency.

- Molecular complexity constraints that impose upper and lower limits on the number of groups of the same type and the total number of groups in a molecule.

The pure component properties are usually calculated with the use a group-contribution method. Some widely used group-contribution approaches are those of Constantinou and Gani [1994], Marrero and Gani [2001], Conte et al. [2008]. Connectivity indices have also been developed for CAMD [Camarda and Maranas, 1999] as structural and

property descriptors. The various optimisation CAMD methods usually proposed differ in the way that constraints are expressed; some alterations that may be made to the "standard" constraints, that are mentioned above, or to the group-contribution method that is used. Different methods have also been proposed for the solution of CAMD problems, such as decomposition approaches or stochastic or deterministic algorithms. A few highlights of some optimisation CAMD methods and applications are given below, followed by specific applications to solvent mixtures and solvents for reactions, in the next sections.

2.2.4 Applications of the optimisation concept

Various applications of optimisation CAMD methods can be found in the literature, primarily for the design of solvents [Pistikopoulos and Stefanis, 1998, Sinha et al., 1999, Karunanithi et al., 2005], refrigerants [Sahinidis and Tawarmalani, 2000, Apostolakou and Adjiman, 2003b] and polymers [Maranas, 1996, Patkar and Venkatasubramanian, 2003a]. The focus is usually either on improving connectivity information of the designed molecule, or developing methods for solving the resulting MINLP, or taking into account specific criteria in the design, such as environmental impact. Several of these approaches are described below with a few details to highlight the differences.

When designing molecules using group contribution methods, which is usually the case in optimisation CAMD, good description of the connectivity of the designed molecules is important. This is what has been addressed by Churi and Achenie [1996] and Apostolakou and Adjiman [2003b] who focused on incorporating full connectivity information in the CAMD formulation. Churi and Achenie [1996] proposed a molecular representation technique for CAMD based on discrete variables that provide structural and connectivity information and enable a very good description of the molecule; i.e. distinguish between isomers, control multiple bonds, specify the design of cyclic/acyclic molecules. With their formulation, they made possible the use of group contribution methods that require connectivity information (e.g. Constantinou and Gani [1994]). The approach was successfully applied to the design of refriger-

ants. Apostolakou and Adjiman [2003b] developed an MINLP CAMD approach, also focusing on the connectivity, based on the high-order group contribution method of Marrero and Gani [2001]. In their approach, Marrero and Gani established a larger set of functional groups, which enables a better representation of molecular structures and used a large data set to estimate the contributions of these functional groups. The main innovations of their work include the assignment of a valency for each bond type to each first-order group (according to Marrero and Gani [2001]), in order to give a better description of molecular structures (three bond types suffice to describe all first-order groups), and the introduction of a systematic methodology to identify forbidden bonds between the molecules, in order to avoid multiple description of a molecule. Their methodology was applied to the design of an aromatic compound, subject to a number of constraints and, also, to the case of refrigerant design, in an attempt to design an environmentally friendly refrigerant [Apostolakou and Adjiman, 2003a].

As mentioned before, CAMD problems are commonly formulated as mixed-integer nonlinear programming problems and, naturally, the more detailed the constraints or demanding the design requirements, the more challenging the solution of the MINLP. Therefore, many researchers have tried to develop methods to facilitate the solution of complex MINLPs for computer-aided molecular design in reasonable computational time. Maranas [1996] adopted, for property estimation, ratio expressions of linear functions of the form

$$p_j(\mathbf{n}) = \frac{\sum_{i=1}^{N} A_{ij} n_i}{\sum_{i=1}^{N} B_{ij} n_i}, \quad j = 1, ..., M \tag{2.6}$$

and

$$p_j(\mathbf{n}) = \left(\frac{\sum_{i=1}^{N} A_{ij} n_i}{\sum_{i=1}^{N} B_{ij} n_i} \right)^{d_j}, \quad j = 1, ..., M \tag{2.7}$$

where p_j is the property, $\mathbf{n} = (n_1, ..., n_N)$ is a vector of integer variables $n_i \in \{$ 0, 1, 2, ...$\}$ describing the number of times the ith group participates in the molecule, A_{ij} and B_{ij} are given parameters associated with molecular group i and property j and d_j is a real positive number. Thus, supposing that the desired properties can be described by these non-linear relations, a methodology to transform them to mixed-integer linear

expressions was proposed, which can be addressed much easier. The methodology was applied to polymer design with various property targets and solutions consistent with experimental data were found, with small CPU requirements. Another work in similar scope is that of Sinha et al. [1999], who proposed a branch and bound algorithm for global optimisation of MINLP CAMD problems where the properties are calculated by expressions of the form:

$$p_k = \frac{f_{NL}^1\left(\sum_j n_j \theta_j^a\right)}{f_{NL}^2\left(\sum_j n_j \theta_j^b\right)}, \tag{2.8}$$

where f_{NL}^1 and f_{NL}^2 are non-linear functions. Although the branch and bound approach [Floudas, 1995] is effective in locating the global solution in MINLP problems, it can be computationally expensive, because all binary variables are used as branching variables, and their number can reach several hundreds. In order to reduce the computational time, instead of using all binary variables for branching, they use branching functions, or "spliting functions", which result in a smaller number of branching nodes. Furthermore, they introduced the "sweep method", a method for constructing linear underestimators and linear overestimators. They applied their branch and bound approach to the design of globally optimal solvents for cleaning in lithographic processes and the results were very satisfactory [Sinha et al., 1999, 2003b]. Decomposition-based approaches are also popular in solving CAMD problems. For example, Karunanithi et al. [2005] introduced a decomposition-based CAMD methodology for the design of solvents. According to their approach the original MINLP problem can be divided to five smaller non-linear sub-problems, each of which requires only the solution of a part of the constraints of the original problem. As each sub-problem is solved, the search space is reduced, and the size of the resulting MINLP problem is significantly smaller, and, consequently, the problem is more easily solved. The decomposition method was applied successfully to two case studies: to the design of solvents for liquid-liquid extraction and to the design of solvents for the formulation of pharmaceutical compounds. A recent decomposition-based approach for the solution of MINLP CAMD problems was proposed by Samudra and Sahinidis [2013]. Here the problem is decomposed over three steps: composition design,

structure design and extended design. The main innovation of this approach is that, in the first step, the composition design problem is formulated as a MILP problem, significantly accelerating (by orders of magnitude) the solution of the problem. It is worth mentioning that the authors also developed novel optimisation and graph models for systematic identification of isomers and fast solution of the subproblems. The applicability of the methodology was demonstrated through three case studies; a refrigerant design application and two solvent design applications for metal decreasing and crystallisation.

A different category in optimisation methods for CAMD are genetic algorithms. Patkar and Venkatasubramanian [2003b] applied a genetic CAMD method to two simple polymer case-studies, where they demonstrated the multiple advantages of genetic programming in CAMD: (a) it is a multiple point search technique, where a set of solutions is examined instead of one solution, (b) it is not derivative-based and, as a result, the maths are simpler, and (c) the methodology can be easily extended to more complex molecules. The latter was further emphasized in [Patkar and Venkatasubramanian, 2003a], where the methodology was applied to a considerably bigger polymer case study. In this case, instead of the standard genetic design, a knowledge-augmented genetic design was added. The knowledge-augmented genetic design renders the algorithm more successful, as only chemically feasible, stable, less complex candidates are produced. It also increases the efficiency of the search by decreasing the number of candidates in the genetic design. The algorithm was not as successful in solving the larger case study, but it succeeded in finding the target molecule for eight out of nine target polymers, even though the search space was significantly increased compared to the smaller case study.

Over the last decades, the need to care about the environment has led researchers to focus on reducing the environmental impact of processes as well as looking for environmentally benign solvents, also known as green solvents. Pistikopoulos et al. [1994] proposed a design methodology for the assessment and the minimisation of the environmental impact of process systems (MEIM) and later [Pistikopoulos and Stefanis, 1998] combined it with a CAMD model to design solvents with minimum

environmental impact. GlaxoSmithKline has developed the Solvent Selection Guide (SSG)[Curzons et al., 1999], where widely used solvents are ranked according to environmental, health and safety criteria. The Guide has been also expanded [Jiménez-González et al., 2005b], by taking into account the life cycle of the solvents. Weis and Visco [2010], based on the Solvent Selection Guide of Curzons et al. [1999], developed a CAMD approach to identify additional green solvents to the solvent list of SSG. Solvent candidates were built using the Signature molecular descriptor [Visco et al., 2002] and then, quantitative structure-property relationships [Camarda and Sunderesan, 2005] were created in order to rank the solvent candidates, according to the environmental criteria of SSG. An investigation of alternative green solvents has also been undertaken by Clark and Tavener [2007] taking again under consideration the solvent life cycle. Five alternative solvent categories were considered; supercritical CO_2, ionic liquids, fluorous solvents, water and organic solvents derived from renewable resources (e.g. bioethanol, biodiesel, glycerol). The aforementioned solvents were classified according to their physical properties, process and cost criteria, as well as health and safety criteria. Supercritical CO_2 and water are judged to be the most suitable alternative solvents for the majority of cases. The renewables also make a very good choice, when organic solvents cannot be avoided, thanks to their low toxicity and their very good performance in terms of the process and cost criteria.

2.3 CAMD for Mixtures

The tunable properties of mixed solvents, depending on the composition of the mixture, or the fact that they can be environmentally benign (e.g. CO_2-expanded solvents), has attracted the attention of industry and academics. Computer aided mixture/blend design is defined as, "Given a set of chemicals and a specified set of property constraints, determine the optimal mixture." [Karunanithi et al., 2005]. CAMD studies of mixtures are usually addressed by adding to a typical MINLP CAMD problem additional constraints, relevant to the mixture properties, such as the miscibility of the co-solvents. In most cases, one of the co-solvents is pre-defined and only the sec-

ond co-solvent is designed, or the co-solvents are chosen from a specific set of solvents. However, there are a few works where the simultaneous design of all components in the mixture is performed [Buxton et al., 1999, Papadopoulos et al., 2013]. Although CAMD of mixtures is not an area as much studied as CAMD of pure components, there are a few significant contributions that are worth mentioning here. The main focus of these works has been not only to include the mixture related constraints, but also to develop methodologies, usually decomposition-based algorithms, in order to address the increased complexity of the MINLP problem formulation.

Klein et al. [1992] presented a work on computer-aided mixture design, a concept which was firstly introduced by Nielsen et al. [1990], where an optimisation problem is formulated having as the objective function the cost of the mixture and as constraints the solubility parameters (linear constraints) and the boiling-point temperature (non-linear constraint) of the mixture. For the solution of the resulting NLP problem, they proposed a successive regression and linear programming (SRLP) algorithm.

Buxton et al. [1999] introduced a systematic decomposition-based approach for the optimal CAMD of solvent blends for nonreactive, multicomponent absorption processes, taking into account environmental impact. Here, both the components of the mixture and their composition are optimised. This work is, in fact, an extension of a previous work of Pistikopoulos and Stefanis [1998], which referred to the design of pure solvents. The framework consists of three basic parts: (a) identification of agent-based process operations, (b) determination of solvent candidates obeying to specific property and environmental constraints, and (c) verification of the candidates' performance on a plant-wide basis. In their method, they apply the principles of the Methodology for Environmental Impact Minimization (MEIM) developed by Pistikopoulos et al. [1994], a methodology that estimates and minimises the various damaging effects of processing systems to the environment. In addition, they propose a solution procedure for the resulting complex NLP problem, adopting a step-wise decomposition based algorithm. The initial problem is decomposed into seven smaller problems, and, in this way, the resulting NLP problem becomes smaller in size and easier to solve. They applied their approach to both single and multicomponent

absorption tasks and managed to find suitable solvent blends.

Sinha et al. [2003a] developed an interval-analysis based optimisation framework to describe the blend design problem, a MINLP problem, where the co-solvents are chosen from a set of pure-component solvents. In their optimisation framework, they introduced new acceleration strategies and extended the standard interval-based optimisation algorithm [Vaidyanathan and El-Halwagi, 1994] to enable the solution of MINLP problems. They developed an eight-step interval-based domain reduction algorithm, LIBRA, and they applied it to the design of environmentally acceptable blanket wash mixtures, managing to find the globally optimal blend.

Karunanithi et al. [2005] proposed a decomposition-based computer aided molecular/mixture design methodology. In their approach, the original MINLP problem is divided into five smaller non-linear sub-problems: (a) structural constraints, (b) pure component constraints, (c) mixture component constraints, (d) miscibility constraints, and (e) process model constraints and the objective function. Each sub-problem takes into account only a part of the constraints of the starting problem. The advantage of this method is the reduced size of the resulting MINLP problem (sub-problem (e)), which makes the solution of the problem much simpler. The presented applications considered a solvent mixture, where water was the one co-solvent, while the other co-solvent was designed.

A recent work on CAMD of mixtures is that of Papadopoulos et al. [2013] where a CAMD formulation is developed for optimal binary mixtures for organic rankine cycles. In this work, the challenge of both designing the two co-solvents and also determining their optimal composition in the mixture is addressed. The CAMD problem is formulated as a multiobjective optimisation problem (MOO), where the objectives are several operating and safety features, decomposed into two stages. In the first stage the optimal molecular structure of one of the components in the mixture is designed. This is done by requiring only by one component to satisfy the optimisation constraints and allowing the other component to violate them. After the first stage, a number of feasible and infeasible mixtures is designed, which are further screened in the second stage, where now the second component has to satisfy

all constraints and the optimal composition of the mixture is also determined.

2.4 CAMD of Solvents for Chemical Reactions

Solvents play various roles in chemical reactions. They control heat transfer when the reaction is endothermic or exothermic, by having a high heat capacity and by absorbing heat, they affect the rate of the reaction, change the pressure and/or temperature of a gas-phase reaction when taking place in the liquid phase. There has been a growing body of work in the area of solvent design for reactions, using chemometrics [Carlson, 1992, Buncel et al., 2003] as well as computer-aided molecular design techniques. Here the focus is on CAMD and several breakthroughs in the area are presented.

The work of Folić et al. [2004] can be considered as the first attempt to apply a CAMD methodology to solvents for reactions. Folić et al. [2004, 2007] developed a hybrid experimental/computer-aided methodology for the design of solvents for reactions and their approach can be summarised in the following steps.

- Starting with a small number of solvents and the relevant experimental data for these solvents, a reaction model is developed.
- A computer-aided solvent design problem (MILP or MINLP) is formulated, which is based on that reaction model.
- A number of candidate solvents that maximise the rate constant is generated.
- Meanwhile, in order to investigate the uncertainty of the problem, they use global sensitivity analysis to evaluate its effect.
- The last step is the verification of the results, where the predicted rate constants are compared to experimental data.

If the results are not compatible with the data, the measured rate constants for the candidate solvents are added to the initial solvents and a new reaction model is built. The procedure is repeated until the prediction is satisfactory. The algorithm

concludes when the optimal solvent under the specific uncertainty conditions is found. For the connection of the rate constant with the solvent properties, they use the solvatochromic equation and for the calculation of solvent properties, they apply group-contribution methods. In a following work [Folić et al., 2008a], the methodology is extended to more complex reaction systems. An important advantage of their framework is that while kinetic data of a small number of solvents is required, a very large solvent design space can be investigated. In this work, in addition to the structure, property and chemical feasibility constraints, they include in their problem formulation process model constraints in order to take into consideration more complex reaction schemes, such as competing or consecutive reactions. They also focus on the subject of sensitivity of parameters and a systematic framework to handle the uncertainty is proposed and applied to a Menschutkin reaction.

Gani et al. [2005] focused on organic reactions and combined industrial practice and computational tools for property estimation, in order to develop their methodology. They use databases for solvents and for reactions and apply the hybrid computer-aided molecular design technique of Harper et al. [2003] to generate solvent candidates. The solvents from the databases and the generated solvents from CAMD are ranked according to scores, defined dependent on their properties. In a second stage, these candidate solvents are further evaluated by more detailed calculations and a final set of candidate solvents, that not only satisfy chemical properties, but also environmental, safety and health requirements, is given.

The solvent selection methodology of Gani et al. [2005] was extended by Gani et al. [2008] to handle multi-step reactive systems and also solvent substitution for specific reactive steps in existing processes. The methodology was applied to solvent selection for an enzymatic glycerolysis reaction and also for a multi stage reactive system (also in [Folić et al., 2008b]). Thus, the methodology was evaluated in actual industrial practice, and it proved to be successful both in predicting appropriate solvent substitutes for the first case study and in predicting potential candidates for the different reaction steps for the second case study. Promising solvent candidates that could possibly be used in all reaction steps were also proposed.

Another work on CAMD for solvents in reactions is that of Stanescu and Achenie [2006], where they have made a theoretical study on the kinetics of the Kolbe-Schmitt reaction. In their approach, they firstly use DFT (Density Functional Theory) calculations to study the reaction mechanism. Then, a CAMD methodology is applied to generate additional solvents to those proposed from literature. They use ProCAMD within the Integrated Computer Aided System (ICAS)[ICAS documentation]. The most promising solvents are further tested using DFT solvation calculations and, finally, the rate constants of the best final candidates are calculated applying DFT.

A recent CAMD study on reaction kinetics by Struebing et al. [2013] applies an *ab initio* methodology to the design of solvents for reactions. The innovation of this approach is that no experimental data are used for the solvent design. The methodology proposed is as follows:

Step 1: An initial set of solvents is chosen (6-7 solvents). Step 2: The reaction rate constant in specific solvent(s) is calculated using Transition State Theory and Quantum Mechanical (QM) calculations with a continuum solvation model.

Step 3: A reaction model is built by regressing the solvatochromic equation, an empirical model that connects the rate constant with solvent properties, that can predict the reaction rate constant for a larger number of solvents.

Step 4: The optimal solvent is found by using an optimisation CAMD formulation, based on the above reaction model.

Step 5: If a new solvent has been designed, go to Step 2, otherwise terminate.

The algorithm iterates until no new solvent is found. This approach differs from that of Folić et al. [2007] in the second step; instead of using experimental data to build the reaction model, the rate constants are predicted from *ab initio* calculations. The methodology was applied to the design of optimal solvent that maximises the rate

constant for a Menschutkin reaction and the results were also verified by experiments.

2.5 Computer-Aided Molecular and Process Design

Although there is a lot of work in the area of computer-aided molecular design (CAMD) of solvents, concerning the solvent performance in the context of solvent property targets or solvent performance measures (e.g. selectivity, capacity), there are only a few approaches focused on process performance. In this section, some of these approaches, known as computer-aided molecular and process design (CAMPD), are discussed.

The work of Buxton et al. [1999] that was described in section 2.3 applies a computer-aided molecular and process design approach, since process requirements are included in the optimisation problem. As mentioned before, a methodology for the optimal CAMD of solvent blends for non-reactive, multicomponent absorption processes was proposed. Marcoulaki and Kokossis [2000] also presented an integrated optimisation approach to solvent design. Based on already existing methods for the calculation of the properties, they proposed a framework for the synthesis of novel solvents, considering not only property objectives, but also process operation objectives. They demonstrate applications to the design of solvents for liquid-liquid extraction, extractive distillation and gas absorption.

Giovanoglou et al. [2003] introduced a mixed-integer dynamic optimisation (MIDO) methodology for the design of solvent in batch processes. In order to simplify the problem, a decomposition-based algorithm is used. The overall problem is decomposed into

- A primal, dynamic optimisation problem, which, in its turn, is decomposed into three parts: physical property tests, process model initialisation and simulation and, finally, process model optimisation, and
- A master mixed-integer linear problem.

The algorithm was applied to a separation system from a batch process; the dehydration-decantation separation unit for the removal of water from the feed mixture. The model is divided into five parts: (i) the dynamic process model, (ii) structure-property relationships, (iii) process specifications, (iv) physical property tests, and (v) a process performance measure to be optimized, as the objective function. Setting an economic criterion as the objective function, the most feasible and profitable solvent for the particular batch process was found.

Papadopoulos and Linke [2006] combined multiobjective optimisation with CAMD for the design of solvents for chemical processes. Based on the representation and optimisation framework of the solvent molecules introduced by Marcoulaki and Kokossis [2000], they proposed a decomposition based approach. The first stage includes multi-objective optimisation in order to define the optimal solvent candidates for a set of particular molecular design objectives. In the next step, these solvents are further examined in a process synthesis problem, and then are optimised to finally find the optimal solvent-process system. Their methodology was applied to the design of solvents for liquid-liquid extraction and gas-absorption processes, similarly to the work of Marcoulaki and Kokossis [2000].

Lek-Utaiwan et al. [2008] worked on the integrated design of solvent-based extractive separation processes. They proposed a systematic methodology that combines solvent design with extractive separation requirements. The optimal process needs to satisfy all product-process constraints as well as economic and environmental constraints. The methodology was extended [Lek-Utaiwan et al., 2009] including experimental validation, in order to validate the predicted performance of the optimal solvents, the process design and the cost analysis, and make a more realistic design.

Cheng and Wang [2010] developed a computer-aided process/solvent design of a biocompatible solvent for an integrated extractive fermentation and distillation process. In order to solve the resulting MINLP problem, a two-part scheme was proposed; the mixed-integer hybrid differential evolution (MIHDE) [Liao et al., 2001] is firstly used to find a feasible solution, and, as a second step, that feasible solution is used as the initial starting point for a trust region sequential quadratic programming

algorithm, MISQP [Exler and Schittkowski, 2007].

Bardow et al. [2010] have also proposed a two-step procedure to address MINLP problems in the integrated solvent and process design, the *continuous-molecular-targeting* approach to computer-aided molecular design (CoMT-CAMD). In the first step, the discrete model parameters, that describe the solvent, are relaxed and both process variables and solvents parameters are optimised continuously. The solution of the relaxed problem is a hypothetical target molecule, described by a set of solvent parameters. In the second step, the hypothetical target molecule is mapped onto real molecules performing a database search, in order to find the real molecule with the closest parameters. PCP-SAFT equation of state is used for the prediction of the solvent parameters. The approach was successfully applied to a absorption/desorption process for CO_2 capture and solvents that improved the performance of the process were identified.

Pereira et al. [2011] have developed a computer-aided molecular and process design (CAMPD) methodology, based on SAFT-VR equation of state [Gil-Villegas et al., 1997], [Galindo et al., 1998] and they applied it to the separation of CO_2 and CH_4 through physical absorption of the CO_2 into an n-alkane solvent. They approached the problem by dividing it into two parts: (a) the identification of the design space, where the different types of separation techniques are examined, as well as the the types of solvents that are to be used, and (b) the formulation and solution of an optimisation problem that describes the problem posed. The model, that is introduced, consists of three parts:

- A thermodynamic model (SAFT-VR), in order to predict the properties of the mixtures of interest
- A steady-state process model for the flowsheet that they propose and
- A cost evaluation model.

A typical MINLP problem results, which, when considering only mixtures of n-alkanes as the solvent, becomes an NLP problem, since the solvent mixture can be represented

by continuous variables only. In this work, the significance of applying an advanced equation of state in CAMPD problems has been shown.

2.6 Conclusions

Computer-aided molecular design theory and applications have been briefly reviewed. The two wide categories of CAMD have been presented and several methodologies and applications of both have been discussed. The review has been mainly focused on solvent studies, as this is the scope of this work. Although solvents have been widely studied computationally applied CAMD methods, only a few works have been reported concerning solvent design for reactions. This is the area where this work is focused. Optimisation CAMD methods for optimal solvents for reactions have been developed and are presented in the rest of the thesis.

Chapter 3

Integrated design of a gas-expanded liquid and reactive system

3.1 Introduction

The advantages of mixed solvents in chemical processes are widely acknowledged and very interesting computational studies have been published over the last years, as discussed in section 2.3. Mixed solvents could also be advantageous in reactive processes, as their impact on the reaction rate can be significant [Ford et al., 2008b]. Furthermore, the urge to reduce the environmental impact of solvents has resulted to the continuously increasing interest in alternative, environmentally friendly solvents. In this chapter, a methodology is developed for the design of solvent mixtures for reactions and, specifically gas-expanded liquids, that are mixed solvents usually composed of an organic solvent and the environmentally benign CO_2. Process criteria are also taken into account in the design.

An introduction to gas-expanded liquids is given in section 3.2. The general design problem is discussed in section 3.3 and the proposed methodology and the models that are used are described in detail in section 3.4. Finally, the application of the

methodology in a Diels-Alder reaction is discussed in section 3.5.

3.2 Gas-expanded liquids

An interesting category of "green" solvents is that of gas-expanded liquids (GXLs) [Jessop and Subramaniam, 2007], which are mixed solvents composed of an organic solvent and a compressible gas, usually CO_2 due to the low risk associated with its use and its economic advantages. GXLs have recently generated great interest, because of their distinct behaviour, which is due to the combination of gas and liquid characteristics; for example, CO_2 enhances gas solubility and mass transfer, while organic solvents augment the solubility of liquid and solid solutes. The properties of a gas-expanded solvent can be tuned by exploiting the properties of the organic solvent and those of the gas, simply by varying the pressure of the system. Ye et al. [2012] studied and successfully predicted the VLE phase behaviour of twelve multicomponent systems including CO_2-expanded liquids using cubic equations of state with excess Gibbs free energy based mixing rules. GXLs have been shown to be effective solvents for many processes such as oil recovery, where Hwang and Ortiz [2000] showed that by adding an amount of organic solvent in supercritical CO_2 the solvation power of the resulting mixture was significantly enhanced, leading to increased extraction efficiency and reduced asphaltene deposits. They have also been investigated in the context of gas recrystallization [Chang and Randolph, 1991], as mobile phases for HPLC (high-performance liquid chromatography) [Wen and Olesik, 2001], as solvents for post-reaction separations, e.g. in homogeneous catalysis where efficient recovery and recycle of the catalyst is required and where GXLs offer mild operating temperatures and pressures [West et al., 2004], or for particle formation [Jung and Perrut, 2001, Fages et al., 2004]. GXLs have also been studied in the context of reactions [Wei et al., 2002], where their performance is also remarkable, as discussed in two extensive reviews on this subject [Akien and Poliakoff, 2009, Subramaniam, 2010]. Their advantages include recovery and recycle of both the organic compound and CO_2 through depressurization, which is less energy intensive

than standard separation techniques, moderate operating pressures relative to the use of supercritical CO_2, and enhanced transport rates and reaction rates, compared to pure organic solvents. Overall, they are environmentally friendly, as the amount of the organic species in a given volume of GXL solvent is reduced, thanks to the addition of CO_2. Thus, gas-expanded solvents satisfy several green chemistry and process engineering requirements.

When designing a GXL for a given process that includes reaction and separation tasks, one must decide both on the nature of the organic solvent and the composition of the GXL. While there is a growing body of work demonstrating the benefits of GXLs for specific processing steps, and assessing the economic and environmental performance of supercritical fluid-based processes [Fang et al., 2007, Gong et al., 2008, Ghanta et al., 2012a,b], the question of how to identify the best GXL has not been addressed, although there have been few systematic comparisons of the performance of a GXL compared to that of a pure organic solvent [Akien and Poliakoff, 2009]. In fact, the optimal choice is closely linked to the process of interest, as trade-offs must be made between the productivity of the reactor and the cost and effectiveness of any subsequent separations, in order to achieve the best overall process performance. This has been amply demonstrated in the literature on solvent design for separations, in which organic solvents have been the main focus [Buxton et al., 1999, Marcoulaki and Kokossis, 2000, Giovanoglou et al., 2003, Eden et al., 2004, Karantzi et al., 2007, Lek-Utaiwan et al., 2008, Bardow et al., 2010, Pereira et al., 2011]. In order for a systematic approach to GXL design to be feasible, one must be able to relate mathematically process performance to solvent properties and thus to quantify the impact of solvent choice on physico-chemical phenomena such as reaction rates and phase equilibrium. In this context, there have been several advances in the design of mixtures of organic solvents or aqueous organic solutions [Klein et al., 1992, Buxton et al., 1999, Sinha et al., 2003a, Gani, 2004, Karunanithi et al., 2005, Akula et al., 2012], in which both the nature of the co-solvent(s) and the composition of the mixture are considered as part of the design problem.

Despite the advances in the area of solvent design for reactions, discussed in

section 2.4, existing approaches cannot be applied directly to the design of GXLs for use in reactive processes. A specific challenge arises in the presence of solid reactants, where it is necessary to trade-off the increase in reaction rate as the CO_2 content of the GXL increases, against the resulting decrease in the solubility of the reactants. The objective of this chapter is thus to develop a methodology for the design of GXLs based on the consideration of GXL performance within a process. The best GXL is chosen based on the overall economic performance of a conceptual process, where the key processing steps in which the GXL is to be involved are taken into account. The effect of the addition of CO_2 on the organic solvent inventory can also be investigated, and compared to the case where pure organic solvent is used. The nature of the organic co-solvent and the composition of the GXL are key decisions, whose optimal values are affected by reactor volume and energy requirements. Although the design of mixed solvents has been addressed in the literature [Klein et al., 1992, Buxton et al., 1999, Sinha et al., 2003a, Karunanithi et al., 2005], attention has so far been focused mainly on low-pressure separation processes and high pressure and reactive systems remain challenging. In particular, the effect of GXL design on the reaction rate, through changes in the reaction rate constant and in the solubility of the reactants as a function of composition or pressure, cannot be quantified using the models that form the basis of existing CAMD approaches. These issues are tackled here by embedding within the design problem empirical models that link the properties of the mixed solvent to the reaction rate constant as well as a predictive equation of state that can capture pressure effects. The proposed methodology is illustrated on the design of a process to produce a Diels-Alder adduct; it is equally applicable to the design of any other mixed solvent.

3.3 General design problem formulation

The proposed methodology for the integrated design of a CO_2-expanded solvent for a given reaction and the conceptual design of the associated reactor-separation system

is based on an optimisation framework. The design problem can be expressed as:

$$\min_{\mathbf{x,y}} f(\mathbf{x,y})$$

subject to

$$\begin{aligned} \mathbf{g}(\mathbf{x,y}) &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}) &\leq \mathbf{0} \\ \mathbf{x} &\in X \subseteq \mathbb{R}^n \\ \mathbf{y} &\in \{0,1\}^q \end{aligned} \tag{3.1}$$

where $\mathbf{x}$ is a n-dimensional vector of process variables such as flowrates, volumes, compositions and $\mathbf{y}$ is a q-dimensional vector of binary variables used to specify the choice of organic co-solvent. The objective function, $f(\mathbf{x,y})$, is an overall performance index, such as the total annualised cost of the process, which needs to be minimized for a specific production rate. The equality constraints, $\mathbf{h}(\mathbf{x,y})$, correspond to the property, process and cost models and the inequalities, $\mathbf{g}(\mathbf{x})$, represent the design constraints. The property constraints include the relationship between the GXL design (nature of the co-solvent and composition) and the reaction rate constant and phase equilibria, while the process model constraints include conservation equations.

A key challenge in GXL design is to develop a modelling framework which allows all the quantities necessary to obtain the performance index to be computed for a range of design choices. A conceptual flowsheet is used to link the GXL make-up to process performance, taking into account the most significant contributions to performance in reaction and separation steps. To facilitate exposition of the model, it is introduced based on the consideration of a single bimolecular reaction:

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{C}.$$

The flowsheet is shown in figure 3.1. It consists of a CSTR, a simple separation system and a compressor, illustrating the need to capture solvent and pressure effects on kinetics and phase equilibria. In the separation section, it is assumed that CO_2 is

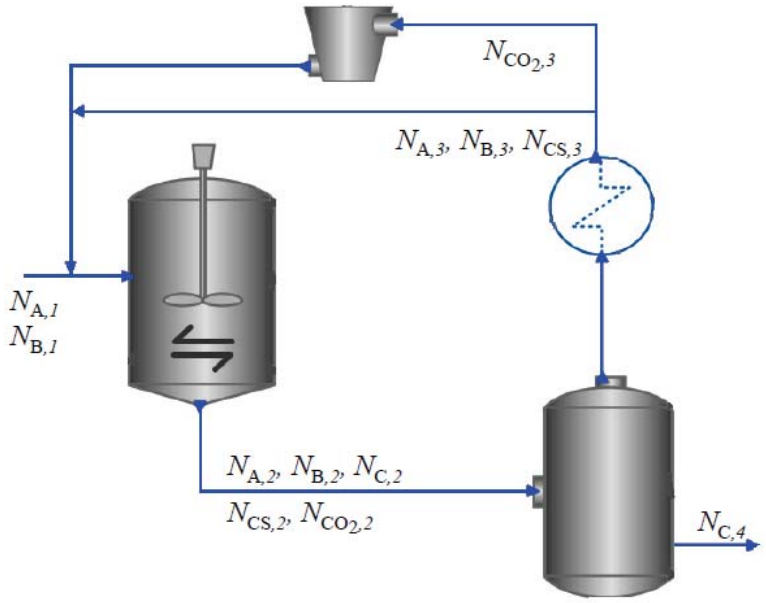


Figure 3.1: The conceptual flowsheet used for GXL design, consisting of a CSTR, a separation unit and a compressor.

recovered by simple depressurization, and that the separation of the organic solvent and reactants and products is effected by evaporation and subsequent condensation of the solvent. In practice, more cost-effective ways to recover the components may be suitable, such as the further addition of CO_2 in order to take advantage of its anti-solvent properties. However, given the high energy requirements associated with the proposed evaporation/condensation, it provides a significant cost penalty on the use of organic solvent and hence a best-case assessment of the performance of a GXL relative to the pure organic solvent. The identification of the optimal GXL that results in the minimum total cost of the process as a function of operating pressure, reactor volume and energy costs, is investigated.

3.4 Model development

The model used for GXL design is divided into several sub-models that capture different physical aspects:

- a kinetic model, that captures the dependence of the reaction rate constant on the GXL;
- a thermodynamic relationship between the solvent make-up, pressure, temperature and density;
- the material and energy balances;
- the dependence of the performance indices (cost and organic solvent inventory) on process variables.

These different components of the model are described in more detail in the remainder of this section.

3.4.1 Reaction Rate Constant

In order to study solvent effects on the reaction rate, an empirical model, the solvatochromic equation, has previously been applied [Folić et al., 2007, 2008a], [Struebing et al., 2013]. It correlates the reaction rate constant with several solvent properties, typically for a pure solvent. However, in our case, a model is needed that relates the rate constant to solvent mixtures. For this reason, a preferential solvation model is used, that connects the solvent properties with the composition of the mixed solvent. These properties are then used within the solvatochromic equation. Solvatochromic equation and preferential solvation models are discussed in detail in this section.

Solvatochromic equation

Beginning in the '70s, Kamlet and Taft [1976] and Taft and Kamlet [1976] carried out a series of studies of solvent effects on spectra. They derived scales of solvent hydrogen-bond acidities, hydrogen-bond basicities and solvent dipolarity/polarisability by using UV/Vis spectral data of solvatochromic compounds and measuring the frequency shifts. They used the results to derive an empirical *linear free-energy relationship* (LFER), known as the *solvatochromic equation*. They included in their equation, not only the bulk polarity of the solvent, but also the ability of the solvent to act

as a donor or acceptor in hydrogen bonding. In their approach, the bulk polarity is represented by parameter π^*, while parameters α and β correspond to the acidity and basicity, respectively, of the solvent. This is related to the property of interest, XYZ, by the following relation for their solvatochromic equation:

$$XYZ = XYZ_0 + s\pi^* + a\alpha + b\beta \tag{3.2}$$

where XYZ is the property being studied, XYZ_0 is the property XYZ in a reference solvent, π^*, α and β are the aforementioned solvent properties and s, a and b are sensitivity coefficients for the property XYZ to the corresponding solvent property. When the property of interest is the reaction rate constant k, equation (3.2) becomes

$$k = k_0 + s\pi^* + a\alpha + b\beta. \tag{3.3}$$

In the literature on LFER, the solvent properties π^*, α and β are referred to as *solvatochromic parameters*. The solvatochromic parameters are estimated from the long length wavenumbers of maximum absorption of solvatochromic probes (otherwise known as indicators), using linear relationships. Here, we report some examples of how the solvatochromic parameters are estimated. Rosés and co-workers carried out a series of studies [Ràfols et al., 1995, Bosch et al., 1996b, Ortega et al., 1996, Bosch et al., 1996a, Ràfols et al., 1997, Rosés et al., 1997, Buhvestov et al., 1998, Herodes et al., 1999] on the Kamlet and Taft solvatochromic parameters, in which they used different relations to estimate them, depending on the solvatochromic probes that they used. In parts 4 and 5 of their work [Bosch et al., 1996a, Ràfols et al., 1997], they calculate the parameters according to the following equations:

$$\pi^* = \frac{34.12 - v_B}{2.343}, \tag{3.4}$$

$$\beta = \frac{1.035v_D + 2.64 - v_C}{2.80}, \tag{3.5}$$

$$\alpha = 0.198v_A - 2.091 - 0.899(\pi^* - 0.211\delta) - 0.148\beta, \tag{3.6}$$

where v_A, v_B, v_C, c_D are the long length wavenumbers of absorption band of indicators A (2,6-dimethyl-4-(2,4,6-triphenyl-pyridin-1-io)-1-phenolate, or Reichardt's betaine), B (4-nitroanisole), C (4-nitroaniline), D (N,N-diethyl-4-nitroaniline), respectively and δ is a polarizability correction term. In Part 6 [Rosés et al., 1997], they made use of the equations:

$$\pi^*_A = 0.433(37.67 - v_A), \tag{3.7}$$

$$\pi^*_B = 0.427(34.12 - v_B), \tag{3.8}$$

$$\pi^*_C = 0.412(32.56 - v_C), \tag{3.9}$$

$$\beta_D = 0.346(35.045 - v_D) - 0.57\pi^*, \tag{3.10}$$

$$\beta_E = 0.358(31.10 - v_E) - 1.125\pi^*, \tag{3.11}$$

$$\alpha_F = -0.186(10.91 - v_F) - 0.72\pi^*, \tag{3.12}$$

$$\alpha_G = -0.208(11.63 - v_G) - 0.72\pi^*, \tag{3.13}$$

where the subscripts A-G denote the indicator, 1-ethyl-4-nitrobenzene, 4-nitroanisole, 2-nitroanisole, 4-nitrophenol, 4-nitroaniline, Reichardt's betaine dye and Reichardt's water-soluble betaine dye, respectively. In this case, they measured the solvatochromic parameters from different indicators and then took the average value. Finally, in part 7 [Buhvestov et al., 1998], the solvatochromic parameters were calculated by the following linear free-energy relationship:

$$v = v_0 + s\pi^* + b\beta + a\alpha, \tag{3.14}$$

where v_0, s, b, a are coefficients that depend on the indicator. The solvatochromic parameters were obtained from different indicators; 1-ethyl-4-nitrobenzene, 4-nitroanisole, 2-nitroanisole, 4-nitrophenol, 4-nitroaniline, Reichardt's betaine dye and Reichardt's water-soluble betaine dye for π^*, 4-nitrophenol and 4-nitroaniline for β, and Reichardt's water-soluble betaine dye for α. The mean value was taken for the solvatochromic parameters obtained for the same mixture from different indicators.

In these two latter works [Rosés et al., 1997, Buhvestov et al., 1998] the authors claim that obtaining the solvatochromic parameters using different indicators and consider the average value as the solvatochromic parameter for the mixture is more reliable than using the measurement of a single indicator. This is indeed reasonable, as, for example, the π^* values obtained from the three different indicators reported in Rosés et al. [1997] vary considerably. This discrepancy is observed because each indicator is affected differently from the solute-solvent interactions, which, in the case of mixtures, are also affected by the solvent-solvent interactions [Marcus, 2002].

Many solvatochromic parameters have been obtained experimentally for pure solvents [Abraham et al., 1991, Abraham, 1993a,b, Zissimos et al., 2002] and predictive methods have been proposed for solvents for which no data are available [Platts et al., 2006, Sheldon et al., 2005]. There has also been some interest in mixed solvents [Barbosa et al., 1996, Reta et al., 2001, Ray and Bagchi, 2005] but the prediction of mixed solvent parameters as a function of composition, which are needed for GXL design, remains challenging. In parallel to these efforts, there have been some successful attempts to correlate the solvatochromic parameters of binary solvent mixtures to their composition, based on preferential solvation models, and these will be discussed in the next section.

Mixed Solvents: Preferential Solvation model

The study of solute-solvent and solvent-solvent interactions in solvent mixtures is of great scientific interest [Marcus, 2002]. The physico-chemical properties of the mixture are affected not only by the solute-solvent interactions, but also by the interaction between unlike solvent molecules and this is what makes the behaviour of

mixtures highly non-ideal. It has been found [Marcus, 2002, Reichardt and Welton, 2011] that the ratio of the solvent components in the solvent shell may differ from that in the bulk; the solute is surrounded preferentially by the co-solvent that leads to a more negative Gibbs energy of solvation. The phenomenon where the solvent shell, the solvent area around the solute, has a different composition from the bulk composition is called *preferential solvation* [Reichardt and Welton, 2011]. As for pure solvents, a method that is often applied to study interactions in mixtures is by means of solvatochromic indicators, which offer information about solvent properties, such as polarity (E_T^{sat}) [Reichardt and Welton, 2011], polarisability or hydrogen bonding capabilities (solvatochromic parameters). The focus here is on the analysis of the preferential solvation models that have been proposed for the calculation of the solvatochromic parameters, the Kamlet-Taft parameters of dipolarity/polarizability (π^*), hydrogen-bond acceptor basicity (β) and hydrogen-bond donor acidity (α), of solvent mixtures. Two key approaches for the prediction of the solvatochromic parameters are those of Rosés and co-workers [Ràfols et al., 1995, Bosch et al., 1996b, Ortega et al., 1996, Bosch et al., 1996a, Ràfols et al., 1997, Rosés et al., 1997, Buhvestov et al., 1998, Herodes et al., 1999] and Harifi-Mood et al. [2006, 2007].

Ràfols et al. [1995] investigated solute-solvent and solvent-solvent interactions in binary solvent mixtures, using solvatochromic indicators. This initial work has so far been followed by seven further publications from the same group on this subject. They started their work with the study of the polarity of the solvents, by means of E_T polarity. They investigated the behaviour of E_T polarity for various solvent mixtures and various indicators [Ràfols et al., 1995, Bosch et al., 1996b, Ortega et al., 1996]. Next, they started using solvatochromic parameters to quantify solute-solvent and solvent-solvent interactions [Bosch et al., 1996a, Ràfols et al., 1997, Rosés et al., 1997, Buhvestov et al., 1998]. Investigating different preferential solvation models, they proposed a general model, that is based on two solvent exchange processes, as shown below:

$$I(S_i)_m + mS_j \quad \leftrightarrows \quad I(S_j)_m + mS_i \tag{3.15}$$

$$I(S_i)_m + \frac{m}{2}S_j \quad \leftrightarrows \quad I(S_{ij})_m + \frac{m}{2}S_i \tag{3.16}$$

where S_i and S_j indicate the two pure solvents and S_{ij} represents a hypothetical solvent formed, by the interaction of the two solvents S_i and S_j, in the microsphere of the solvation. This new solvent contains the same number of molecules of solvents i and j and it can have different properties from those of pure solvents i and j. They have demonstrated that the value of m that gives the best results for the majority of binary systems is close to 2, in which case the general model becomes the two-step model proposed by Skwierczynski and Connors [1994],

$$I(S_i)_2 + 2S_j \quad \leftrightarrows \quad I(S_j)_2 + 2S_i \tag{3.17}$$

$$I(S_i)_2 + S_j \quad \leftrightarrows \quad I(S_{ij}) + S_i \tag{3.18}$$

They defined the preferential solvation parameters $f_{j/i}$ and $f_{ij/i}$, which are composition-independent, were introduced to measure the tendency of the indicator (or solute) to be solvated by solvents j and ij, respectively, rather than solvent i. They are defined as :

$$f_{j/i} = \frac{x_j^s/x_i^s}{(x_j/x_i)^2}, \tag{3.19}$$

$$f_{ij/i} = \frac{x_{ij}^s/x_i^s}{x_j/x_i}. \tag{3.20}$$

where x_i^s , x_j^s , x_{ij}^s are the mole fractions of solvents i, j, ij, respectively, in the microsphere of the indicator and x_i, x_j are the mole fractions of solvents i, j, respectively, in the bulk mixed solvent.

In the preferential model discussed here it is considered that the Y values of the mixture, where Y is an appropriate solvatochromic property, are an average of the Y values in the solvents that compose the solvation microsphere of indicator, according to their mole fractions in this sphere. This can be expressed as below:

$$Y = x_i^s Y_i + x_j^s Y_j + x_{ij}^s Y_{ij} \tag{3.21}$$

Considering that

$$x_i + x_j = x_i^s + x_j^s + x_{ij}^s = 1 \tag{3.22}$$

and substituting x_i^s , x_i^s , x_{ij}^s from equations (3.19) and (3.20), the mole fractions in the microsphere can be calculated from the equations below:

$$x_i^s = \frac{(1-x_j)^2}{(1-x_j)^2 + f_{j/i}(x_j)^2 + f_{ij/i}(1-x_j)x_j} \tag{3.23}$$

$$x_j^s = \frac{f_{j/i}(x_j)^2}{(1-x_j)^2 + f_{j/i}(x_j)^2 + f_{ij/i}(1-x_j)x_j} \tag{3.24}$$

$$x_{ij}^s = \frac{f_{ij/i}(1-x_j)x_j}{(1-x_j)^2 + f_{j/i}(x_j)^2 + f_{ij/i}(1-x_j)x_j} \tag{3.25}$$

Therefore, substituting equations (3.23) - (3.25) into equation (3.21), we conclude that:

$$Y = \frac{Y_i(1-x_j)^2 + Y_j f_{j/i}(x_j)^2 + Y_{ij} f_{ij/i}(1-x_j)x_j}{(1-x_j)^2 + f_{j/i}(x_j)^2 + f_{ij/i}(1-x_j)x_j} \tag{3.26}$$

However, although this model proved to be very successful with many solvent mixtures [Bosch et al., 1996a, Ràfols et al., 1997], it could not be applied to solvent mixtures of water and alcohols [Rosés et al., 1997]. The reason was the enhancement of the water structure caused by the presence of alcohol. In order to take into account this behaviour, Rosés et al. [1997] added an extra term to equation (3.26), ΔY, to represent the effect of the enhancement. Assuming that the enhancement of the water structure depends on the mole fraction of alcohol molecules and already structured water clusters, they assumed that the modification (ΔY) is proportional to the product of the mole fractions. The mole fraction of structured water in the microsphere of solvation of the indicator is x_j^s, where water is component j. But the alcohol molecules i exist in alcohol (x_i^s) and water-alcohol (x_{ij}^s) clusters. Therefore,

$$\Delta Y = c x_j^s (x_i^s + \frac{x_{ij}^s}{2}), \tag{3.27}$$

where c is a proportionality constant. Substituting equations (3.23) - (3.25) into equation (3.27), we obtain

$$\Delta Y = \frac{c f_{j/i} (x_j)^2 [(1 - x_j)^2 + f_{ij/i}(1 - x_j) x_j / 2]}{[(1 - x_j)^2 + f_{j/i}(x_j)^2 + f_{ij/i}(1 - x_j) x_j]^2}. \tag{3.28}$$

Finally, the combination of equations (3.27) and (3.28) gives the appropriate relation for the calculation of the solvatochromic properties for solvent mixtures of water and alcohols,

$$Y = \frac{Y_i (1 - x_j)^2 + Y_j f_{j/i} (x_j)^2 + Y_{ij} f_{ij/i} (1 - x_j) x_j}{(1 - x_j)^2 + f_{j/i} (x_j)^2 + f_{ij/i} (1 - x_j) x_j} + \Delta Y. \tag{3.29}$$

Another model that has been applied for the correlation of the solvatochromic parameters with solvent composition is the CNIBS/R - K model, used by Harifi-Mood et al. [2006] for the prediction of the solvatochromic parameters of binary solvent mixtures of an ionic liquid, 1-(1-butyl)-3-methylimidazolium tetrafluoroborate ([bmin]BF_4), with water, ethanol and methanol. Moreover, it has been applied by Habibi-Yangjeh [2004] to various aqueous and organic binary solvent systems. According to the CNIBS/R - K model, the solvatochromic parameters (Y) in a binary solvent mixture, can be expressed as

$$Y_m = x_i Y_i + x_j Y_j + x_i x_j \sum_{l=0}^{3} A_l (x_i - x_j)^l \tag{3.30}$$

where SP_m, SP_i, SP_j are the solvatochromic parameters determined in mixed and pure solvents i and j, respectively, and x_i, x_j are the mole fractions of the binary solvent mixture. A_l are the equation coefficients.

In this work, the preferential solvation model of Rosès and co-workers is applied and, particularly, equation (3.26) is used throughout for the calculation of the solvatochromic parameters for the mixed solvents.

Predicted Solvatochromic Parameters

Equation (3.26) has three parameters, $f_{j/i}$, $f_{ij/i}$ and Y_{ij}, which are estimated based on experimental data, in order to calculate the solvent property Y at different compositions. Unfortunately, there are currently limited data for the solvatochromic parameters of GXLs in the literature. We consider three organic solvents in combination with CO_2 here: acetonitrile [Ford et al., 2008b], acetone and methanol [Ràfols et al., 1997]. In figures 3.2-3.4, results from parameter estimation for the model

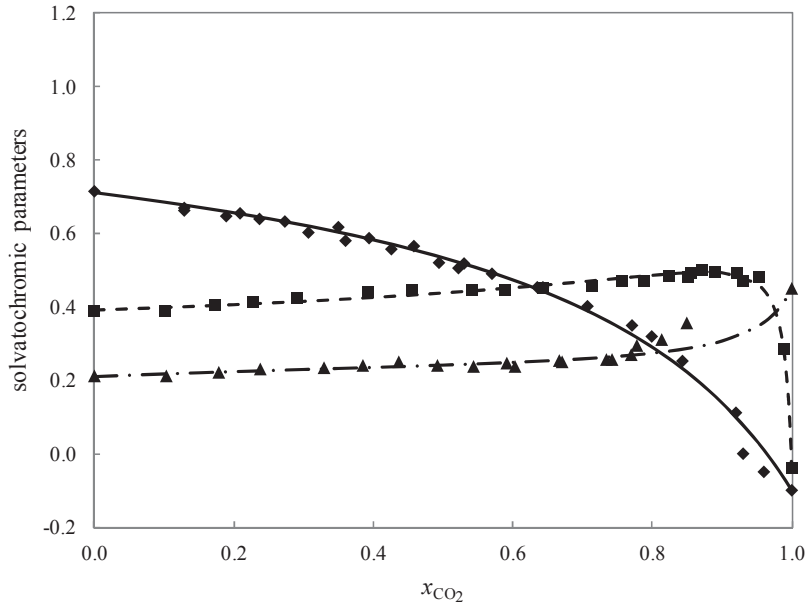


Figure 3.2: Calculated solvatochromic parameters for CO_2 + acetonitrile at $T = 40°C$ (curves) compared to experimental data (symbols), [Ford et al., 2008b] . Solid curve, diamonds: π^*, dashed curve, squares: β, dash-dot curve, triangles: α.

described are shown for the three mixed solvents. The solvatochromic parameters are given as a function of mole fraction of CO_2 in the mixture (i.e., $j = CO_2$), at a temperature $T = 40$°C. The globally optimal values for parameters $f_{j/i}$, $f_{ij/i}$ and Y_{ij}, as determined by the BARON software [Tawarmalani and Sahinidis, 2005, GAMS Development Corporation, 2011] and using a lower bound 0.01 for the parameters $f_{j/i}$ and $f_{ij/i}$, are given in Table 3.1; an excellent fit is obtained over the entire range of CO_2 compositions for the three organic solvents, even in the case of α for

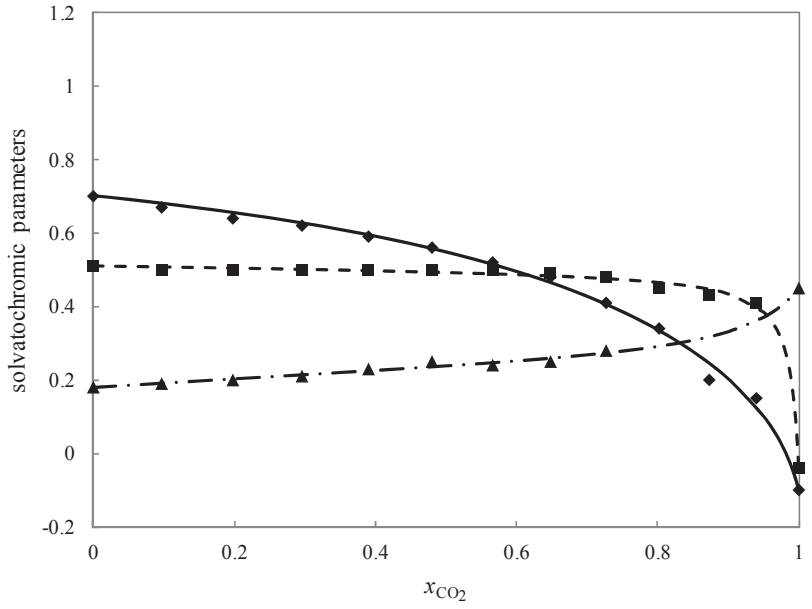


Figure 3.3: Calculated solvatochromic parameters for CO_2 + acetone at $T = 40°C$ (curves) compared to experimental data (symbols), [Wyatt et al., 2005]. Solid curve, diamonds: π^*, dashed curve, squares: β, dash-dot curve, triangles: α.

methanol + CO_2, which exhibits highly nonlinear behaviour.

	acetonitrile (i) + CO_2 (j)			acetone (i) + CO_2 (j)			methanol (i) + CO_2 (j)		
	π^*_{ij}	β_{ij}	α_{ij}	π^*_{ij}	β_{ij}	α_{ij}	π^*_{ij}	β_{ij}	α_{ij}
$f_{j/i}$	0.2525	0.0037	0.1192	0.0100	0.0100	0.1125	0.0100	0.0100	0.0100
$f_{ij/i}$	1.2532	0.3493	2.2782	0.344	0.8031	1.4788	0.7801	0.7801	15.9348
Y_{ij}	0.5194	0.5746	0.2441	0.1211	0.4764	0.2624	0.0385	0.0385	1.0011

Table 3.1: Estimated parameters for the preferential solvation model for the mixed solvents.

3.4.2 Thermodynamic model

The use of GXLs or other solvents is especially desirable when some of the reactants are solid at the reaction conditions. In such cases, when operating at maximum

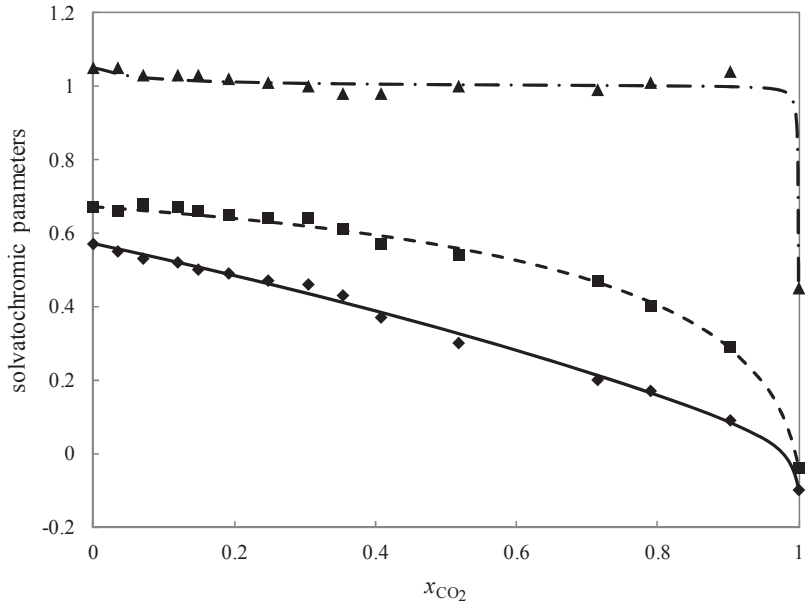


Figure 3.4: Calculated solvatochromic parameters for CO_2 + methanol at $T = 40°C$ (curves) compared to experimental data (curves), [Wyatt et al., 2005]. Solid curve, diamonds: π^*, dashed curve, squares: β, dash-dot curve, triangles: α.

capacity, the reactor mixture is expected to be at solid-vapour-liquid equilibrium (SVLE) so that the limiting (solid) reactant is present at maximum concentration in the liquid phase, as defined by its solubility. For simplicity, it is assumed that reactant A is a solid.

The isofugacity condition is applied for vapour-liquid (VLE) and solid-liquid equilibrium (SLE), expressed in terms of fugacity coefficients:

$$\hat{y}_i \phi_i^v(T, P, \underline{\hat{y}}) = \hat{x}_i \phi_i^l(T, P, \underline{\hat{x}}), \quad i = \mathrm{CO_2}, \mathrm{CS} \tag{3.31}$$

and

$$f_{\mathrm{A}}^s(T, P) = \hat{x}_{\mathrm{A}} \phi_{\mathrm{A}}(T, P, \underline{\hat{x}}), \tag{3.32}$$

where the index i runs over the compounds present in the liquid and vapour phases (here CO_2 and the organic co-solvent, referred to as CS above), ϕ^v denotes the fugacity coefficient of component i for the vapour phase and ϕ^l that for the liquid phase. Component A, the solid, is assumed involatile and is only present in the solid and liquid phase. The isofugacity condition in this case can be written as shown in

equation (3.32), where, f_{A}^{s} is the fugacity of species A in the solid phase (dependent only on the temperature and pressure as it is assumed to be pure), $\underline{\hat{x}}$ is the vector of liquid phase mole fractions and $\underline{\hat{y}}$ that of vapour phase mole fractions, T is the temperature and P the pressure. The fugacity coefficients are calculated using the group-contribution volume translated Peng-Robinson equation of state (GC-VTPR EoS) [Ahlers et al., 2004].

The fugacity of the solid phase can be expressed as a function of the fugacity coefficient of the pure solid, ϕ_{A}^{s}, the sublimation pressure of reactant A, P_{A}^{sub} and the Poynting correction factor, Po_{A}, so that

$$\phi_{\mathrm{A}}^{s}(T, P^{sub}) P_{\mathrm{A}}^{sub} Po_{\mathrm{A}} = \hat{x}_{\mathrm{A}} \phi_{\mathrm{A}}^{l}(T, P, \underline{\hat{x}}). \tag{3.33}$$

The Poynting factor, Po_{A}, is given by the relation:

$$Po_{\mathrm{A}} = \exp\left(\frac{v_{\mathrm{A}}^{s}(P - P_{\mathrm{A}}^{sub})}{RT}\right), \tag{3.34}$$

where v_{A}^{s} is the molar volume of component A in the solid phase. The fugacity coefficient of the pure solid is assumed to be unity at the very low pressure P_{A}^{sub}.

We note that the solubility of the solid could alternatively be obtained based on a thermodynamic route that involves a solid-liquid transition [Sandler, 1999, Poling et al., 2007], rather than the solid-gas transition used above. Both routes are applicable provided that data for the relevant transitions can be found.

Group-Contribution Volume Translated Peng-Robinson Equation of State

The fugacity coefficients are calculated using the group contribution Volume Translated Peng Robinson equation of state (GC-VTPR EoS) [Ahlers et al., 2004]. The GC-VTPR EoS is a predictive model that requires only the UNIFAC interaction parameters of the atom groups that make up the components in the mixture and the critical values for pressure, temperature and volume of the pure components. It is based on the VTPR equation of state

$$P = \frac{RT}{v + c - b} - \frac{a}{(v + c)(v + c + b) + b(v + c - b)}, \tag{3.35}$$

where a, b and c are mixture-specific parameters that depend on composition, P is the pressure, T is the temperature, v is the molar volume and R is the gas constant. The three parameters a, b and c that describe the mixture can be obtained from mixing rules. The mixing rules used in the model for the parameters a and b are those suggested by Chen and Gmehling [2002]:

$$\frac{a}{b} = \sum_{i=1}^{n} \frac{a_i}{b_i} + \frac{g_{res}^E}{A}, \quad A = -0.53087, \tag{3.36}$$

$$b = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j b_{ij}, \tag{3.37}$$

$$b_{ij}^{0.75} = \frac{b_i^{0.75} + b_j^{0.75}}{2}, \tag{3.38}$$

where a_i, b_i are the pure component parameters, n is the number of components, x_i, x_j are the mole fractions of components i and j respectively, and g_{res}^E is the residual part of the excess Gibbs free energy and is calculated with the modified-UNIFAC (Dortmund) model [Gmehling et al., 2002]. The volume translation parameter, c, does not affect vapour-liquid [Peneloux et al., 1982] nor solid-vapour-liquid equilibrium calculations [Ahlers et al., 2004]. For c the following mixing rule is used [Peneloux et al., 1982, Sandler, 1999]:

$$c = \sum_{i=1}^{n} x_i c_i, \tag{3.39}$$

where c_i is the pure component parameter.

The parameters, a_i and b_i, for a pure component are given by the following relations:

$$a_i = 0.45724 \frac{R^2 T_{c,i}^2}{P_{c,i}} \alpha_i(T), \tag{3.40}$$

$$b_i = 0.07780 \frac{R T_{c,i}}{P_{c,i}}, \tag{3.41}$$

and the translation parameter, c_i,

$$c_i = -0.252 \frac{R T_{c,i}}{P_{c,i}} (1.5448 z_{c,i} - 0.4024), \tag{3.42}$$

where $T_{c,i}$, $P_{c,i}$ are the critical temperature and pressure of component i, respectively, and $z_{c,i}$ is the compressibility factor at the critical point:

$$z_{c,i} = \frac{P_{c,i} v_{c,i}}{R T_{c,i}}, \tag{3.43}$$

where $v_{c,i}$ is the critical volume of component i.

The value of $\alpha_i(T)$ can be obtained from the expression of Twu et al. [1995] for any pure component i for which data are available [Ahlers and Gmehling, 2002, Ahlers et al., 2004]:

$$\alpha_i(T) = T_{r,i}^{N_i(M_i-1)} \exp\left[L_i(1 - T_{r,i}^{N_i M_i})\right] \tag{3.44}$$

where $T_{r,i}$ is the reduced temperature, equal to $T/T_{c,i}$, and L_i, M_i, N_i are compound-specific parameters. When no experimental data are available, the following generalised expressions are used [Twu et al., 1995]:

$$\alpha_i(T) = \alpha_i^{(0)} + \omega_i(\alpha_i^{(1)} - \alpha_i^{(0)}), \tag{3.45}$$

where ω_i is the acentric factor of component i and, for $T_{r,i} < 1$,

$$\alpha_i^{(0)} = T_{r,i}^{-0.1883273} \exp[0.1048767(1 - T_{r,i}^{2.1329765})], \tag{3.46}$$

$$\alpha_i^{(1)} = T_{r,i}^{-0.6029386} \exp[0.5113343(1 - T_{r,i}^{2.2059312})], \tag{3.47}$$

while for $T_{r,i} \geq 1$,

$$\alpha_i^{(0)} = T_{r,i}^{-0.792651} \exp[0.401219(1 - T_{r,i}^{-0.992615})], \tag{3.48}$$

$$\alpha_i^{(1)} = T_{r,i}^{-1.98471} \exp[0.024955(1 - T_{r,i}^{-9.98471})]. \tag{3.49}$$

Returning to the calculation of the fugacity coefficients, the general expression for the fugacity coefficient is [Sandler, 1999]:

$$\ln \phi_i = \frac{1}{RT} \int_0^P \left(v_i - \frac{RT}{P} \right) dP. \tag{3.50}$$

In our case, the fugacity coefficient for component i from VTPR EoS, and with equations (3.36) to (3.37) for the mixing rules, is given by the following equation:

$$\ln \phi_i = \frac{S_2}{b}(z-1) - \ln(z-B) - \frac{a}{bRT\sqrt{8}} \left[\frac{S_1}{a} - \frac{S_2}{b} \right] \ln \left[\frac{z + (1+\sqrt{2})B}{z + (1-\sqrt{2})B} \right], \tag{3.51}$$

where

$$S_1 = S_2 \frac{a}{b} + b \left(\frac{a_i}{b_i} + \frac{1}{A} RT \ln \gamma_i \right), \tag{3.52}$$

$$S_2 = 2\sum_j^n x_j b_{ij} - b, \tag{3.53}$$

$$B = \frac{bP}{RT}, \tag{3.54}$$

where z is the compressibility factor of the mixture

$$z = \frac{P \cdot (v + c)}{RT}, \tag{3.55}$$

and γ_i is the activity coefficient of component i calculated with the modified-UNIFAC (Dortmund) model [Gmehling et al., 2002].

3.4.3 Process model

The process flowsheet is given in figure 3.1. For this conceptual process, it is assumed that the separation is perfect, so that no product is recycled and there are no losses of reactants or solvents. Thus, there is no need for a fresh solvent feed. The desired production flowrate, $N_{\mathrm{C},4}$, and the reaction temperature are taken as fixed. The optimal co-solvent, defined by vector $\underline{y}$, the optimal composition of the GXL, as given by the mole fraction of CO_2 relative to that of co-solvent, x_{CO_2}, and the optimal single-pass conversion, ϵ, are sought.

The reaction rate r, in mol m^{-3} s^{-1}, is given by

$$r = r(k, C_{\mathrm{A},2}, C_{\mathrm{B},2}) \tag{3.56}$$

where k is the reaction rate constant given by equation (3.3), in m^3 mol s^{-1}, and $C_{\mathrm{A},2}$, $C_{\mathrm{B},2}$ are the concentrations of reactants A and B in the reactor in mol m^{-3}. The reactor mole balance for species i is

$$N_{i,1} + N_{i,3} - N_{i,2} + \nu_i r V_R = 0, \quad i = \mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{CO_2}, \mathrm{CS}, \tag{3.57}$$

where $N_{i,1}$ is the fresh feed of i (with $N_{\mathrm{C},1} = N_{\mathrm{CO_2},1} = N_{\mathrm{CS},1} = 0$), $N_{i,2}$ is the molar flowrate of i in the stream leaving the reactor, $N_{i,3}$ is the recycle flowrate of i ($N_{\mathrm{C},3} = 0$). All molar flowrates are in mol s^{-1}. ν_i is the stoichiometric coefficient for component i (with $\nu_{\mathrm{CO_2}} = \nu_{\mathrm{CS}} = 0$) and V_R is the reactor volume in m^3.

The single-pass conversion, ϵ, is defined by

$$N_{\mathrm{A},2} = (1-\epsilon)(N_{\mathrm{A},1} + N_{\mathrm{A},3}) \tag{3.58}$$

Reactant B is assumed to be present in excess by a factor f_e

$$N_{\mathrm{B},2} = f_e N_{\mathrm{A},2} \tag{3.59}$$

The mole fractions in the reactor are the same as in the reactor outlet and given by

$$x_{R,i} = \frac{N_{i,2}}{\sum\limits_{i\in\{\mathrm{A,B,C,CO_2,CS}\}} N_{i,2}}, \quad i = \mathrm{A,B,C,CO_2,CS}. \tag{3.60}$$

The concentrations of A and B in the reactor, $C_{\mathrm{A},2}$ and $C_{\mathrm{B},2}$, are given by the relations:

$$C_{i,2} = \frac{x_{R,i} M_{R,i}}{V_R}, \quad i = \mathrm{A,B}, \tag{3.61}$$

where $M_{R,i}$ is the number of moles of component i in the reactor and $x_{R,i}$ is the mole fraction of component i in the reactor. The GXL composition is given by

$$x_{\mathrm{CO_2}} = \frac{x_{R,\mathrm{CO_2}}}{x_{R,\mathrm{CO_2}} + x_{R,\mathrm{CS}}}. \tag{3.62}$$

For the phase equilibrium calculations, it is assumed that species B and C behave ideally and are present only in the liquid phase, so that the phase equilibrium calculations are carried out for three components only: component A, the organic co-solvent, and CO_2, on the basis of scaled mole fractions $\hat{x}_i$, such that $\hat{x}_{\mathrm{A}} + \hat{x}_{\mathrm{CO_2}} + \hat{x}_{\mathrm{CS}} = 1$. The mixture is denoted by $\mathrm{A} + GXL$. The scaled mole fractions are therefore defined as

$$\hat{x}_i = \frac{x_{R,i}}{x_{R,\mathrm{A}} + x_{R,\mathrm{CO_2}} + x_{R,\mathrm{CS}}}, \quad i = \mathrm{A,CO_2,CS}. \tag{3.63}$$

The mole numbers in the reactor (in mol), $M_{R,i}$ for component i, M_R for total number, are related by

$$M_{R,i} = x_{R,i} M_R, \quad i = \mathrm{A,B,C,CO_2,CS}. \tag{3.64}$$

The volume of the mixture of A and GXL (in m^3), $V_{\mathrm{A}+GXL}$, is

$$V_{\mathrm{A}+GXL} = (M_{\mathrm{A}} + M_{\mathrm{CO_2}} + M_{\mathrm{CS}})\, v_{\mathrm{A}+GXL}, \tag{3.65}$$

where $v_{\mathrm{A}+GXL}$ is the molar volume of the liquid phase of the A+GXL mixture (in $\mathrm{m}^3\ \mathrm{mol}^{-1}$), as calculated in the phase equilibrium model, and $V_{\mathrm{A}+GXL}$ is the corresponding volume in the reactor (in m^3). V_{B} and V_{C}, the volumes of reactant B and the product C, respectively, are calculated from their pure component densities under the assumption of ideal mixture behaviour:

$$V_i = \frac{M_{R,i} Mr_i}{\rho_i}, \quad i = \mathrm{B}, \mathrm{C}, \tag{3.66}$$

where Mr_i is the molar mass of i (in $\mathrm{g.mol}^{-1}$) and ρ_i is the mass density of i (in $\mathrm{g.m}^{-3}$). The volume of the reactor is therefore given by

$$V_R = V_{\mathrm{A}+GXL} + V_{\mathrm{B}} + V_{\mathrm{C}}. \tag{3.67}$$

The mass balances for the separator are as follows:

$$\begin{aligned} &N_{i,2} = N_{i,3}, \quad i = \mathrm{A}, \mathrm{B}, \mathrm{CO_2}, \mathrm{CS}, \\ &N_{\mathrm{C},2} = N_{\mathrm{C},4} \end{aligned} \tag{3.68}$$

The specific co-solvent used is determined by defining binary variables y_{ACN}, y_{DMK} and y_{MeOH}, which take a value of 1 if acetonitrile, acetone or methanol, respectively, is present in the GXL, and a value of 0 otherwise. To ensure that exactly one co-solvent is selected, they are such that

$$y_{\mathrm{ACN}} + y_{\mathrm{DMK}} + y_{\mathrm{MeOH}} = 1. \tag{3.69}$$

All coefficients relating to the co-solvent CS in the model are set through linear relations to these binary variables. For example,

$$f_{\mathrm{CO_2/CS}} = y_{\mathrm{ACN}} f_{\mathrm{CO_2/ACN}} + y_{\mathrm{DMK}} f_{\mathrm{CO_2/DMK}} + y_{\mathrm{MeOH}} f_{\mathrm{CO_2/MeOH}}. \tag{3.70}$$

3.4.4 Estimating process costs

The results of the material balances can be used to estimate the size of the units and the utility requirements. On this basis, the total cost of the process can be evaluated by considering capital and operating costs. For the flowsheet considered here, the main items of capital cost consist of the costs arising from the reactor, the separation

unit, the compressor and the organic solvent, while the cost of the steam and the cooling water for the separation and of the electricity for the compressor contribute to the operating cost.

Correlations from Douglas [1988] are used to obtain the relevant costs and are listed here for completeness. The annualised installed cost of the reactor, C_R, is thus given by

$$C_R(\$\text{pa}) = \left(\frac{M\&S}{280}\right) 482.37 V_r^{0.6287}(2.18 + F_p F_m), \tag{3.71}$$

where $M\&S$ is the Marshall and Swift equipment index for 2010, F_p is a correction factor that depends on the pressure. The value of F_m is set to be 1 (assuming carbon steel), while for F_p, a second order polynomial is used:

$$F_p = 0.0255P^2 + 0.0387P + 1.0136. \tag{3.72}$$

The annualised cost of the organic solvent is calculated using the following expression:

$$C_{\text{CS}}(\$\text{pa}) = \frac{1}{3} c_{\text{CS}} m_{\text{CS}}, \tag{3.73}$$

where subscript CS corresponds to the organic co-solvent, m_{CS} is inventory of co-solvent (in kg), derived by multiplying the mass of co-solvent in the reactor by 2, and c_{CS} is the cost of the solvent per kg, which is obtained from Sigma-Aldrich. Values for cost coefficients and other design data are listed in table 3.2.

The capital cost of the separator, C_{SEP}, is calculated as the cost of two heat exchangers, namely an evaporator to evaporate the organic solvent and recover the solid product and a condenser to condense the organic solvent again. The overall temperature change is from the reactor temperature of 40^{o}C to the boiling point of the solvent. The installed cost of each heat exchanger is calculated from the following equation [Douglas, 1988]

$$C_{SEP}(\$\text{pa}) = \left(\frac{M\&S}{280}\right) 23.71 \left(A_{evap}^{0.65} + A_{cond}^{0.65}\right), \tag{3.74}$$

where A_{evap} and A_{cond} are the areas of the evaporator and condenser respectively, in m^2. The areas are obtained by assuming that evaporation is achieved using saturated steam at 15 psig, and condensation is achieved with cooling water.

Quantity	Value	Units
c_{ACN}, acetonitrile price	108	\$ kg^{-1}
c_{DMK}, acetone price	40	\$ kg^{-1}
c_{MeOH}, methanol price	36	\$ kg^{-1}
Heat transfer coefficients [Douglas, 1988]		
condensing gas to vaporising liquid	1400	W m^{-2} K^{-1}
condensing gas to liquid	850	W m^{-2} K^{-1}
liquid to liquid	300	W m^{-2} K^{-1}
$M\&S$ cost factor [Chemical Engineering Magazine, December 2010]	1457.4	–
c_{st} steam price	0.03	\$ kWh^{-1}
c_w cooling water price	0.005	\$ kWh^{-1}
c_{ele} electricity price	0.06	\$ kWh^{-1}
Cooling water inlet temperature	298.15	K
Cooling water outlet temperature	313.15	K
η, compressor efficiency	0.9	–
Operating hours	8000	h per year

Table 3.2: Design and cost data for the case study. The utility prices were provided by an industrial user of utilities.

The annualised installed capital cost of the compressor, C_{COMP}, is given as follows [Douglas, 1988]

$$C_{COMP}(\$\mathrm{pa}) = \left(\frac{M\&S}{280}\right) 536.475\mathrm{bhp}^{0.82}. \tag{3.75}$$

The brake horsepower, bhp, in units of hp, is given by

$$\mathrm{bhp} = \frac{P_{hp}}{\eta}, \tag{3.76}$$

where P_{hp} is the utility requirement assuming 100% efficiency in hp units, and η is the efficiency of the compressor. The utility requirement is related to the flowrate of CO_2 as follows

$$P_{hp} = \frac{3.0310^{-5}}{\gamma} P_{in} Q_{in} \left(\left(\frac{P_{out}}{P_{in}}\right)^{\gamma} - 1\right), \tag{3.77}$$

where Q_{in} is the flow that enters the compressor in ft^3/min (based on $N_{CO_2,3}$, P_{in} and P_{out} are the initial and final pressures in lbf/m^2 and $\gamma = 0.23$ [Douglas, 1988].

The operating cost includes the cost of the steam, C_{st}, and the cooling water, C_w, for the separation unit and the electricity, C_{ele}, used by the compressor on an annual basis:

$$C_{st}(\$) = Q_{st} c_{st}, \tag{3.78}$$

$$C_w(\$) = Q_{cw} c_{cw}, \tag{3.79}$$

$$C_{ele}(\$) = P_{ele} c_{ele}, \tag{3.80}$$

where Q_{st} is the heating requirement for evaporation, Q_{cw} the cooling requirement for condensation and Q_{ele} the compressor energy requirement.

The total cost, C_{total}, in \$ per annum, is

$$C_{total} = C_R + C_{SEP} + C_{COMP} + C_{\mathrm{CS}} + C_{st} + C_w + C_{ele}. \tag{3.81}$$

3.5 Case study

The methodology presented for the design of a GXL is applied to the Diels-Alder reaction of anthracene with 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) to form the adduct (8,9,10,11-dibenzo-4-phenyl-2,4,6-triaza[5,2,2,0] tricyclo-undeca-8,10-diene-3,5-dione) (figure 3.5), based on a production rate of 1 mol s^{-1} of adduct. The kinetics of this reaction have been studied in acetonitrile + CO_2 mixtures by Ford et al. [2008b]. Data were obtained at a temperature of $T = 40$ °C, under pseudo first-order conditions, by using an excess of PTAD. Anthracene, a non-polar compound, has limited solubility in the polar co-solvents considered here and its solubility is a determining factor in identifying the optimal GXL composition and equipment size. Given the higher expected solubility of PTAD, it is assumed to be present in excess in the reactor by a factor of five so that $f_e = 5$. The density of the adduct is assumed to be equal to 1.2 g cm^{-3}. Other physical properties are listed in table 3.3.

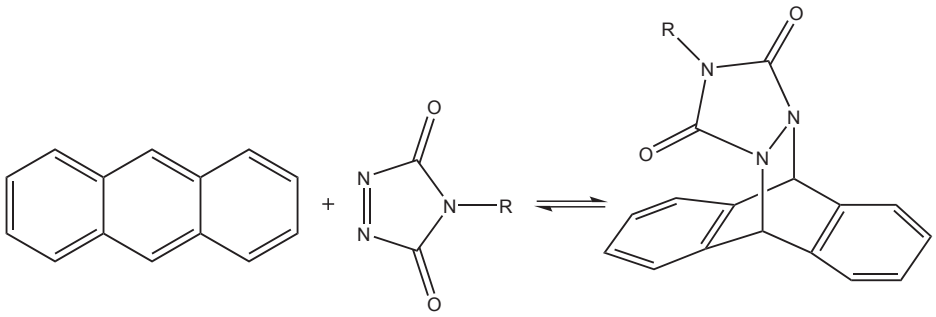


Figure 3.5: The Diels-Alder reaction of anthracene and 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD)to form 8,9,10,11-dibenzo-4-phenyl-2,4,6-triaza[5,2,2,0] tricy- clo-undeca-8,10-diene-3,5-dione (adduct).

	ΔH_{vap} (J mol^{-1})	T_c (K)	P_c (MPa)	z_c	ω	T_b (K)	C_p (J mol^{-1} K^{-1})
acetonitrile	29840^a	545.5^b	4.83^b	0.184^c	0.321^d	354.71^a	82^e
acetone	29100^f	509.5^g	4.76^g	0.2348^g	0.311^g	329.45^f	131^h
methanol	35200^k	512.6^l	8.096^l	0.2242^l	0.559^l	337.8^k	88^m
water	40670^n	–	–	–	–	–	75.3^o

Table 3.3: Physical properties used. The heat of vaporisation, ΔH_{vap}, the critical temperature, T_c, the critical pressure, P_c, the critical compressibility factor, z_c, the acentric factor, ω, the normal boiling point, T_b, the liquid heat capacity, C_p, parameters D, E, F of Antoine equation ($\log P^{sub}$[mmHg] $= D - E/(F + T[^oC])$) for the calculation of the sublimation pressure P^{sub} of anthracene, molar volume v^s. a: Antosik et al. [2004], b: Ewing and Ochoa [2004], c: Simmrock et al. [1986], d: Khurma et al. [1983], e: Mirzaliev et al. [1987], f: Hopfe [1990], g: Liessmann et al. [1995], h: Rastorguev and Ganiev [1967], k: Matyushov and Schmid [1994], l: Ahlers et al. [2004], m: Davila and Trusler [2009], n: Antosik et al. [2004], o: Anouti et al. [2009], p: Ahlers et al. [2004]

3.5.1 Reaction rate constant

The solvatochromic equation proposed by Ford et al. [2008b], based on the form of equation (3.82) and their kinetic data in acetonitrile + CO_2, is used to obtain the

pseudo first-order rate constant k_i as a function of the properties of a GXL containing the co-solvent i and mole fraction x_{CO_2}

$$\ln k_i(x_{CO_2}) = 1.9 - 2.62\pi_i^*(x_{CO_2}) - 4.68\alpha_i(x_{CO_2}) + 1.58\beta_i(x_{CO_2}), \tag{3.82}$$

The application of this equation to calculate the reaction rate constant in the three mixed solvents is shown in figure 3.6, where k_i is given as a function of the mole fraction of CO_2 in the mixture. The solvatochromic equation provides a good fit for

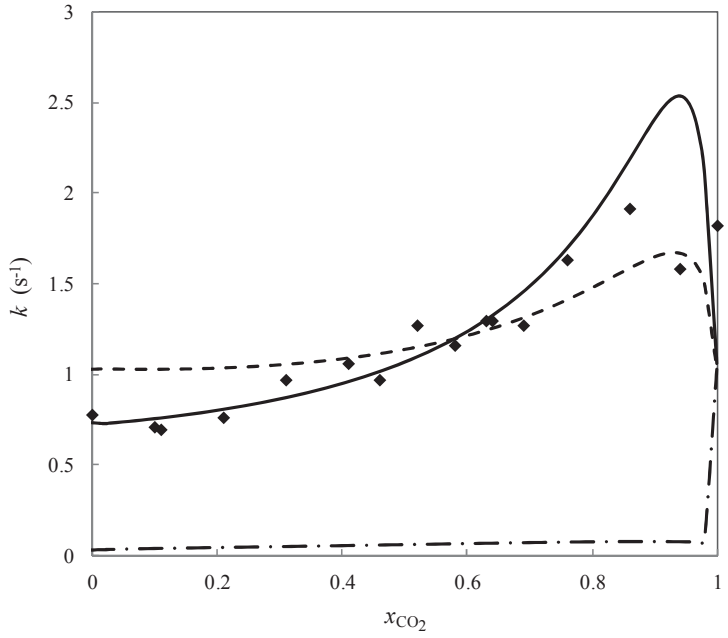


Figure 3.6: Calculated pseudo first-order reaction rate constant, k_i, in mixed solvents, as a function of CO_2 mole fraction, x_{CO_2}. Solid curve: acetonitrile + CO_2 , dashed curve: acetone + CO_2, dash-dot curve: methanol + CO_2. Symbols: experimental data for acetonitrile + CO_2 [Ford et al., 2008b].

the acetonitrile co-solvent data for CO_2 mole fractions up to 0.85, but significant differences can be seen for the two data points with the highest CO_2 content. Due to the low solubility of anthracene at such high concentrations, larger errors can be expected in the reported rate constants; indeed, this would explain the surprising decrease in rate constant at about 0.95 CO_2 mole fraction. This indicates that greater model uncertainty can be expected at high CO_2 concentration, and care should be taken

in interpreting any solution of the design problem around such values. In all mixed solvents, the rate constant is predicted to increase with increasing the mole fraction of CO_2, up to about 0.95. The calculated increase is greater when the co-solvent is acetonitrile than when it is acetone. In contrast, the Diels-Alder reaction is predicted to be very slow in CO_2 + methanol at most concentrations, with a sharp increase to as the pure CO_2 limit approaches, consistent with the behaviour of α for this mixture. It should be noted that the coefficients of the solvatochromic equation are based only on acetonitrile + CO_2 data and, although the solvatochromic parameters of the other solvent mixtures are accurate (figures 3.3 and 3.4), the extrapolated rate constants for the other solvents can be expected to be less accurate.

3.5.2 Thermodynamic model

The reactor contains a gas-expanded liquid which consists of anthracene, PTAD, the adduct, CO_2 and the organic co-solvent. Anthracene is assumed to be at its solubility limit, so that the solid-liquid equilibrium equation is applied (equation (3.33)). The excess reactant, PTAD, and the adduct are assumed to be present only in the liquid phase and to behave ideally. CO_2 and the organic co-solvent are at vapour-liquid equilibrium.

The predictions of the group-contribution VTPR EoS are compared against available experimental data for vapour-liquid equilibrium for the three GXLs of interest are shown in figure 3.7. The VLE predictions are in reasonable agreement with experimental data over the whole range of CO_2 compositions. The percentage average absolute deviations in pressure for the three mixtures are as follows: 15.3% for acetonitrile + CO_2, 1.8% for acetone + CO_2 and 17.2% for methanol + CO_2. In the case of the methanol + CO_2 mixture, the onset of vapour-liquid-liquid equilibrium (VLLE) is seen at high pressures, for an overall CO_2 mole fraction of approximately 0.6. In figure 3.8, the solubility of anthracene in each binary solvent, as calculated by applying the solid-vapour-liquid equilibrium model to several three component mixtures, is shown as a function of pressure (equivalently x_{CO_2}). The model provides a good prediction of the solubility of anthracene in pure acetone and in pure ace-

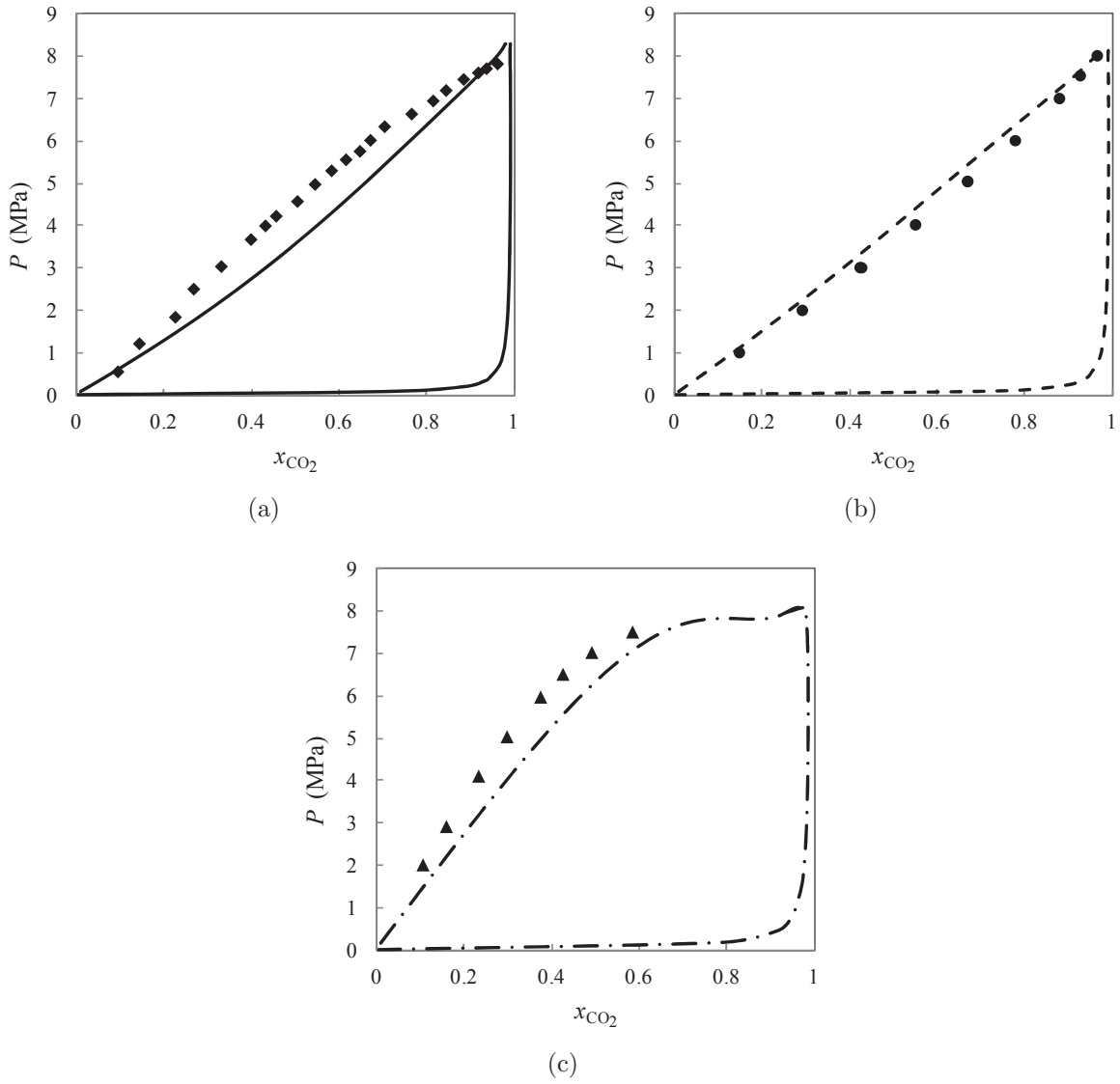


Figure 3.7: Vapour-liquid equilibrium for binary mixtures at $T = 40°C$. (a) acetonitrile + CO_2 (data from Kordikowski et al. [1995]); (b) acetone + CO_2 (data from Adrian and Maurer [1997]); (c) methanol + CO_2 (data from Kodama et al. [1996]). The curves are the predicted phase envelopes with the GC-VTPR EoS, symbols represent experimental data.

tonitrile. These results clearly show that the behaviour of solubility as a function of pressure is the opposite of that of the rate constant, with solubility tending to decrease with increasing CO_2 mole fraction. Slight solubility maxima are predicted in

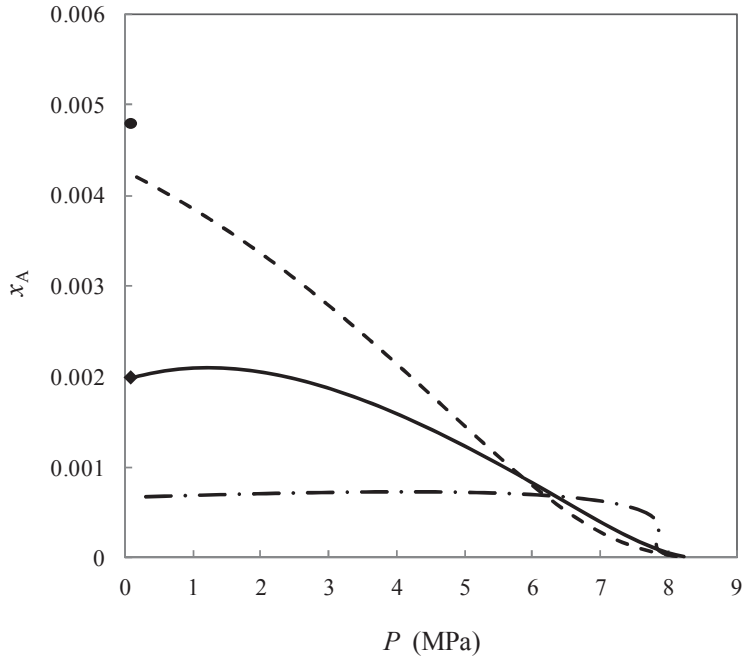


Figure 3.8: Predicted and measured solubility of anthracene in mixed solvents at $T = 40°C$, shown with curves and symbols respectively. Acetonitrile + CO_2: solid curves and diamond [Cepeda and Diaz, 1996]; acetone + CO_2: dashed curve and circle [Petrova, 1974]; methanol + CO_2: dash-dot curve.

the case of CO_2 with acetonitrile and with methanol. This indicates that the optimal GXL design may require a trade-off between solubility and rate constant.

3.5.3 GXL design

The mixed-integer design problem consists of identifying the optimal co-solvent, composition for the GXL and conversion for a fixed production rate of adduct. Since in this case study, only three organic solvents are considered, the MINLP is solved by enumeration. An implementation of the model in gPROMS [Process Systems Enterprise, 1997-2009] is used.

The total cost of the process is given in figure 3.9, for the case of a single-pass conversion of 50%. Mole fractions of CO_2 of up to 0.6 are investigated in the case of methanol + CO_2, in order to avoid the occurrence of VLLE. The cost with methanol +

CO_2 is considerably higher than with the other two mixed solvents, since the reaction in methanol + CO_2 is much slower than in the other solvents (figure 3.6) and the solubility of anthracene is lower (figure 3.8). In all mixed solvents, the cost is seen to increase with increasing CO_2 content in the GXL, although shallow minima are exhibited in the case of acetonitrile + CO_2 and methanol + CO_2. These occur at a low CO_2 mole fraction of $x_{CO_2} = 0.04$; for the acetonitrile co-solvent a minimum cost of approximately \$3.9 million per annum is found, while for the methanol co-solvent the minimum cost is approximately \$12.9 million per annum. In the case of the acetone co-solvent, the minimum occurs when no CO_2 is used. However, the acetone exhibits a competitive cost over a large region of mole fraction of CO_2 (up to 0.8), when compared to the other co-solvents.

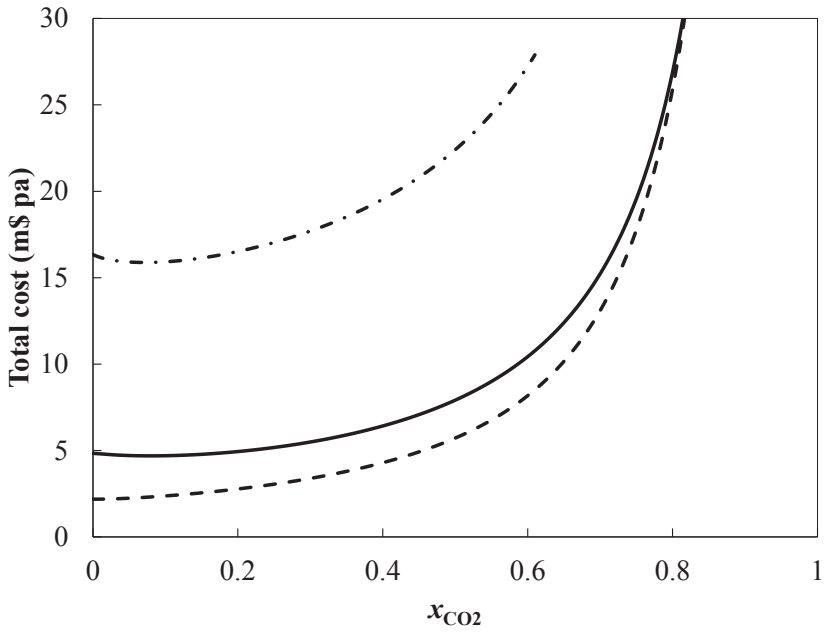


Figure 3.9: Calculated total process cost as a function of CO_2 mole fraction in the GXL for a single-pass conversion of $\epsilon = 0.5$. Acetonitrile + CO_2: solid curve; acetone + CO_2: dashed curve; methanol + CO_2: dash-dot curve.

The performance of the process can be investigated for different single-pass conversions, as shown in figure 3.10 for the case of acetone + CO_2. The total cost decreases with increasing conversion, since a smaller amount of organic co-solvent is needed and thus the costs of the separation unit and the compressor, which dominate the total cost, decrease. Qualitatively, the overall dependence of the cost on CO_2 content remains the same at all conversions.

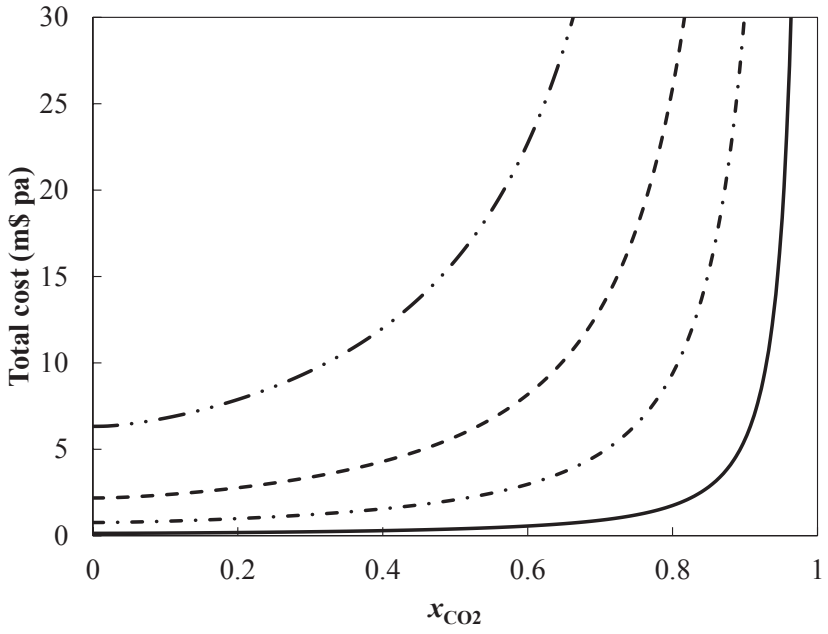


Figure 3.10: Calculated process cost as a function of CO_2 mole fraction for acetone + CO_2 for different single-pass conversions: $\epsilon = 0.25$, dash-dot-dot curve; $\epsilon = 0.50$, dashed curve; $\epsilon = 0.75$, dash-dot curve; $\epsilon = 0.95$, solid curve.

The analysis of the overall cost indicates that the cost is minimised when little or no CO_2 is present (tables 3.4 and 3.5).

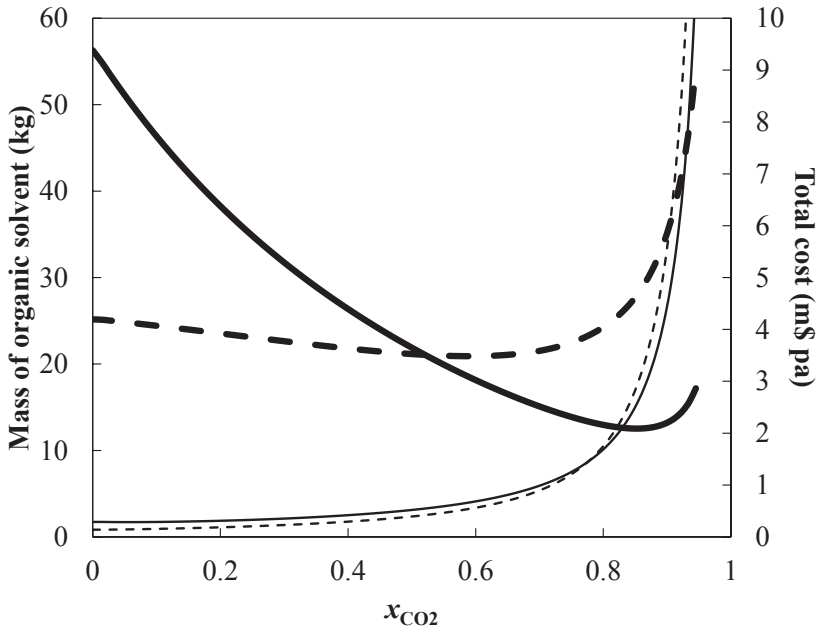


Figure 3.11: Calculated mass of co-solvent (thick curves) and process cost (thin curves) as a function of CO_2 mole fraction at 95% single-pass conversion: acetonitrile + CO_2, solid curves; acetone + CO_2, dashed curves.

In this particular case study, pure acetone leads to the best performance. For methanol and acetonitrile, the use of a small amount of CO_2 (4-7 mol %) brings an economic benefit. Nevertheless, in order to design a process with low environmental impact, it is desirable to find a trade-off between the amount of organic solvent and the total costs. This is investigated by considering a single-pass conversion of 95%, which affords the best economic performance, and examining the total cost and the solvent inventory as a function of GXL composition. For methanol, the minimum solvent mass is reached at the upper bound on CO_2 mole fraction of 0.6 (tables 3.6 and 3.7). The results for acetone and acetonitrile are shown in figure 3.11. As can be seen, the mass of solvent shows a minimum towards higher CO_2 mole fractions, specifically, $x_{CO_2} = 0.595$ for acetone and $x_{CO_2} = 0.855$ for acetonitrile. In the latter

Co-solvent	x_{CO_2}	$x_{R,A}$	P (MPa)	V_R (m^3)	m_{CS} (kg)	A (m^2)	Q_{SEP} (kW)	P_{COMP} (kW)
acetonitrile	0.065	0.00172	0.41	0.06	49.6	51.1	1568	9
acetone	0	0.00308	0.10	0.03	25.2	37.2	734	–
methanol	0.04	0.00064	0.58	2.01	2491.9	203	5517	22

Table 3.4: Key physical and design variables for each co-solvent for single-pass conversion 95% at minimum total cost. In the case of acetone, the use of pure solvent is better economically. x_{CO_2} is the CO_2 mole fraction in the GXL, $x_{R,A}$ the mole fraction of anthracene in the reactor, P the reactor pressure, V_R the reactor volume, m_{CS} the solvent inventory, A the total area for heat exchange, Q_{SEP} the total heat duty in the separation unit, P_{COMP} the compressor power.

case, the high mole fraction of CO_2 corresponds to the region of larger uncertainty in kinetic model (cf figure 3.6), which may lead to an underestimation of the amount of solvent required. Nevertheless, the trend in acetonitrile is clear and the use of a GXL has a significant impact on the solvent requirement, with a decrease of around 77 % for acetonitrile and of around 17 % in the case of acetone. Details of the key design and economic metrics for processes based on minimising the use of organic solvent are listed in tables 3.6 and 3.7. The reduction in solvent use comes at a significant cost, as can be seen by comparing tables 3.5 and 3.7, and incurs a large increase in energy consumption (cf tables 3.4 and 3.6). The capital and operating costs associated with the use of the compressor are seen to be the largest contributors to the overall cost based on the simple process analysis here. We highlight that accounting for solvent losses and introducing less costly separation techniques may also have an impact on the overall behaviour of the process. The data in figure 3.11 provides a useful illustration of the trade-off between economic and environmental performance indicators.

Co-solvent	Capital Costs (10^3\$ pa)				Operating Costs (10^3\$ pa)			Total Cost (10^3\$ pa)
Co-solvent	C_R	C_{SEP}	C_{COMP}	C_{CS}	C_{st}	C_w	C_{ele}	
acetonitrile	1.4	44.5	13.2	1.8	188.1	31.4	4.3	285
acetone	0.9	34.2	0	0.3	88.0	14.7	0	138
methanol	12.9	106.6	27.4	29.6	662.0	110.3	10.4	959

Table 3.5: Cost breakdown for each co-solvent for a single-pass conversion 95% at a GXL composition corresponding to the design with the minimum total cost, as shown in table 3.4. C_R is the reactor cost, C_{SEP} the separation unit cost, C_{COMP} the compressor cost, C_{CS} the co-solvent cost, C_{st} the steam cost, C_w the cooling water cost, and C_{ele} the electricity cost.

Co-solvent	x_{CO_2}	$x_{R,A}$	P (MPa)	V_R (m^3)	m_{CS} (kg)	A (m^2)	Q_{SEP} (kW)	P_{COMP} (kW)
acetonitrile	0.855	0.00042	6.90	0.065	12.5	37.6	1151	2367
acetone	0.595	0.00157	4.47	0.039	20.9	36.2	715	330
methanol	0.60	0.00057	7.22	1.297	632.8	96.3	2620	1239

Table 3.6: Key physical and design variables for each co-solvent for single-pass conversion 95% at minimum inventory of co-solvent. For methanol, the maximum allowable mole fraction of CO_2 is restricted to 0.6 due to the onset of VLLE. Symbols are as in table 3.4.

Moderate amounts of CO_2, up to 10-15 mol % are found to yield a reasonable trade-off. Naturally, in order to establish a firm comparison of the environmental performance of the three co-solvents, quantitative cradle-to-grave environmental impact analysis should be considered, e.g., following the approach adopted for GXL systems [Fang et al., 2007, Ghanta et al., 2012a,b] and supercritical CO_2 systems [Gong et al., 2008].

Indeed, greater advantages may be derived when applying the proposed approach to other reaction systems, as enhanced reaction rates have also been observed for a Menschutkin reaction in CO_2-expanded acetonitrile [Ford et al., 2008a]. However, further application of the design approach requires predictive models not only for the rate constant, but also for the phase equilibria of the reactants and GXL mixtures. Such a model is not yet available for methyl *p*-nitrobenzenesulfonate (MNBS), one of the reactants in the reaction studied by [Ford et al., 2008a].

Co-solvent	Capital Costs (10^3\$ pa)				Operating Costs (10^3\$ pa)			Total Cost (10^3\$ pa)
	C_R	C_{SEP}	C_{COMP}	C_{CS}	C_{st}	C_w	C_{ele}	
acetonitrile	2.1	36.4	1,291.5	0.5	138.2	23.0	1,135.9	2,628
acetone	1.3	33.7	256.7	0.3	85.8	14.3	158.3	550
methanol	14.3	65.8	759.7	7.5	314.5	52.4	594.7	1,809

Table 3.7: Cost breakdown for each co-solvent for a single-pass conversion 95% at a GXL composition corresponding to the design with the minimum inventory of co-solvent, as shown in table 3.6. Symbols are defined as in table 3.5.

3.6 Conclusions

Solvents play a vital role in industry and novel solvent classes such as gas-expanded liquids have been the subject of growing interest, thanks especially to their relatively benign environmental impact. An assessment of their benefits must necessarily include process considerations, and take into consideration changes in capital and operating costs. A methodology for the design of a CO_2-expanded solvent and an associated conceptual process design were presented and applied to a case study for which kinetic data are available, namely the Diels-Alder reaction of anthracene and PTAD. Three organic co-solvents were considered: acetonitrile, acetone and methanol. The effect of co-solvent choice on reaction kinetics was modelled by using a solvatochromic equation in combination with a preferential solvation model, while solid-vapour-liquid phase equilibrium was modelled using the group-contribution volume-translated Peng Robinson (GC-VTPR) equation of state for the fluid phases, and sublimation data for the solid phase. Model calculations were compared to available data and found to offer a good description of the kinetic and thermodynamic properties. On this basis, the GXL designs that achieve minimum process cost or minimum solvent inventory were considered and the impact of solvent composition investigated.

It was found that the use of pure acetone as a solvent results in a lower cost than any GXL, but both acetonitrile and acetone offer good performance for the process over a range of CO_2 concentrations, giving the designer the option to decrease organic solvent use by tuning the operating pressure of the reactor. Effective designs are based on balancing anthracene solubility (highest in the organic solvent) and rate constant (highest at high CO_2 content). Methanol was found to be an inappropriate co-solvent for the reaction studied, as the rate of the reaction is predicted to be very low, leading to very high costs. The proposed methodology highlights the importance of taking multiple process performance indicators into account when designing GXLs and assessing their benefits. The approach can be applied to other processes and can be used to guide the investigation of improved solvent mixtures. A current bottleneck in its application is the limited availability of data on mixed solvents, which makes

the prediction of reaction kinetics challenging. In the rest of this thesis, *ab initio* methods for the prediction of reaction rate constants are investigated and two *ab initio* solvent design methodologies for reactions, in which no experimental data are required, are presented.

Chapter 4

Predicting liquid-phase reaction rate constants

4.1 Introduction

When looking for a suitable solvent for a reaction, predictive models for the reaction rate constant can be very useful, in order to consider a wide range of solvents. Since the 19^{th} century, many attempts to understand the reaction mechanisms have been reported [Laidler, 1987]. One of the early breakthroughs was based on the experimental observations of Arrhenius and the development of his equation in 1889, for the temperature dependence of rate constants [Arrhenius, 1889, Back and Laidler, 1967]

$$k = Ae^{-E_a/RT}, \tag{4.1}$$

where E_a is the activation energy and A is a constant. k, R and T, as defined in earlier chapters, are the reaction rate constant, ideal gas constant and absolute temperature, respectively. According to Arrhenius, the reactants, in order to transform into the products, have to overcome an energy barrier, or activation energy, E_a, where the energy reaches a maximum, which today we call *transition state*, TS. The molecules at that maximum are nowadays known as *activated complexes*.

Arrhenius suggested that the reactants and activated complexes are in equilibrium and that the rate of the reaction is proportional to the concentration of the activated

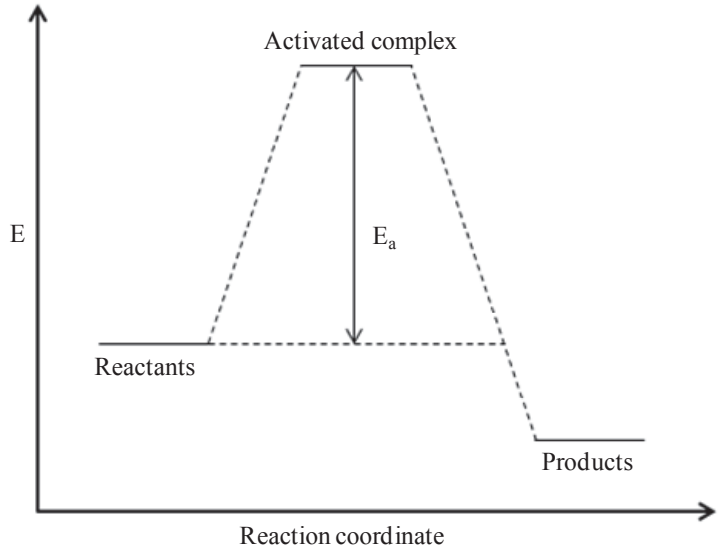


Figure 4.1: Representation of the activation energy and the activated complex.

complexes (figure 4.1). Although this is a simplified view of reactions, it captures the main concepts and it became the basis of future kinetic theories [Laidler, 1987] for studying both gas-phase and liquid-phase reactions. A convenient representation of the activation barrier is through the *potential energy surface* (PES) [Laidler, 1987], [Cramer, 2004], where the total energy of any atom arrangement encountered during a chemical reaction is represented as a multidimensional surface, with atomic positions as variables. A theoretical interpretation of equation 4.1 is given by the well known *transition state theory* [Eyring, 1935], [Evans and Polanyi, 1935] and will be discussed later in the chapter.

According to the nature of the activated complex (figure 4.1), the solvent effects may vary. Organic reactions can be categorised as *isopolar*, *dipolar* and *free-radical* transition-state reactions, based on the charge distribution of reactants and activated complex [Reichardt and Welton, 2011]. In isopolar transition state reactions, the reactants and activated complex have similar charge distribution/separation, while dipolar transition-state reactions are reactions in which the charge distribution or charge separation of the activated complex differs from that of the reactants. Free-radical activated complexes are formed by the creation of unpaired electrons during

homolytic bond cleavage. In the case of dipolar transition state reactions, solvent effects can be very significant, while the effect in isopolar and free-radical transition state reactions is small or even negligible. According to the Hughes-Ingold rules [Hughes and Ingold, 1935], [Cooper et al., 1948], [Reichardt and Welton, 2011] for aliphatic nucleophilic substitution and elimination reactions, which are typical and, when increasing the solvent polarity, the reaction rate will increase if the charge density of activated complex is greater than that of the reactants; in such a case the activated complex is stabilised in the solvent. On the other hand, when the charge density of the activated complex is lower than that of the reactants, an increase of the solvent polarity will decrease the reaction rate. In the case of similar reactants and activated complex charge densities, a change in solvent polarity will not affect the reaction rate. These observations are summarised in Tables 4.1 and 4.2, adapted from Reichardt and Welton [2011], for nucleophilic substitution and β-elimination reactions, respectively.

Reaction type	Initial reactants	Activated complex	Charge alteration during activation	Effect of increased solvent polarity on rate
S_N1	R-X	$R^{\delta+} \cdots X^{\delta-}$	separation of unlike charges	large increase
S_N1	$R\text{-}X^+$	$R^{\delta+} \cdots X^{\delta+}$	dispersal of charge	small decrease
S_N2	Y + R-X	$Y^{\delta+} \cdots R \cdots X^{\delta-}$	separation of unlike charges	large increase
S_N2	$Y^- + R\text{-}X$	$Y^{\delta-} \cdots R \cdots X^{\delta-}$	dispersal of charge	small decrease
S_N2	$Y + R\text{-}X^+$	$Y^{\delta+} \cdots R \cdots X^{\delta+}$	dispersal of charge	small decrease
S_N2	$Y^- + R\text{-}X^-$	$Y^{\delta-} \cdots R \cdots X^{\delta-}$	destruction of charge	large decrease

Table 4.1: Solvent effects on rates of nucleophilic substitution reactions

Reaction type	Initial reactants	Activated complex	Charge alteration during activation	Effect of increased solvent polarity on rate
E1	H-C-C-X	$\text{H-C-C}^{\delta+} \cdots \text{X}^{\delta-}$	separation of unlike charges	large increase
E1	H-C-C-X^{+}	$\text{H-C-C}^{\delta+} \cdots \text{X}^{\delta+}$	dispersal of charge	small decrease
E2	Y: + H-C-C-X	$\text{Y}^{\delta+} \cdots \text{H} \cdots \text{C}\therefore\text{C} \cdots \text{X}^{\delta-}$	separation of unlike charges	large increase
E2	Y:^{-} + H-C-C-X	$\text{Y}^{\delta-} \cdots \text{H} \cdots \text{C}\therefore\text{C} \cdots \text{X}^{\delta-}$	dispersal of charge	small decrease
E2	Y: + H-C-C-X^{+}	$\text{Y}^{\delta+} \cdots \text{H} \cdots \text{C}\therefore\text{C} \cdots \text{X}^{\delta+}$	dispersal of charge	small decrease
E2	Y:^{-} + R-X^{-}	$\text{Y}^{\delta-} \cdots \text{H} \cdots \text{C}\therefore\text{C} \cdots \text{X}^{\delta+}$	destruction of charge	large decrease

Table 4.2: Solvent effects on rates of β-elimination reactions

In general, the solvent affects the reaction rate constant through stabilisation of the reactants or transition structure by altering the potential energy surface of the reaction and consequently the activation energy. Therefore, a solvent that decreases the activation energy accelerates a reaction while a solvent that increases the activation barrier, decreases the reaction rate constant. Nevertheless, in engineering applications one is interested in the rate of the reaction and not only the rate constant. The rate of a reaction depends also on the concentration of the reactants in the solvent (rate = rate constant $\times$ concentration). Thus, when considering solvent effects on a reaction, ideally the effect on rate constant $\times$ solubility should be taken into account, however, in this thesis the focus is only on the rate constant.

The methods that have been developed to capture the solvent effects on reactions and predict rate constants of liquid-phase reactions are either empirical, the so-called *linear free-energy relationships*, or based on theoretical chemistry. The basic concepts of linear free-energy relationships and those most widely-used are discussed in section 4.2, while an overview of the most commonly applied kinetic theory [Santiso and Gubbins, 2004], the transition state theory (TST), is given in section 4.3. Expressions for the rate constant in the gas and the liquid phase are derived using TST in sections 4.4 and 4.5. Finally, in section 4.5.3 an overview of available solvation models needed in the TST calculations for liquid phase rate constants is given.

4.2 Linear free-energy relationships

Linear free-energy relationships (LFER) or *linear Gibbs energy relationships* are linear correlations between the logarithm of the rate constant or equilibrium constant for one series of reactions and the logarithm of the rate constant or equilibrium constant for a related series of reactions. The use of "free energy" in the name of these models results from the fact that equilibrium constants and rate constants are related to the Gibbs free energy of reaction, or of activation, or to changes in spectral frequencies, which are proportional to energy changes.

The first attempt to develop a linear free-energy relationship for reactions was

that by Hammett [1937], who described the effect of a substituent position in the benzene ring upon the rate or equilibrium of a reaction, by studying various reactions involving *meta*- and *para*-substituted benzene derivatives. He introduced a relation now known as the *Hammett equation*

$$\log K = \log K^0 + \sigma\rho, \tag{4.2}$$

where K refers to the reaction rate constant or equilibrium constant for a substituted reactant and K^0 is the corresponding property for the unsubstituted reactant, σ is a substituent coefficient and ρ is a reaction coefficient.

The difficulty with theoretical approaches that try to capture solvent effects on reactions, is that the solvent physical properties (polarity, polarizability) or chemical properties (acidity, basicity) are not independent and, as a result, they do not act alone. Considering, for example, the definition of polarity that Reichardt gave in 1965, "the polarity of a solvent is determined by its solvation capability for reactants and activated complexes as well as for molecules in their ground and excited states, which (capability) depends on the action of all possible, specific and non-specific, intermolecular forces between solvent and solute molecules", it becomes obvious that solvent polarity cannot be measured by an individual physical quantity. Therefore, simple, single-term linear Gibbs energy relationships such as the Hammett equation are not always sufficient to fit sets of experimental series of reactions. Many attempts have been made to express the parallel effects of solvent properties and these attempts usually take the form

$$A = A_0 + bB + cC + dD + \cdots, \tag{4.3}$$

where A is a solvent-dependent physicochemical property such as rate constant or equilibrium constant in a given solvent, A_0 is the value of this property in the gas phase or a reference solvent, B, C, D are solvent properties and b, c, d are regressed coefficients that describe the sensitivity of the corresponding solvent property to property A. A widely known and widely applied equation of this type is the solvatochromic equation presented in Kamlet and Taft [1976] and Taft and Kamlet [1976], and which was described in section 3.4.1. Further equations have been developed by apply-

ing statistics to treat multivariate chemical data; this type of approach is adopted in *chemometrics* [Buncel et al., 2003]. The most popular chemometrics method is *principal component analysis* (PCA) [Eliasson et al., 1982], where the various solvent properties are expressed as linear combinations of a set of new independent (or orthogonal) variables. The original variables P_j can be written as a linear function of the orthogonal set, such as

$$P_j = \sum_r a_{jr} Z_r, \tag{4.4}$$

where Z_r are the new variables or principal components, whose number is equal to the number of properties in the original set. The main feature of PCA is that the first principal component Z_1 covers the largest part of the variation of the original data, Z_2 covers the next largest, and so on. It should be mentioned that PCA considers a larger number of solvent properties (e.g. 17 properties in Stairs and Buncel [2006]) than the solvatochromic equation (e.g. 5 properties in Struebing et al. [2013]), naturally with the cost of increased complexity.

Applications of empirical methods for a wide range of reactions can be found in the literature. For example, the solvatochromic equation has been shown to correlate successfully the rate constant of numerous reactions, including the solvolysis of t-butyl chloride [Abraham et al., 1987], Diels-Alder reactions [Cativiela et al., 1997], reactions of 2-substituted cyclohex-1-enylcarboxylic and 2-substituted benzoic acids [Nikolić et al., 2007], Menschutkin reactions [Folić et al., 2007, 2008a, Struebing et al., 2013], ring-closing metathesis reactions [Adjiman et al., 2008], and the reaction of 1-fluoro-2,4-dinitrobenzene with anilines [Jamali-Paghaleh et al., 2011]. Principal component analysis is also popular. It has recently been applied and provided useful insight on kinetic studies to ten different reactions (including S_NAr, S_EAr, addition, cycloaddition, thermolysis, Diels-Alder reactions) by Stairs and Buncel [2006], coal pyrolysis [Khare et al., 2011] and the reaction of amlodipine and 1,2-naphthoquinone-4-sulfonate [Shariati-Rada et al., 2013]. It is notable that when the required parameters are or can be available, these simple expressions are often comparable in accuracy with the computationally intensive theoretical methods [Jalan et al., 2010]. However, the applicability of empirical methods is generally limited to reactions for which data

are available. In cases where no data are available or *ab initio* design is desirable, predictive models are required, like transition-state theory that is discussed in the next section.

4.3 Conventional transition state theory

Conventional transition state theory (CTST) is a theory based on the proposed view of the reaction mechanism proposed by Arrhenius, that allows the prediction of the rate of a reaction, using statistical mechanical information; it that was published almost simultaneously by Eyring [1935] and by Evans and Polanyi [1935].

Consider a bimolecular reaction of the type

$$\nu_{\mathrm{A}}\mathrm{A} + \nu_{\mathrm{B}}\mathrm{B} \rightarrow \text{products}, \tag{4.5}$$

where A and B are the reactants and ν_i is the stoichiometric coefficient of species i. In transition state theory an activated complex $(\mathrm{AB})^{\ddagger}$ is considered, as a population of molecules in equilibrium with one another and also in equilibrium with the reactants. The activated complex population either reacts irreversibly to form the products, or deactivates back to the reactants. Thus, reaction (4.5) is rewritten as

$$\nu_{\mathrm{A}}\mathrm{A} + \nu_{\mathrm{B}}\mathrm{B} \overset{K_C^{\ddagger}}{\rightleftharpoons} \nu_{(\mathrm{AB})^{\ddagger}}\,(\mathrm{AB})^{\ddagger} \overset{k}{\rightarrow} \text{products}, \tag{4.6}$$

where the reactants A and B are in a rate-determining quasi-equilibrium with the activated complex $(\mathrm{AB})^{\ddagger}$, $K_C^{\ddagger}$ is the quasi-equilibrium constant between the activated complex and the reactants, and k is the reaction rate constant. In the space of the reaction coordinate and the potential energy surface (PES), the transition state corresponds to a first-order saddle point on the PES [Jensen, 2007]; a maximum in the reaction coordinate direction and a minimum in all other directions.

The key further assumption of conventional transition-state theory (CTST) is that that no multiple crossings occur in the potential energy surface, i.e., all the molecules that pass through the transition state transform into products. This means that the

rate predicted from CTST is the maximum value that the rate can take and that it is thus most probably an overestimation of the real reaction rate. To overcome this limitation a transmission coefficient κ is usually introduced to account for any "re-crossings". The transmission coefficient also allows for the phenomenon of tunnelling in which molecules that do not have sufficient energy to overcome the activation barrier may however "tunnel" through the barrier and transform into products.

Another commonly used form of transition state theory is variational transition state theory (VTST) [Keck, 1960], [Garrett and Truhlar, 1979a], [Garrett and Truhlar, 1979b]. The basic concept of VTST is that instead of considering the transition state as the maximum of a potential-energy surface, as is done in CTST, various dividing surfaces along the reaction path are considered. The TS of the surface that gives the lowest rate is chosen, acknowledging the fact that in CTST the rate constant is usually overestimated because of the re-crossings that are not taken into account. For more details about VTST the reader is referred to Laidler [1987] and Cramer [2004].

4.4 Gas phase reaction rate constants from CTST

In this section, an expression for the rate constant of a bimolecular reaction in the ideal gas phase is derived applying CTST, following the derivation of Struebing [2011].

According to CTST a bimolecular reaction can take the following form

$$\nu_{\mathrm{A}}\mathrm{A} + \nu_{\mathrm{B}}\mathrm{B} \overset{K_C^{\ddagger,IG}}{\rightleftharpoons} \nu_{(\mathrm{AB})^{\ddagger}}\,(\mathrm{AB})^{\ddagger} \overset{k^{IG}}{\rightarrow} \text{products}, \tag{4.7}$$

where $(\mathrm{AB})^{\ddagger}$ is the activated complex or the transition state structure (TS), $\nu_{(\mathrm{AB})^{\ddagger}}$ is its associated stoichiometric coefficient, $K_C^{\ddagger,IG}$ is the concentration-based quasi-equilibrium constant and k^{IG} is the ideal gas-phase reaction rate constant.

The ideal gas reaction rate constant k^{IG} is related to the quasi-equilibrium constant $K_C^{\ddagger,IG}$ between the reactants and TS structure [Wynne-Jones and Eyring, 1935], [Laidler, 1987] by

$$k^{IG} = \kappa \frac{k_B T}{h} K_C^{\ddagger,IG}, \tag{4.8}$$

where κ is the transmission coefficient presented earlier and h is Plank's constant in

J s. The quasi-equilibrium constant is obtained as,

$$K_C^{\ddagger,IG} = \prod_i \left(c_i^{IG}\right)^{\nu_i}, \tag{4.9}$$

where c_i is the concentration of species i in mol dm^{-3} , where i = A,B,TS.

4.4.1 $K_C^{\ddagger,IG}$ in terms of $K_A^{\ddagger,IG}$

The activity-based equilibrium constant, $K_A^{\ddagger,IG}$, is defined as the product of the reaction species activities to the power of their respective stoichiometry coefficients

$$K_A^{\ddagger,IG} = \prod_i \left(a_i^{IG}\right)^{\nu_i}, \tag{4.10}$$

where a_i^{IG} is the activity for species i. Considering a system at specified pressure P, temperature T and concentration x_i for species i, a_i^{IG} can be expressed as [Sandler, 1999]

$$a_i^{IG} = \frac{p_i^{IG}}{p^\circ}, \tag{4.11}$$

where p_i^{IG} is the partial pressure of species i and p° is the standard state pressure (1 atm). By substituting Equation (4.11) into Equation (4.10) the following expression is obtained

$$K_A^{\ddagger,IG} = \prod_i \left(\frac{p_i^{IG}}{p^\circ}\right)^{\nu_i}. \tag{4.12}$$

Using the ideal gas equation of state

$$p_i^{IG} V = N_i RT \tag{4.13}$$

to relate pressure to concentration, $c_i = N_i/V$, equation (4.12) becomes

$$K_A^{\ddagger,IG} = \prod_i \left(\frac{c_i^{IG} RT}{p^\circ}\right)^{\nu_i} \tag{4.14}$$

$$= K_C^{\ddagger,IG} \prod_i \left(\frac{RT}{p^\circ}\right)^{\nu_i}. \tag{4.15}$$

Rearranging equation (4.15), $K_C^{\ddagger,IG}$ is given as a function of $K_A^{\ddagger,IG}$ by

$$K_C^{\ddagger,IG} = K_A^{\ddagger,IG} \prod_i \left(\frac{RT}{p^\circ}\right)^{-\nu_i}. \tag{4.16}$$

4.4.2 $K_A^{\ddagger,IG}$ in terms of the free energy of reaction

For single-phase chemical equilibrium problems, the partial molar Gibbs free energy of a molecular species i at P, T and x_i (the chemical potential of species i) can be expressed in terms of the standard state of each species [Sandler, 1999] as follows

$$\begin{aligned}\underline{G}_i^{IG} &= \underline{G}_i^{\circ,IG}(p^\circ,T) + \left[\underline{G}_i^{IG} - \underline{G}_i^{\circ,IG}(p^\circ,T)\right] \\ &= \underline{G}_i^{\circ,IG}(p^\circ,T) + RT\ln\left[\frac{f_i^{IG}}{f_i^{\circ,IG}(p^\circ,T)}\right] \\ &= \underline{G}_i^{\circ,IG}(p^\circ,T) + RT\ln a_i^{IG}, \end{aligned} \tag{4.17}$$

where f_i^{IG} and $f_i^{\circ,IG}$ are the fugacity and standard state fugacity, respectively, of species i. From the equilibrium condition at P and T

$$\sum_i \nu_i \underline{G}_i^{IG} = 0, \tag{4.18}$$

it follows that

$$\begin{aligned}\sum_i \nu_i \underline{G}_i^{IG} &= \sum_i \nu_i \underline{G}_i^{\circ,IG}(p^\circ,T) + RT\sum_i \nu_i \ln a_i^{IG} \\ &= \Delta^\ddagger \underline{G}^{\circ,IG} + RT\sum_i \nu_i \ln a_i^{IG}, \end{aligned} \tag{4.19}$$

and at equilibrium,

$$-\frac{\Delta^\ddagger \underline{G}^{\circ,IG}}{RT} = \ln\left(\prod_i \left(a_i^{IG}\right)^{\nu_i}\right), \tag{4.20}$$

where $\Delta^\ddagger \underline{G}^{\circ,IG}$ is the standard state activation Gibbs free energy. Rearranging equation (4.20) and combining with equation (4.10), the activity-based equilibrium constant is given by

$$K_A^{\ddagger,IG} = \exp\left(-\frac{\Delta^\ddagger \underline{G}^{\circ,IG}}{RT}\right). \tag{4.21}$$

In the remainder of this section, an expression for $K_A^{\ddagger,IG}$ is derived using statistical mechanics. According to statistical mechanics, in the canonical ensemble (N,V,T) the Helmholtz free energy, A, is related to the partition function $Q(N,V,T)$ as [McQuarrie, 2000]

$$A = -k_B T \ln Q(N,V,T). \tag{4.22}$$

The definition of the chemical potential μ_i of species i is [Sandler, 1999]

$$\mu_i(N,V,T) = \left(\frac{\partial A}{\partial N_i}\right)_{N_{j\neq i},V,T} \tag{4.23}$$

and from equation (4.22) it follows that

$$\mu_i(N,V,T) = -k_B T \left(\frac{\partial Q(N,V,T)}{\partial N_i}\right)_{N_{j\neq i},V,T}. \tag{4.24}$$

The canonical partition function of an ideal gas can be expressed as [McQuarrie, 2000]

$$Q^{IG}(N,V,T) = \prod_i \frac{\left(q_i^{IG}(V,T)\right)^{N_i}}{N_i!}, \tag{4.25}$$

where q_i^{IG} is the dimensionless molecular partition function for species i, N is the total number of particles (i.e. $N = \sum_i N_i$). When considering a mixture of ideal gases, the species are independent and distinguishable [McQuarrie, 2000] and, thus, the chemical potential of each species i, using also Striling's approximation, is given by the following equation [McQuarrie, 2000]

$$\mu_i^{IG}(N,V,T) = -k_B T \ln \frac{q_i^{IG}(V,T)}{N_i}. \tag{4.26}$$

The system is in equilibrium and equation (4.18) can be written also in terms of chemical potential

$$\sum_i \nu_i \mu_i^{IG} = 0. \tag{4.27}$$

Thus, by substituting equation (4.26) into equation (4.27) it results that

$$\prod_i \left(q_i^{IG}(V,T)\right)^{\nu_i} = \prod_i (N_i)^{\nu_i}. \tag{4.28}$$

When substituting the ideal gas law expression

$$p_i^{IG} V = N_i k_B T \tag{4.29}$$

for p_i^{IG} in equation (4.12) the following expression is obtained

$$K_A^{\ddagger,IG} = \prod_i \left(\frac{p_i^{IG}}{p^\circ}\right)^{\nu_i} = \prod_i \left(\frac{N_i k_B T}{V p^\circ}\right)^{\nu_i} \tag{4.30}$$

$$= \prod_i (N_i)^{\nu_i} \prod_i \left(\frac{k_B T}{V p^\circ}\right)^{\nu_i}. \tag{4.31}$$

Substituting equation (4.28) in (4.31), it follows that

$$K_A^{\ddagger,IG} = \prod_i \left(q_i^{IG}(V,T)\right)^{\nu_i} \prod_i \left(\frac{k_B T}{V p^\circ}\right)^{\nu_i} \tag{4.32}$$

$$= \prod_i \left(\frac{q_i^{IG}(V,T) k_B T}{V p^\circ}\right)^{\nu_i}. \tag{4.33}$$

The molecular partition function for species i, q_i^{IG}, consists of four contributions; translational, rotational, vibrational and electronic

$$q_i^{IG}(V,T) = q_{t,i}^{IG}(V,T)\, q_{r,i}^{IG}(T)\, q_{v,i}^{IG}(T)\, q_{e,i}^{IG}(T), \tag{4.34}$$

where $q_{t,i}^{IG}$ is the translational molecular partition function, $q_{r,i}^{IG}$ is the rotational molecular partition function, $q_{v,i}^{IG}$ is the vibrational molecular partition function and $q_{e,i}^{IG}$ is the electronic molecular partition function. The derivations of all terms are well documented (e.g. Hill [1986], McQuarrie [2000]) and expressions for each term are reported here for completeness.

The translational molecular partition function is given by

$$q_{t,i}^{IG} = \left(\frac{2\pi \left(\sum_k m_{k,i}\right) k_B T}{h^2}\right)^{\frac{3}{2}} V, \tag{4.35}$$

where $m_{k,i}$ is the mass of each atom k in molecule i in atomic mass units. For linear molecules, the rotational molecular partition function is obtained by

$$q_{r,i}^{IG} = \frac{8\pi^2 I_i k_B T}{\sigma_i h^2}, \tag{4.36}$$

where σ_i is a symmetry number and I_i is the moment of inertia in kg m^2. For non-linear molecules, the rotational molecular partition function is

$$q_{r,i}^{IG} = \frac{\pi^{\frac{1}{2}}}{\sigma_i} \left(\frac{8\pi^2 I_{A,i} k_B T}{h^2}\right)^{\frac{1}{2}} \left(\frac{8\pi^2 I_{B,i} k_B T}{h^2}\right)^{\frac{1}{2}} \left(\frac{8\pi^2 I_{C,i} k_B T}{h^2}\right)^{\frac{1}{2}}, \tag{4.37}$$

where σ_i is a symmetry number and $I_{A,i}$, $I_{B,i}$ and $I_{C,i}$ are principal moments of inertia in kg m^2. The vibrational molecular partition function is calculated as follows

$$q_{v,i}^{IG} = \prod_{j=1}^{\alpha_i} \frac{\exp\left(-\frac{\Theta_{i,j}}{2T}\right)}{1 - \exp\left(-\frac{\Theta_{i,j}}{T}\right)}, \tag{4.38}$$

where $\alpha_i = 3N - 6$ for non-linear molecules i, $\alpha_i = 3N - 5$ for linear molecules i and $\Theta_{i,j}$ is a characteristic vibrational temperature for species i in K and vibrational mode j and is defined as

$$\Theta_{i,j} = \frac{h\nu_{i,j}}{k_B}, \tag{4.39}$$

where $\nu_{i,j}$ is the vibrational frequency of species i and vibrational mode j in s^{-1}. The electronic molecular partition function is calculated from the following expression

$$q_{e,i}^{IG} = \omega_{e1,i} \exp\left(-\frac{\epsilon_{e1,i}^{IG}}{k_B T}\right), \tag{4.40}$$

where $\omega_{e1,i}$ is the degeneracy, i.e. the number of electronic energy levels in the ground state, which for ordinary chemically saturated molecules is equal to 1 [Hill, 1986], and $\epsilon_{e1,i}^{IG}$ is the potential energy in J particle^{-1}, with a reference of zero energy taken when all atoms in the molecule are separated in their ground electronic states. At the minimum of the potential energy surface, $\epsilon_{e1,i}^{IG}$ corresponds to the ideal gas electronic energy $E_i^{el,IG}$ which can be calculated using quantum mechanical (QM) software packages. The electronic molecular partition function becomes

$$q_{e,i}^{IG} = \omega_{e1,i} \exp\left(-\frac{E_i^{ele,IG}}{k_B T}\right). \tag{4.41}$$

Finally, the total molecular partition function can be written as a function of the electronic energy

$$q_i^{IG} = q_i^{',IG}(V,T) \exp\left(-\frac{E_i^{ele,IG}}{k_B T}\right), \tag{4.42}$$

where $q_i^{',IG}$ is defined for convenience to include all components of the partition function except the potential energy contribution and is given by

$$q_i^{',IG}(V,T) = q_{t,i}^{IG}(V,T)\, q_{r,i}^{IG}(T)\, q_{v,i}^{IG}(T). \tag{4.43}$$

4.4.3 Reaction rate constant for the ideal gas phase

The activity-based quasi-equilibrium constant from equations (4.33) and (4.42) becomes

$$K_A^{\ddagger,IG} = \prod_i \left(\frac{q_i^{',IG}(V,T)\, k_B T}{V p^{o,IG}}\right)^{\nu_i} \exp\left(-\frac{\Delta^{\ddagger} E^{ele,IG}}{k_B T}\right), \tag{4.44}$$

where $\Delta^{\ddagger}$ is the difference of the stoichiometric coefficients of the species, that is, $\Delta^{\ddagger} = \nu_{TS} - \nu_A - \nu_B$. In order to make the molecular partition function a function of temperature only, the translational partition function is written as

$$q_i^{',IG}(V,T) = \tilde{q}_i(T) \cdot V \tag{4.45}$$

where $\tilde{q}_i(T)$ is the temperature dependent partition function. Now the activity-based quasi-equilibrium constant becomes

$$K_A^{\ddagger,IG} = \prod_i \left(\tilde{q}_i(T) \cdot \frac{k_B T}{p^{\circ}} \right)^{\nu_i} \exp\left[-\frac{\Delta^{\ddagger} E^{ele,IG}}{k_B T} \right] \tag{4.46}$$

$$= \prod_i \left(q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T \right) \right)^{\nu_i} \exp\left[-\frac{\Delta^{\ddagger} E^{ele,IG}}{k_B T} \right] \tag{4.47}$$

where

$$q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T \right) = \tilde{q}_i(T) \cdot \frac{k_B T}{p^{\circ}}. \tag{4.48}$$

QM software packages (e.g. Gaussian 03 [Frisch et al., 2004] and Gaussian 09 [Frisch et al., 2009], GAMESSPLUS [Higashi et al., 2008]) can be used to calculate $q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T\right)$ for $p^{\circ} = 1$ atm. The concentration-based quasi-equilibrium constant can now be expressed by incorporating equation (4.47) into equation (4.16)

$$K_C^{\ddagger,IG} = \prod_i \left(\frac{RT}{p^{\circ}} \right)^{-\nu_i} \prod_i \left(q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T \right) \right)^{\nu_i} \exp\left[-\frac{\Delta^{\ddagger} E^{ele,IG}}{k_B T} \right] \tag{4.49}$$

$$= \prod_i \left(\frac{RT}{p^{\circ}} \right)^{-\nu_i} \prod_i \left(q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T \right) \right)^{\nu_i} \exp\left[-\frac{\Delta^{\ddagger} \underline{E}^{ele,IG}}{RT} \right], \tag{4.50}$$

where $E^{ele,IG}$ is the electronic energy per particle and $\underline{E}^{ele,IG}$ is the molar electronic energy. The reaction rate constant in the ideal gas phase k^{IG} is finally given in terms of known quantities in $\mathrm{dm^3\ mol^{-1}\ s^{-1}}$, by substituting equation (4.50) into equation (4.8)

$$k^{IG} = \kappa \frac{k_B T}{h} \prod_i \left(\frac{RT}{p^{\circ}} \right)^{-\nu_i} \prod_i \left(q_i^{',IG}\left(\frac{k_B T}{p^{\circ}}, T \right) \right)^{\nu_i} \exp\left[-\frac{\Delta^{\ddagger} \underline{E}^{ele,IG}}{RT} \right]. \tag{4.51}$$

4.5 Liquid phase reaction rate constants from CTST

In this section, an expression for the liquid phase reaction rate constant of a bimolecular reaction is presented applying conventional transition state theory; here

the derivation of Struebing [2011] is followed.

Similarly to the gas phase, a bimolecular reaction in the liquid phase can be expressed according to CTST as

$$\nu_{\mathrm{A}}\mathrm{A} + \nu_{\mathrm{B}}\mathrm{B} \overset{K_C^{\ddagger,L}}{\rightleftharpoons} \nu_{(\mathrm{AB})^{\ddagger}}\,(\mathrm{AB})^{\ddagger} \overset{k^L}{\rightarrow} \text{products}, \tag{4.52}$$

where k^L is the liquid phase reaction rate constant in $\mathrm{dm}^3\ \mathrm{mol}^{-1}\ \mathrm{s}^{-1}$, $K_C^{\ddagger,L}$ is the concentration-based quasi-equilibrium constant in $\mathrm{dm}^3\ \mathrm{mol}^{-1}$ and ν_A, ν_B, $\nu_{(\mathrm{AB})^{\ddagger}}$ are the stoichiometric coefficients for species A, B and the activated complex (or TS), $(\mathrm{AB})^{\ddagger}$. k^L is then given by

$$k^L = \kappa \frac{k_B T}{h} K_C^{\ddagger,L}. \tag{4.53}$$

In the rest of the section, $K_C^{\ddagger,L}$ is derived in terms of known quantities.

4.5.1 $K_C^{\ddagger,L}$ in terms of $K_A^{\ddagger,L}$

Similarly to the gas phase, the activity-based quasi-equilibrium constant can be expressed as a function of the activities of the reactions species or as a function of the activation free energy

$$K_A^{\ddagger,L} = \prod_i \left(a_i^L\right)^{\nu_i} \tag{4.54}$$

$$= \exp\left(-\frac{\Delta^{\ddagger}\underline{G}^{\circ,L}}{RT}\right), \tag{4.55}$$

where $K_A^{\ddagger,L}$ is the activity-based quasi-equilibrium coefficient and $\Delta^{\ddagger}\underline{G}^{\circ,L}$ is the activation free energy required by the reaction and i refers to species A, B, and TS. Only low to medium pressures are considered, thus the Poynting correction factor is assumed to be equal to unity [Sandler, 1999] and the activity for species i is defined [International Union of Pure and Applied Chemistry, 1993] as

$$a_i^L = \gamma_i^* \frac{m_i^L}{m_i^{\circ,L}}, \tag{4.56}$$

where m_i^L is the molality of species i and is defined as the moles of species i in 1 kg of solvent, $m_i^{\circ,L}$ is the standard state molality of species i also in mol (kg of solvent)$^{-1}$

and γ_i^* is the dimensionless activity coefficient for species i. Infinite dilution is assumed and therefore γ_i^* tends to unity and it can be assumed that the density of the solution is equal to the density of the solvent. It then follows that

$$m_i^L \cdot \rho = c_i^L \tag{4.57}$$

and

$$m_i^{\circ,L} \cdot \rho = c_i^{\circ,L}, \tag{4.58}$$

where ρ is the solvent density in kg dm^{-3}, c_i^L is the molar concentration for species i in mol dm^{-3} and $c_i^{\circ,L}$ is the standard state molar concentration of species i in mol dm^{-3}. By substituting equations (4.56) to (4.58) into equation (4.54), and setting $\gamma_i^* = 1$, $K_A^{\ddagger,L}$ can be expressed as follows

$$K_A^{\ddagger,L} = \prod_i \left(\frac{c_i^L}{c_i^{\circ,L}} \right)^{\nu_i} \tag{4.59}$$

$$= K_C^{\ddagger,L} \prod_i \left(c_i^{\circ,L} \right)^{-\nu_i}, \tag{4.60}$$

where

$$K_C^{\ddagger,L} = \prod_i \left(c_i^L \right)^{\nu_i}. \tag{4.61}$$

Finally, by substituting $K_A^{\ddagger,L}$ from equation (4.55) and solving for $K_C^{\ddagger,L}$, the following expression for the concentration-based quasi-equilibrium constant is obtained

$$K_C^{\ddagger,L} = \prod_i \left(c_i^{\circ,L} \right)^{\nu_i} \exp\left(-\frac{\Delta^{\ddagger} \underline{G}^{\circ,L}}{RT} \right). \tag{4.62}$$

4.5.2 Reaction rate constant for the liquid phase

The partial molar free energy in the liquid phase for species i, $\underline{G}_i^{\circ,L}$, can be partitioned into three components [Ho et al., 2010] as follows

$$\underline{G}_i^{\circ,L} = \underline{G}_i^{\circ,IG} + \Delta \underline{G}_i^{p \to c,IG} + \Delta \underline{G}_i^{\circ,solv}, \tag{4.63}$$

where $\underline{G}_i^{\circ,L}$ is the partial molar Gibbs free energy of component i in the liquid phase at temperature $T = 298.15$ K and concentration $c_i^{o,L} = 1$ mol dm^{-3}, and $\underline{G}_i^{\circ,IG}$ as above, is the partial molar free energy of component i in an ideal gas at temperature

$T = 298.15K$ and pressure $p^{o,IG} = 1$ atm. The $\Delta\underline{G}_i^{p\to c,IG}$ term accounts for the conversion from the gas phase standard state at temperature $T = 298.15K$ and pressure $p^{o,IG} = 1$ atm to temperature $T = 298.15$ K and concentration $c_i^{o,IG} = 1$ mol dm^{-3} and it is given by [Ho et al., 2010]

$$\Delta\underline{G}_i^{p\to c,IG} = RT\ln\left(\frac{RT}{p^{o,IG}}\right) \tag{4.64}$$

$$\approx 1.89 \text{ kcal mol}^{-1}, \tag{4.65}$$

where $p^\circ = 101,325$ Pa. Finally, $\Delta\underline{G}_i^{\circ,solv}$ represents the energy of transfer 1 mole of species i from the ideal gas phase at temperature $T = 298.15$ K and concentration $c_i^{o,IG} = 1$ mol dm^{-3} to the liquid (solvent) phase at temperature $T = 298.15$ K and concentration $c_i^{o,L} = 1$ mol dm^{-3}.

For a bimolecular reaction and according to equation (4.63), the molar activation free energy in the liquid phase is

$$\Delta^\ddagger\underline{G}^{\circ,L} = \Delta^\ddagger\underline{G}^{\circ,IG} + \Delta^\ddagger\Delta\underline{G}^{p\to c,IG} + \Delta^\ddagger\Delta\underline{G}^{\circ,solv} \tag{4.66}$$

and combining equations (4.53), (4.62) and (4.66), the liquid phase reaction rate constant is obtained as:

$$\begin{aligned} k^L =& \kappa\frac{k_BT}{h}\prod_i\left(c_i^{\circ,L}\right)^{\nu_i}\exp\left[-\frac{\Delta^\ddagger\underline{G}^{\circ,IG}}{RT}\right] \\ &\cdot\exp\left[-\frac{\Delta^\ddagger\Delta\underline{G}^{p\to c,IG} + \Delta^\ddagger\Delta\underline{G}^{\circ,solv}}{RT}\right]. \end{aligned} \tag{4.67}$$

The exponential term for the ideal gas term, containing $\Delta^\ddagger\underline{G}^{\circ,IG}$ can be expressed as in equation (4.47)

$$K_A^{\ddagger,IG} = \exp\left(-\frac{\Delta^\ddagger\underline{G}^{\circ,IG}}{RT}\right) = \prod_i\left(q_i^{',IG}\right)^{\nu_i}\exp\left[-\frac{\Delta^\ddagger E^{ele,IG}}{k_BT}\right] \tag{4.68}$$

$$= \prod_i\left(q_i^{',IG}\right)^{\nu_i}\exp\left[-\frac{\Delta^\ddagger\underline{E}^{ele,IG}}{RT}\right]. \tag{4.69}$$

By substituting equation (4.69) in (4.67), a general expression for the liquid phase

rate constant is obtained

$$k^L = \kappa \frac{k_B T}{h} \prod_i \left(c_i^{\circ,L}\right)^{\nu_i} \prod_i \left(q_i^{',IG}\right)^{\nu_i} \exp\left[-\frac{\Delta^\ddagger \underline{E}^{ele,IG}}{RT}\right] \cdot \exp\left[-\frac{\Delta^\ddagger \Delta \underline{G}^{p\to c,IG} + \Delta^\ddagger \Delta \underline{G}^{\circ,solv}}{RT}\right]. \quad (4.70)$$

The liquid free energy of solvation, $\Delta \underline{G}^{\circ,solv}$, can be calculated using a solvation model. The basis of solvation models and some widely used models are reviewed in the next section.

4.5.3 Solvation models

Although gas-phase free energy predictions are nowadays impressively accurate, the prediction of liquid-phase free energies and thus liquid-phase kinetics still remains a challenge [Jalan et al., 2010], as it is hard to capture all solvent effects. Different a priori methods have been proposed for the calculation of the solvation free energy within a quantum mechanical framework; they differ mainly in the representation of the solvent molecules around the solute: from a large number of solvent molecules to an infinite dielectric continuum (figure 4.2).

In figure 4.2 the different approaches that have been considered in order to represent the solvent in developing solvation models are summarised. There are methods that treat the solvent explicitly (a,b,c), where the solute is surrounded by a large number of solvent molecules. In these cases, the solute and solvent can be treated with the same model (a,b) or with different models (c), where the solvent molecules are represented via a classical, molecular mechanics force field, while a quantum view of the solutes is adopted. Another category is that consisting of methods that treat the solvent implicitly (f); here the solvent is considered to be a continuum described by a small number of solvent properties and only the solute is treated with an explicit representation of its molecular and electronic structure. Finally, there are methods that lie in between these two extremes where only a limited number of solvent molecules around the solute are treated explicitly either with molecular mechanics (MM) (d) or with quantum mechanics (QM) (e) and the rest are treated implicitly.

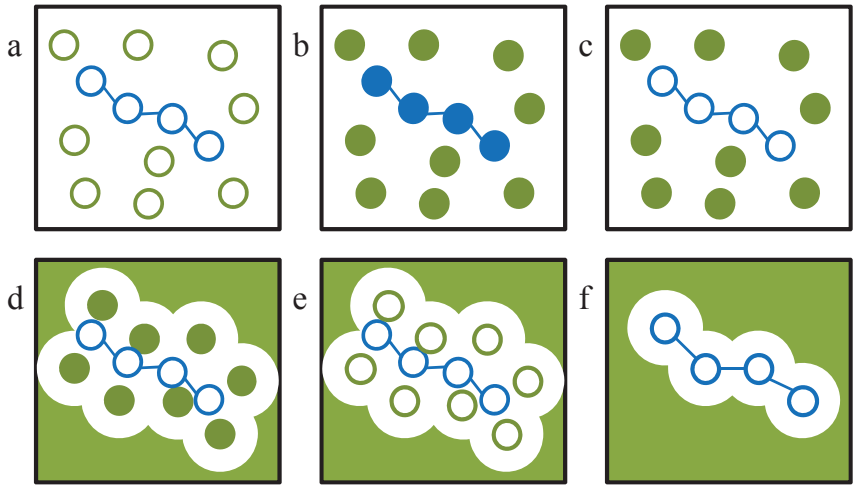


Figure 4.2: Representation of the solvent from different solvation model categories; from discrete to continuum models. White dots represent molecules or atoms calculated by quantum mechanics, coloured dots by force field. Green represents a dielectric continuum. Adapted from Jalan et al. [2010].

Conceptually, the solvation process can be divided into three steps [Orozco and Luque, 2000]: cavitation, dispersion-repulsion and electrostatics (equation (4.71)).

$$\Delta G^{solv} = \Delta G^{cav} + \Delta G^{vw} + \Delta G^{ele}. \tag{4.71}$$

Firstly, a cavity is formed within the solvent to accommodate the solute. The energy corresponding to the formation of this cavity is known as energy of cavitation, denoted by ΔG^{cav}. Once the solute is in the cavity, dispersion-repulsion (also referred to as van der Waals, ΔG^{vw}) forces act between the solute and the solvent molecules. These two contributions, ΔG^{cav} and ΔG^{vw}, are usually referred to as the non-electrostatic contribution. As the last step, a charge distribution of the solute in the solvent is created and this is the electrostatic contribution ΔG^{ele} to the solvation free energy. This term includes two components: the work required to transfer the solute's charge distribution from the gas phase to the liquid phase, and the work required for the polarization of the solute's charge distribution by the solvent.

Explicit solvation models

Explicit solvation models consider the solute surrounded by the solvent molecules, which are modelled discretely. There is an infinite number of ways in which solvent molecules can be arranged around the solute and, in order to sample over all possible conformations and obtain an average value of the solvation free energy and other properties, Monte Carlo or Molecular Dynamics algorithms are used.

The molecular interactions in this kind of systems can be represented at three different levels: pure quantum mechanics (QM) (figure 4.2a), pure classical mechanics (MM) (figure 4.2b), and mixed quantum and classical mechanics (QM/MM) (figure 4.2c), where the solute is treated with QM but the solvent modelled with a MM force field. QM methods treat both the solute and the solvent at the quantum mechanical level. A widely-used QM method is the Car-Parrinello method [Car and Parrinello, 1985]. These methods may be very accurate, but they are usually too computationally expensive due to the large number of MC or MD moves required for a sufficient sampling of the solvent space. On the other hand, it is more common to use pure classical MM methods, where parametrised force fields are used to describe molecular interactions. The classical MM treatment is very useful for providing, not only the free energy of solvation, but also structural and dynamic information which is difficult to obtain from other techniques [Orozco and Luque, 2000].

In cases like chemical reactions where changes in the electronic state take place, e.g. bond-breaking, bond-forming, charge transfer, electronic excitation, MM force fields are unable to provide good predictions and QM calculations are necessary. In order to capture as many solvent effects as possible while limiting the computational cost, QM/MM methods, that combine QM and MM calculations, have been proposed [Lin and Truhlar, 2007, Senn and Thiel, 2009]. A QM/MM method treats a localised region with QM and the surroundings with MM. QM/MM methods are widely used when studying biological systems [Orozco and Luque, 2000], [Senn and Thiel, 2009]. It is an especially relevant time for these methods. The 2013 Nobel prize in Chemistry was given to three scientists that worked on computational studies of

biological systems using mainly QM/MM methods (e.g. Cui et al. [2002], Kamerlin et al. [2009]).

One of the major limitations of explicit solvation models is the lack of classical force field potentials that are able to reproduce satisfactorily the solvent behaviour, such as long-range dielectric effects. Other limitations include the parametrisation of the force field potentials and, naturally, the challenge of sampling the configurational space of all possible solvent molecule positions. Discrete solvation models provide a very detailed description of the system and their predictions are sometimes in better agreement with experimental data compared to continuum solvation models [Jalan et al., 2010], [Acevedo and Jorgensen, 2010]. Nevertheless, it must also be noted that although MM or QM/MM calculations are faster than pure QM calculations, they are still significantly more computationally expensive when compared to continuum solvation methods or empirical methods.

Implicit solvation models

Continuum solvation models treat the solute at the QM level, whereas the solvent is considered as a uniform polarisable medium. The solute is placed in a suitably shaped cavity in the medium. Five aspects differentiate the continuum models [Jensen, 2007]:
(a) How the size and the shape of the cavity are defined.
(b) How the cavity/dispersion contribution is calculated.
(c) How the dielectric medium is described.
(d) How the charge distribution is defined.
(e) How the solute is treated (i.e. with QM or MM).

The first continuum solvation models developed by Born [1920], Kirkwood [1934] and Born [1936] were mainly based on considering the electrostatic effects by characterising the solvent by a scalar, static dielectric constant, and on assuming a linear response of the solvent to a perturbing electric field. These kinds of models provide a satisfactory description of the bulk phase, however, they are not able to represent the first-solvation-shell effects, i.e. the effects occurring in the close area around the solute, as the properties of the solvent molecules in the first-solvation-shell area differ

from the solvent properties in the bulk. One of the approaches that was proposed to tackle this problem and has been widely applied is the solvent-accessible surface area (SASA) approach of Lee and Richards [1971] and Lee and Richards [1972]. The SASA is defined as the area traced out by a centre of a ball, whose radius is the effective half width of the first solvent shell, rolling over the surface of a solute. How this area is calculated differs from model to model [Cramer and Truhlar, 1999].

An important concept when referring to solute polarisation in solution is the reaction field. The reaction field is the electric field that the polarised solvent exerts on the solute. When this is included in the solute Hamiltonian, the electric moments of the solute change, resulting in further changes in the polarisation of the solvent, etc. The approach used to iterate to self-consistency is called the self-consistent reaction field (SCRF) method.

One of the most commonly used methods today is the polarizable continuum model (PCM) (or DPCM for dielectric PCM) introduced by Miertuš et al. [1981] that combines SCRF with a boundary element problem with apparent surface charges (ASC). In PCM, a van der Waals cavity formed by overlapping spheres with empirically determined radii is considered. A more recent version of PCM is the integral equation formalism PCM (IEFPCM) [Cancès et al., 1997], where the ASC are calculated with the use of the electrostatic potential, instead of the normal component of the electric field that is used in DPCM. Another popular model is COSMO (conductor-like screening model) developed by Klamt and Schüürmann [1993]. In COSMO, the surrounding medium is described as a conductor, i.e. it assumes infinite dielectric constant, to determine the ASC, which then are scaled by a function of the real dielectric constant. Klamt [1995] extended COSMO to COSMO-RS, where RS stands for real solvents. In COSMO-RS, interactions between the molecular surfaces of all the molecules in the liquid are modelled using the screening charge of densities. Starting in 1991 Cramer and Truhlar [1991] have developed a series of generalised Born type solvation models [Cramer and Truhlar, 2006], [Cramer and Truhlar, 2008], [Marenich et al., 2009], under the name SMx, where x represents the different versions. One of the basic concepts of SMx models that the partitioning of

the solvation free energy into two parts, instead of the three-contribution expression given in equation (4.71) earlier: the term that accounts for shorter range polarisation and non-electrostatic effects, like cavitation (C), dispersion (D) and changes to the solvent structure (S) caused by the solute, denoted as ΔG^{CDS}, and the term that accounts for the polarisation of the medium, for changes in the electronic structure of the solute and for changes in the nuclear coordinates, denoted here as ΔE^{ele} [1]. A brief description of the latest version of SMx models, SMD [Marenich et al., 2009], is given in section 4.5.4.

QM/MM/Continuum models

In order to predict better the interactions between the solute and nearby solvent molecules, QM/MM/continuum methods represent a number of solvent molecules in the close vicinity of the solute with MM force fields and the rest of the solvent treated as a continuum. The problem that arises here is defining the boundary between the MM and the continuum region. QM/MM/continuum methods include the general liquid optimised boundary [Brancato et al., 2008] (GLOB) model and the solvated macromolecule boundary potential [Benighaus and Thiel, 2009] (SMBP) model developed for biomolecules.

Predictive capabilities of continuum solvation models

Comparing the accuracy of the various solvation models is not a trivial issue [Jalan et al., 2010]. Firstly, all methods use parameters and a wrong choice of parameters may lead to false conclusions [Klamt et al., 2009]. Moreover, the training sets and theories used for the parametrisation are different for each model; it is very important that the QM method used should match the method used for the parametrisation of the non-electrostatic terms [Klamt et al., 2009]. Last but not least is the lack of sufficient experimental data to compare with, as there are systems that are difficult to measure [Jalan et al., 2010].

[1] In Marenich et al. [2009] this term is denoted as ΔG_{ENP}.

Recent methods seem to predict accurately the solvation free energies with average errors of around 0.5-1 kcal/mol for neutral solutes. More specifically, Cramer and Truhlar [2008] report for their SM8 model a mean absolute deviation (MAD) from experimental data of 0.59 kcal/mol for 940 neutral solutes, while Klamt et al. [2009], for the same 940 solutes, report a MAD of 0.48 kcal/mol for COSMO-RS [Klamt, 1995]. For the MST IEF-PCM [Curutchet et al., 2001] method a MAD of 0.64 kcal/mol is reported for non-aqueous solvation and a MAD of 1.01 kcal/mol for aqueous solvation, but only when considering the subset of data for which the model was parametrised. The reported errors are, however, larger for the solvation energies of ions; 4.31 kcal/mol MAD for SM8 and 6.72-12.49 kcal/mol for the non-SMx models tested in Cramer and Truhlar [2008].

The latest version of SMx models, SMD (D standing for density) [Marenich et al., 2009], has proved to be very accurate as well, although slightly less accurate than the SM8 model; SMD gives a MAD of 0.62 kcal/mol versus 0.55 obtained from SM8 for neutral solutes in water, and 0.63 kcal/mol versus 0.57 kcal/mol for neutral solutes in organic solvents. Finally, for ionic solutes the MAD for SMD is 4.30 kcal/mol versus 4.21 kcal/mol of SM8. These deviations refer to calculations with levels of theories and basis sets that the models have been parametrised. The density-based SMD model does not use partial atomic charges like generalised Born approximation based SM8 model, but considers a continuum charge density of the solute. Thus, it is not restricted to theories with available suitable charges. Therefore, although SM8 is slightly more accurate when using theories and basis sets for which it has been developed, the SMD model can be more broadly applied and is bound to be more accurate when using any other level of theory and basis set. A more detailed description of the SMD model is given in the next section and the derivation of an expression for the liquid phase reaction rate constant is presented based on this model.

4.5.4 Reaction rate constant for the liquid phase by CTST and SMD solvation model

As discussed in the previous section, the SMx solvation models express the molar free energy of solvation as the sum of an electrostatic contribution ΔE^{ele} (equal to $E^{ele,L}$ - $E^{ele,IG}$) and a non-electrostatic contribution $G^{CDS,L}$

$$\Delta \underline{G}_i^{\circ,solv} = \underline{E}_i^{ele,L} - \underline{E}_i^{ele,IG} + \underline{G}_i^{CDS,L}. \tag{4.72}$$

In the majority of SMx models, including SM8 [Cramer and Truhlar, 2008], which is claimed to have the highest accuracy among the SMx models, the generalised Born approximation for the bulk electrostatic calculations is used, in which the solute is represented as a collection of partial atomic charges in a cavity. In contrast, in the SMD model [Marenich et al., 2009] a continuum charge density of the solute is used and the non-homogeneous Poisson equation (NPE) with the IEF-PCM protocol [Cancès et al., 1997] is solved. While the applicability and accuracy of SM8 depends on the availability of reliable partial atomic charges, as a density-based model SMD can be applied effectively to cases in which the explicit charges are not defined and hence SM8 cannot be used. Furthermore, SM8 was developed only for density functional theory and Hartree-Fock theory, whereas SMD can be used with any of the electronic structure methods to which the PCM method can be applied. SMD was parametrised for the IEF-PCM method, but the parameters can be used with other models as well that solve the NPE, such as the COSMO algorithm [Klamt and Schüürmann, 1993].

The electrostatic term

According to the electrostatic theory of dielectric media [Cramer, 2004], the medium is described by a dielectric constant ε, which is scalar constant for isotropic homogeneous media and a scalar function of position for isotropic non-homogeneous media. For a linear isotropic homogeneous medium, Poisson's equation is given by [Cramer, 2004]

$$\varepsilon \nabla^2 \Phi = -4\pi \rho_f, \tag{4.73}$$

where ρ_f is the charge density per unit volume (in SMD it corresponds to the solute charge density), also called the free-charge density, and Φ is the electric potential. In the case of an isotropic non-homogeneous medium, the free-charge density and the electric potential become functions of position $\mathbf{R}$, and the non-homogeneous Poisson's equation (NPE) is used [Cramer, 2004]

$$\nabla\left(\varepsilon\nabla\Phi(\mathbf{R})\right) = -4\pi\rho_f(\mathbf{R}). \tag{4.74}$$

From quantum mechanics, the electrostatic contribution to the free energy of solvation is related to the reaction field ϕ, which is the difference between the total electric potential Φ and the electric potential of the gas phase molecule $\Phi^{(0)}$, by the following expression [Cramer, 2004, Marenich et al., 2009]

$$\Delta E^{ele,L} = \langle\Psi|H^{(0)} - \frac{e}{2}\phi|\Psi\rangle + \frac{e}{2}\sum_j Z_j\phi_j - \langle\Psi^{(0)}|H^{(0)}|\Psi^{(0)}\rangle, \tag{4.75}$$

where $H^{(0)}$ is the solute electronic Hamiltonian in the gas phase, $\Psi^{(0)}$ is the the electronic wave function in the gas phase, Ψ is the polarised solute wave function in solution, e is the atomic unit of charge, ϕ_j is the reaction field at atom j, and Z_j is the atomic number of atom j. The reaction field at an arbitrary position $\mathbf{R}$ within the cavity is calculated by [Marenich et al., 2009]

$$\phi(\mathbf{R}) = \sum_m \frac{q_m}{|\mathbf{R} - \mathbf{R}_m|}, \tag{4.76}$$

where $\mathbf{R}_m$ is the position of an element m (known as tessarae) of surface area on the boundary between solute and solvent, and q_m is the apparent surface charge on element m. In the SMD model, the boundary between the solute cavity and the solvent continuum is defined as a superposition of nuclear-centered spheres with intrinsic Coulomb radii ρ_k. This boundary forms the solvent accessible surface area.

The cavity-dispersion-solvent structure term

The non-electrostatic part of the free energy of solvation $G^{CDS,L}$ accounts for changes in energy caused by the cavity formation, changes in dispersion, and changes of the

solvent structure induced by the solute. This term in the SMD model is the same as for the SM8 model and is given by the following expression [Marenich et al., 2009]

$$G^{CDS,L} = \sum_{j=1}^{N_A} \sigma_j A_j(\mathbf{r}, \{r_{Z_j} + r_s\}) + \sigma^{[M]} \sum_{j=1}^{N_A} A_j(\mathbf{r}, \{r_{Z_j} + r_s\}), \tag{4.77}$$

where N_A is the number of atoms, σ_j is the surface tension of atom k in cal mol^{-1} Å^{-2} and $\sigma^{[M]}$ is the molecular surface tension cal mol^{-1} Å^{-2}. A_j is the SASA of atom j in Å^2, which depends on the geometry of the solute $\mathbf{r}$, the atomic van der Waals radii r_{z_j}, and the solvent radius r_s. The atomic surface tension is calculated as follows

$$\sigma_j = \tilde{\sigma}_{Z_j} + \sum_{j'=1}^{N_A} \tilde{\sigma}_{Z_j Z_{j'}} T_j(Z_{j'}, r_{jj'}), \tag{4.78}$$

where $\tilde{\sigma}_{Z_j}$ is a parameter that depends on the atomic number of atom j, $\tilde{\sigma}_{Z_j Z_{j'}}$ is a parameter that depends on the atomic number of both atoms j and j', and $T_j(Z_{j'}, r_{jj'})$ is a geometry-dependent function, called cutoff tanh. The atomic and molecular surface tensions depend on bulk solvent properties

$$\tilde{\sigma}_\theta = \tilde{\sigma}_\theta^{[n_D]} n_D + \tilde{\sigma}_\theta^{[\alpha]} \alpha + \tilde{\sigma}_\theta^{[\beta]} \beta \tag{4.79}$$

and

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]} \gamma + \tilde{\sigma}^{[\varphi^2]} \varphi^2 + \tilde{\sigma}^{[\psi^2]} \psi^2 + \tilde{\sigma}^{[\beta^2]} \beta^2, \tag{4.80}$$

where the subscript θ stands for either Z_j or $Z_j Z_j'$, and the solvent properties are as follows: n_D is the refractive index at 293 K, α is Abraham's hydrogen bond acidity, β is Abraham's hydrogen bond basicity, γ is the macroscopic surface tension at 298.15 K in units of cal mol^{-1} $\mathring{A}^{-2}$, φ is the aromaticity, and ψ is the halogenicity. The coefficients $\tilde{\sigma}_\theta^{[n_D]}$, $\tilde{\sigma}_\theta^{[\alpha]}$, $\tilde{\sigma}_\theta^{[\beta]}$ are regressed parameters that depend on θ, while $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\varphi^2]}$, and $\tilde{\sigma}^{[\psi^2]}$ are regressed parameters independent of θ.

The complete expression for the reaction rate constant in dm^3 mol^{-1} s^{-1}, according to CTST and the SMD model, is obtained by combining equations (4.70) and (4.72)

$$k^L = \kappa \frac{k_B T}{h} \prod_i \left(c_i^{\circ,L}\right)^{\nu_i} \prod_i \left(q_i^{',IG}\right)^{\nu_i} \cdot \exp\left[-\frac{\Delta^\ddagger \Delta \underline{G}^{p\to c,IG} + \Delta^\ddagger \underline{E}^{ele,L} + \Delta^\ddagger \underline{G}^{CDS,L}}{RT}\right]. \tag{4.81}$$

The liquid rate constant in equation (4.81) is only a function of the geometry of the species (reactants and TS) and seven solvent properties: refractive index, Abraham's hydrogen bond acidity, Abraham's hydrogen bond basicity, surface tension, aromaticity and halogenicity.

4.6 Prediction of the rate constant for a Menschutkin reaction using conventional transition state theory

Conventional transition state theory in combination with the SMD solvation model for the Gibbs free energy of solvation has been applied here to predict the rate constant of a Menschutkin reaction. Menschutkin was the first to report in 1890 the solvent effect on the reaction rate based on his studies of the reaction of triethylamine with iodoethane [Menschutkin, 1890a], [Menschutkin, 1890b]. Reactions between an tertiary amine and an alkyl halide that produce a quaternary ammonium salt have since become known as Menschutkin reactions. The Menschutkin reactions are ideal for the study of solvent effects on reaction rate constants, since rate constants can vary orders of magnitude according to the solvent. The Menschutkin reaction of phenacyl bromide with pyridine (figure 4.3) has been chosen to demonstrate the solvent design methodology. The kinetics of this particular reaction have been studied experimentally by many authors: Pearson et al. [1952], Halvorsen and Songstad [1978], Barnard and Smith [1981], Yoh et al. [1981], Forster and Laird [1982], Hwang et al. [1983], Shunmugasundaram and Balakumar [1985], Winston et al. [1996], Koh et al. [2000], Ganase et al. [2013].

The aim of this study is to investigate the predictive capabilities of CTST and SMD, in the context of chemical kinetics, and the effect of the choice of the level of theory and basis set on the predictions. The Menschutkin reaction of phenacyl bromide with pyridine has been previously studied computationally [Struebing et al., 2010], [Struebing, 2011], [Struebing et al., 2013]. Struebing [2011] tested different sol-

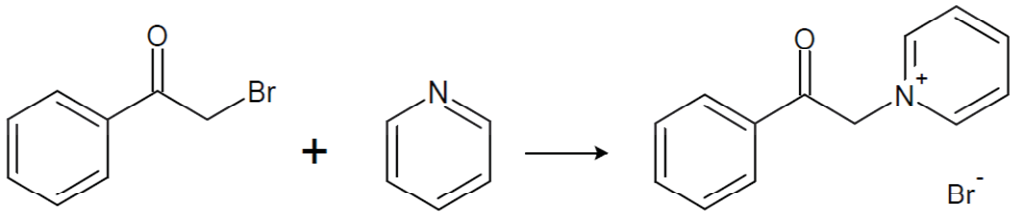


Figure 4.3: The Menschutkin reaction of phenacyl bromide and pyridine.

vation models for the prediction of the reaction rate constant in seven organic solvents (toluene, chloroform, tetrahydrofuran, acetone, methanol, ethanol, acetonitrile). The solvation models considered were SM8, SMD and IEF-PCM and the levels of theory used were B3LYP/6-31+G(d) and M06-2X/6-31+G(d). The author concluded that the SMD model with B3LYP/6-31+G(d) gave the best predictions, followed by SMD with M06-2X/6-31+G(d), but M06-2X seemed to overestimate the rate constants.

In this work, the effect of the level of theory and basis set is further explored using the SMD solvation model. Three levels of theories have been tested; B3LYP, M05-2X and M06-2X. The basis sets considered are: 6-31G(d), 6-31+G(d), 6-31+G(d,p) and 6-311++G(d,p). The rate constant is calculated using equation (4.81). The same level of theory and basis set are used for calculations in both the gas and the liquid phase. In each calculation, the geometries of the reactants and transition state are optimised not only in the gas phase but also in the liquid phase. The Gaussian 09 package was used in all of the calculations [Frisch et al., 2009].

The transmission coefficient κ is calculated using the simple, widely used expression of the Wigner tunnelling correction factor [Garrett and Truhlar, 1979a, Henriksen and Hansen, 2008, Ashcraft et al., 2008, Struebing et al., 2013]

$$\kappa = 1 + \frac{1}{24}\left(\frac{hv^{\ddagger}}{k_B T}\right)^2 \tag{4.82}$$

where $v^{\ddagger}$ is the magnitude of the imaginary frequency of the transition state structure in cm^{-1}. Gas-phase frequencies are used throughout this work for the transmission coefficient κ, as it has been shown that the impact of using liquid phase frequencies is minor [Zhao et al., 2005].

The rate constant is calculated for five organic solvents; toluene, chloroform,

tetrahydrofuran, acetone and acetonitrile. Experimental rate constants are available for all five solvents [Ganase et al., 2013]. The calculations have been performed in Gaussian 09 [Frisch et al., 2009].

Level of theory	$\mathbf{k}^{TST}$ **in dm^3 mol^{-1} s^{-1}**				
	toluene	chloroform	tetrahydrofuran	acetone	acetonitrile
B3LYP/6-31G(d)	7.392E-08	1.032E-06	3.696E-06	1.119E-05	1.714E-05
B3LYP/6-31+G(d)	1.659E-06	2.425E-05	8.930E-05	2.831E-04	4.451E-04
B3LYP/6-31+G(d,p)	1.632E-06	2.178E-05	8.232E-05	2.571E-04	4.035E-04
B3LYP/6-311++G(d,p)	8.477E-08	1.099E-06	4.209E-06	1.238E-05	1.902E-05
M05-2X/6-31G(d)	1.632E-05	1.813E-04	5.810E-04	1.544E-03	2.342E-03
M05-2X/6-31+G(d)	7.368E-04	9.640E-03	3.188E-02	1.119E-05	1.714E-05
M06-2X/6-311++G(d)	1.016E-06	1.081E-05	4.033E-05	1.235E-04	1.875E-04
experimental[a]	1.145E-04	2.079E-04	2.585E-04	1.228E-03	2.613E-03

Table 4.3: Predicted rate constants for the Menschutkin reaction for five organic solvents using CTST with different levels of theory and the SMD solvation model. [a]The experimental values are from Ganase et al. [2013].

The predictions for the rate constant in the five solvents for different levels of theory and basis sets are given in Table 4.3 and illustrated in figure 4.4. The closest predictions to the experimental values are obtained by M05-2X/6-31G(d) for all solvents. The results are consistent with the reported errors in Marenich et al. [2009], where M05-2X/6-31G(d) is found to be the method with the lowest mean unsigned error, 0.64 kcal/mol, among the methods used in the parametrisation of the SMD model. It is important to note that, even though the general rule for basis sets is "the larger the better", when using continuum solvation models that have been parametrised from experimental data, the method for the QM calculations should be consistent with the parametrisation of the solvation model [Ho et al., 2010]. The SMD model has been parametrised with small basis sets, thus large basis sets do not increase the accuracy but, on the contrary, they lead to a deterioration of the predic-

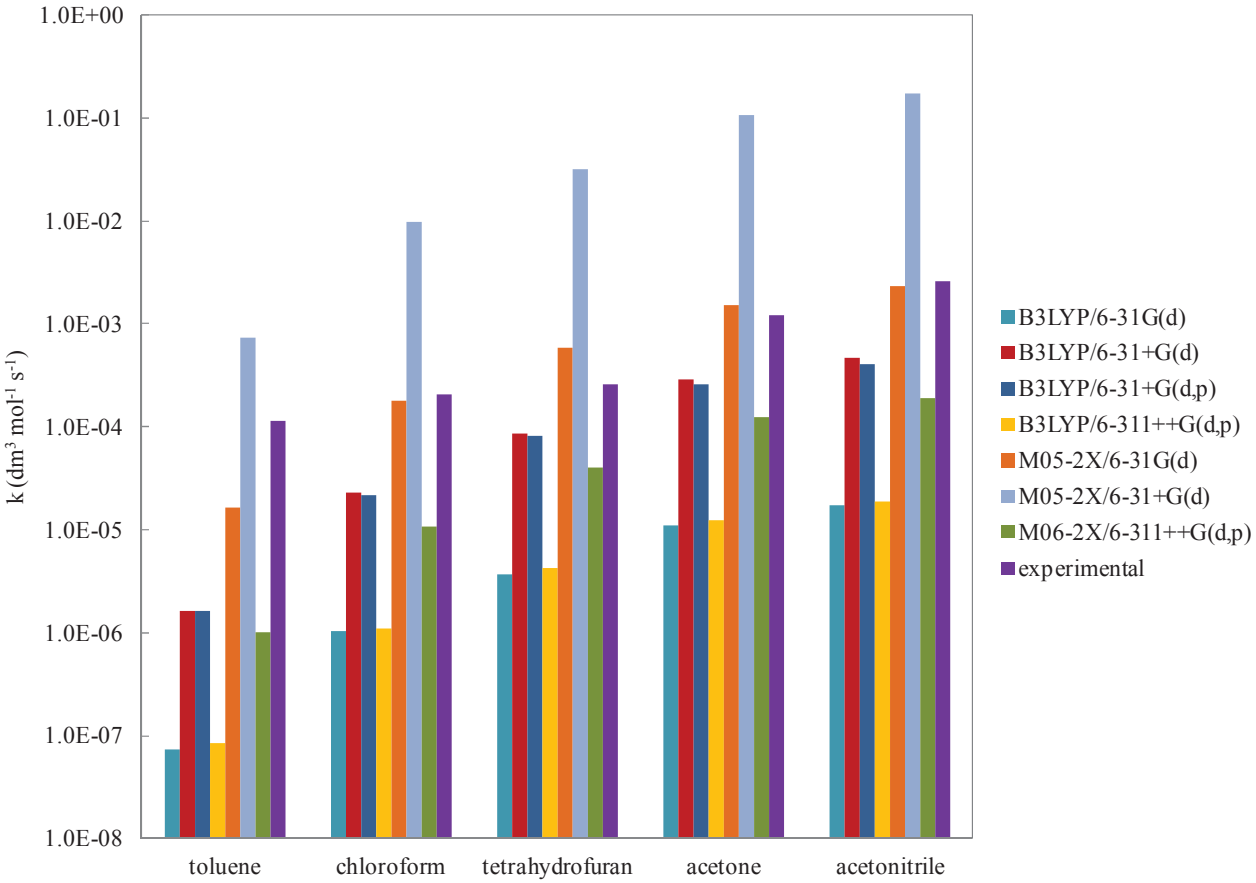


Figure 4.4: Predicted rate constants for the Menschutkin reaction of phenacyl bromide and pyridine in different organic solvents with various methods. Experimental values are from Ganase et al. [2013].

tions in the liquid phase. B3LYP/6-31+G(d) is also satisfactorily accurate, following M05-2X/6-31G(d). Although SMD was parametrised with B3LYP/6-31G(d), here the diffusion function (i.e. the + in the basis set) improves significantly the predictions. This is possibly due to the presence of bromide in the molecules which is large and, especially in the ionic form, would require a diffusion function. This behaviour is not observed with M05-2X. The cause may be that this latter model generally overestimates the solvation free energies [Marenich et al., 2009] and this overestimation compensates for the lack of the diffusion function. Although the predictions for the different methods differ quantitatively, the trend of the rate constants is the same for all methods. The predictions for three most accurate methods, M05-2X/6-31G(d),

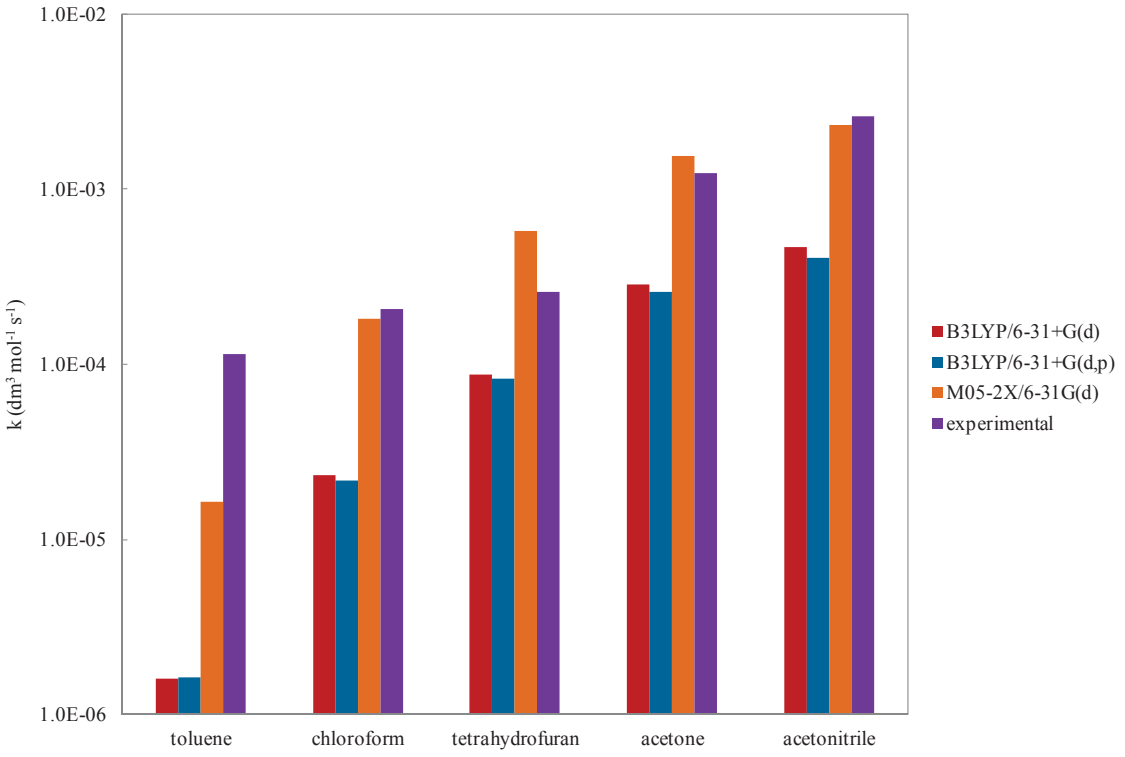


Figure 4.5: Predicted rate constants for the Menschutkin reaction of phenacyl bromide and pyridine in different organic solvents with the three best methods tested. Experimental values are from Ganase et al. [2013].

B3LYP/6-31+G(d), B3LYP/6-31+G(d,p), are shown in figure 4.5. As can be seen from Table 4.4, and was also shown in Struebing [2011] and Struebing et al. [2013], the reaction is faster in polar solvents; the higher the dielectric constant, the higher the rate constant. Suitable solvents for this reaction will be investigated in chapters 5 and 6. Information about the QM results with M05-2X/6-31G(d) are given for the gas phase in Table 4.5 and for the five organic solvents in Table 4.6.

Solvent	ε	$\mathbf{k}^{TST}$ **in dm**^{3} **mol**$^{-1}$ **s**$^{-1}$
toluene	2.374	1.632E-05
chloroform	4.711	1.813E-04
tetrahydrofuran	7.426	5.810E-04
acetone	20.493	1.544E-03
acetonitrile	35.688	2.342E-03

Table 4.4: The dielectric constants, ε, and predicted rate constants for all solvents tested for the Menschutkin reaction of phenacyl bromide and pyridine using M05-2X/6-31G(d) method.

	$v^{\ddagger}$	$\underline{E}^{ele,IG}$	$ZPVE$	$q^{'},^{IG}$
	[cm^{-1}]	[a.u. Particle^{-1}]	[J mol^{-1}]	-
Transition state	-474.657	-3204.15631	589201.1	0.12238E-82
Phenacyl bromide	-	-2955.93152	347002.6	0.52920E-44
Pyridine	-	-248.25171	238567.1	0.62517E-29

Table 4.5: Predicted values for the imaginary frequency $v^{\ddagger}$, the electronic energy $\underline{E}^{ele,IG}$, the zero-point vibrational energy $ZPVE$, and the total partition function $q^{'},^{IG}$, in the gas phase for the Menschutkin reaction of phenacyl bromide and pyridine. All calculations are performed at M05-2X/6-31G(d) level of theory.

Solvent	**Transition state**			**Phenacyl bromide**		**Pyridine**	
	$v^{\ddagger}$	$\underline{E}^{ele,L} + \underline{G}^{CDS,L}$	$\underline{G}^{CDS,L}$	$\underline{E}^{ele,L} + \underline{G}^{CDS,L}$	$\underline{G}^{CDS,L}$	$\underline{E}^{ele,L} + \underline{G}^{CDS,L}$	$\underline{G}^{CDS,L}$
	[cm^{-1}]	[a.u. Particle^{-1}]	[kcal mol^{-1}]	[a.u. Particle^{-1}]	[kcal mol^{-1}]	[a.u. Particle^{-1}]	[kcal mol^{-1}]
toluene	-480.3314	-3204.18453	-8.47	-2955.94583	-6.04	-248.25962	-2.27
chloroform	-479.3762	-3204.19036	-7.11	-2955.94724	-4.85	-248.26177	-2.16
tetrahydrofuran	-488.5044	-3204.19067	-5.87	-2955.94727	-4.49	-248.26096	-1.00
acetone	-490.5976	-3204.19436	-5.61	-2955.94871	-4.28	-248.26229	-1.02
acetonitrile	-490.7449	-3204.19434	-4.83	-2955.94820	-3.59	-248.26238	-0.87

Table 4.6: Predicted values for the imaginary frequency $v^{\ddagger}$, the electronic energy $\underline{E}^{ele,L}$, the $\underline{G}^{CDS,L}$ term of the solvation free energy in various solvents, for the reactants and the transition state structure for the Menschutkin reaction of phenacyl bromide and pyridine.

4.7 Conclusions

An overview of methods for the prediction of liquid phase reaction rate constants has been presented, including empirical methods, and particularly linear free-energy relationships, and conventional transition state theory. Although linear free-energy relationships can correlate very satisfactorily kinetic data, they are restricted to availability of data. Consequently, in cases where no experimental data are available, predictive models are necessary, such as transition state theory. Expressions for the reaction rate constant in both the gas and the liquid phase using conventional transition-state theory have been presented. Different solvation models for the calculation of the free energy of solvation have been discussed. Continuum solvation models are a good compromise between accuracy and computational cost and the continuum solvation model SMD has been chosen to be used in this thesis. Finally, conventional transition-state theory and the SMD model are applied to predict the rate constant for a Menschutkin reaction in various organic solvents and are found to give good qualitative and quantitative predictions. In the rest of the thesis, the integration of the *ab initio* reaction rate constants predicted from quantum mechanics into a solvent design algorithm is discussed.

Chapter 5

The solvent design problem with integrated quantum mechanical calculations

5.1 Introduction

In this chapter, the detailed formulation of a solvent design problem for reactions is presented, where a quantum mechanical approach to calculating rate constants, as described in chapter 4, is used. In the solvent design formulation, we build on the work of several authors. Sheldon et al. [2006] proposed an approach for solvent design where the objective was to design a solvent that minimises the solvation Gibbs free energy. In parallel, but in the context of solvent design for reactions, Folić et al. [2004, 2005, 2006, 2007, 2008a] developed a mixed-integer linear programming CAMD formulation to design solvents that maximise the reaction rate constant, using an empirical expression for the rate constant, requiring rate constant measurements to be made in a few solvents. Following this work, Struebing et al. [2010, 2013] used QM derived data to regress an empirical expression for the rate constant. In this thesis, these three approaches are combined and a QM based methodology for the design of optimal solvents for reactions is presented. The problem is formulated as

a bilevel mixed-integer non-linear programming (MINLP) optimisation problem with black box functions. The CAMD formulation is presented in section 5.2. In section 5.3 the fixed geometry assumption that is considered in this chapter is discussed. The solvent design algorithm and its implementation are described in sections 5.4 and 5.5 respectively. An application of the approach to a Menschutkin reaction is presented in section 5.6, where results and limitations of the method are discussed.

5.2 CAMD formulation for solvents for reactions

A computer-aided molecular design framework for the design of optimal solvents for chemical reactions is described in this section. The CAMD problem is based on the CAMD formulation proposed by Struebing et al. [2010] which is a mixed-integer linear problem (MILP). Here, the rate constant is calculated directly from conventional transition state theory. Non-linear constraints are introduced for the solvent physical properties that increase the size and complexity of the problem considerably. The general problem formulation is introduced in section 5.2.1, the objective function is given in section 5.2.2, and the constraints of the optimisation problem are described in sections 5.2.3 - 5.2.6.

5.2.1 General Problem formulation

The solvent design methodology combines a CAMD formulation with a quantum mechanically derived reaction rate constant. The problem is formulated as a mixed-integer non-linear programming (MINLP) problem as follows

$$\max_{\boldsymbol{\xi},\mathbf{n},\mathbf{y}} f(\boldsymbol{\xi}) \tag{5.1}$$

subject to

$$h_1(\boldsymbol{\xi},\mathbf{n},\mathbf{y}) = \mathbf{0} \tag{5.2}$$

$$g_1(\boldsymbol{\xi},\mathbf{n},\mathbf{y}) \leq \mathbf{0} \tag{5.3}$$

$$h_2(\mathbf{n},\mathbf{y}) = \mathbf{0} \tag{5.4}$$

$$g_2(\mathbf{n}, \mathbf{y}) \leq \mathbf{0} \tag{5.5}$$

$$d(\boldsymbol{\xi}, \mathbf{n}, \mathbf{y}) \leq \mathbf{0} \tag{5.6}$$

$$\boldsymbol{\xi} \in \mathbb{R}^m \tag{5.7}$$

$$\mathbf{n} \in \mathbb{R}^{\mathbf{q}} \tag{5.8}$$

$$y_i \in \{0, 1\}^u \quad i = 1, ..., q \tag{5.9}$$

where f is the objective function, h_1 is a set of structure-property equality constraints, g_1 is a set of structure-property inequality constraints, h_2 is a set of chemical feasibility and molecular complexity equality constraints, g_2 is a set of chemical feasibility and molecular complexity inequality constraints and d is a set of design constraints. $\boldsymbol{\xi}$ is a m-dimensional vector of continuous variables such as kinetic properties and solvent physical properties, $\mathbf{n}$ is a q-dimensional vector of variables denoting the number of groups in a molecule and $\mathbf{y}$ is a $q \times u$ matrix of binary variables used to activate groups and constrain continuous variables to integer values.

5.2.2 Objective function

The objective function of the problem is some measure of performance of the reactive system. In this thesis, a single reaction is considered and the focus is on kinetics, thus the objective is to maximise the rate constant of a particular reaction. Naturally, the rate of a reaction depends also on concentration, since rate = $k \times$ concentration, but concentration is not taken into account here. The reaction rate constant is calculated from conventional transition state theory, therefore

$$f(\boldsymbol{\xi}) = k^{TST}, \tag{5.10}$$

where k^{TST} is the reaction rate constant from conventional transition state theory.

Reaction rate constant

The rate constant is calculated from Conventional Transition State Theory (CTST) using the SMD solvation model [Marenich et al., 2009] as described in section 4.5.4, where it is given by the following expression:

$$k^{TST} = \kappa \frac{k_B T}{h} \prod_i \left(c_i^{\circ,L}\right)^{\nu_i} \prod_i \left(q_i^{',IG}\right)^{\nu_i} \cdot \exp\left[-\frac{\Delta^\ddagger \Delta \underline{G}^{p\to c,IG} + \Delta^\ddagger \underline{E}^{ele,L}(\mathbf{r}^*,\varepsilon,\alpha) + \Delta^\ddagger \underline{G}^{CDS,L}(\mathbf{r}^*, n_D, \alpha, \beta, \gamma, \varphi, \psi)}{RT}\right], \tag{5.11}$$

where $\mathbf{r}^*$ is a vector of optimised geometries, ε is the dielectric constant, α is Abraham's hydrogen bond acidity, β is Abraham's hydrogen bond basicity, γ is the surface tension, φ is the aromaticity, and ψ is the halogenicity. Recall that the symbol $\Delta^\ddagger$ denotes the difference of the stoichiometric coefficients between the transition state and the reactants ($\nu_{TS} - \nu_A - \nu_B$). Thus, for the electrostatic energy

$$\Delta^\ddagger \underline{E}^{ele,L}(\mathbf{r}^*,\varepsilon,\alpha) = \nu_{TS}\underline{E}^{ele,L}_{TS}(\mathbf{r}^*_{TS},\varepsilon,\alpha) - \nu_A \underline{E}^{ele,L}_A(\mathbf{r}^*_A,\varepsilon,\alpha) - \nu_B \underline{E}^{ele,L}_B(\mathbf{r}^*_B,\varepsilon,\alpha), \tag{5.12}$$

where

$$\underline{E}^{ele,L}_i(\mathbf{r}^*_i,\varepsilon,\alpha) = \min_{r_i} \underline{E}^{ele,L}_i(\mathbf{r}_i,\varepsilon,\alpha), \tag{5.13}$$

and the vector of optimised geometries $\mathbf{r}^*$ is defined as

$$\mathbf{r}^*_i = \arg\min_{r_i} \underline{E}^{ele,L}_i(\mathbf{r}_i,\varepsilon,\alpha), \quad i = TS, A, B. \tag{5.14}$$

For $\underline{G}^{CDS,L}$:

$$\begin{aligned}\Delta^\ddagger \underline{G}^{CDS,L}(\mathbf{r}^*, n_D, \alpha, \beta, \gamma, \varphi, \psi) &= \underline{G}^{CDS,L}_{TS}(\mathbf{r}^*_{TS}, n_D, \alpha, \beta, \gamma, \varphi, \psi) \\ &- \underline{G}^{CDS,L}_A(\mathbf{r}^*_A, n_D, \alpha, \beta, \gamma, \varphi, \psi) - \underline{G}^{CDS,L}_B(\mathbf{r}^*_B, n_D, \alpha, \beta, \gamma, \varphi, \psi),\end{aligned} \tag{5.15}$$

where $\mathbf{r}^*_\mathbf{i}$ is the optimised geometry for the corresponding minimum $\underline{E}^{ele,L}_i$. The analytical expressions of $\underline{G}^{CDS,L}$ are given in equations (4.77) - (4.80).

In this thesis, the expression of the reaction rate constant in the optimisation problem, as it requires quantum mechanical calculations, is treated as a black box function with only input the seven solvent properties (ε, n_D, α, β, γ, φ, ψ).

5.2.3 Structure-property constraints

Structure - property constraints of the form $h_1(\boldsymbol{\xi}, \mathbf{n}, \mathbf{y}) = \mathbf{0}$ and $g_1(\boldsymbol{\xi}, \mathbf{n}, \mathbf{y}) \leq \mathbf{0}$ are presented in this section. They relate the molecular structure of the solvent to its properties and, in turn, the properties of the solvent to the rate constant(s).

The solvent properties, required in the $G^{CDS,L}$ term of the liquid phase free energy, are calculated with the group-contribution methods developed by Sheldon et al. [2005] and Folić et al. [2007]. The reported regression statistics for the properties that are used here are shown in Table 5.1. It is worth mentioning that in the case of the dielectric constant, ε, the regression is significantly better ($R^2 = 0.714$) for the set of compounds with $\varepsilon \leq 40$, which is mostly the case in this thesis.

Property	Data points	AAE	AAPE(%)	SD	R^2
n_D	797	0.03	2.18	0.046	0.556
α	350	0.017			
β	350	0.043			
γ	214	3.08	10.90	4.160	0.737
ε	337	3.03	18.50	13.47	0.350
μ	870	0.50	36.85	0.85	0.459
δ_H	664	0.079	3.96	1.48	0.826

Table 5.1: Regression statistics from Sheldon et al. [2005] for refractive index (n_D), macroscopic surface tension (γ) in dynes cm^{-1}, dielectric constant (ε), dipole moment (μ) and Hildebrand solubility parameter δ_H in MPa$^{0.5}$, and from Folić et al. [2007] for acidity (α) and basicity (β).

The group-contribution methods of Constantinou et al. [1995] and Marrero and Gani [2001] are applied for the calculation of the liquid molar volume and enthalpy of vaporisation, respectively. In the rest of this section, group-contribution expressions for the solvent physical properties [Constantinou et al., 1995, Marrero and Gani, 2001, Sheldon et al., 2005, Folić et al., 2007] are presented for completeness.

Dielectric constant, ε

The dielectric constant is given by the following expression

$$\varepsilon = \begin{cases} n_D^2 + 0.1 & \text{if } \mu < 0.5D \\ 0.91\,(48\mu^2 - 15.5\mu^3)\,V_m^{-0.5} + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 & \text{otherwise.} \end{cases} \tag{5.16}$$

The terms ε_1, ε_2, ε_3 are correction factors and are given from the following equations

$$\varepsilon_1 = \begin{cases} 70\left(\sum_{i\in G_{1-9}} n_i + 4.5\right)^{-1} & \text{if } \sum_{i\in G_1} n_i \geq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.17}$$

where G_1 = { OH, CH_3CO, CH_2CO, CHO, CH_2CN, CH_2NO_2, $CHNO_2$ } and G_{1-9} = { CH_3, CH_2, CH, C, CH_2=CH, CH=CH, CH_2=C, CH=C, C=C }.

$$\varepsilon_2 = \begin{cases} -16\left(\sum_{i\in G_{1-9}} n_i + 3\right)^{-1} & \text{if } n_{COOH} \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.18}$$

$$\varepsilon_3 = \begin{cases} 2.5 & \text{if } \sum_{i\in G_3} n_i \geq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.19}$$

where G_3 = { CH_2Cl, CHCl, $CHCl_2$ }. In order to include the above equations in the CAMD formulation, they have to be reformulated in an algebraic form. The binary variable, y_ϵ is introduced,

$$y_\epsilon = \begin{cases} 0 & \text{if } \mu \leq 0.5 \\ 1 & \text{otherwise,} \end{cases} \tag{5.20}$$

which algebraically is expressed as follows

$$\mu - My_\varepsilon \leq 0.5 \tag{5.21}$$

$$M(y_\varepsilon - 1) - \mu \leq -0.5, \tag{5.22}$$

where M is a large positive number; the value of 100 is used throughout here. The dielectric constant, ε , is calculated with the use of four slack variables $s^+_{\varepsilon,1}$, $s^-_{\varepsilon,1}$, $s^+_{\varepsilon,2}$ and $s^-_{\varepsilon,2}$ and, instead of equation (5.16), the following expressions are used

$$s^+_{\varepsilon,1} + s^-_{\varepsilon,1} - My_\varepsilon \leq 0 \tag{5.23}$$

$$s^+_{\varepsilon,2} + s^-_{\varepsilon,2} - M(1 - y_\varepsilon) \leq 0 \tag{5.24}$$

$$\varepsilon - (n_D^2 + 0.1) + s^+_{\varepsilon,1} - s^-_{\varepsilon,1} = 0 \tag{5.25}$$

$$\varepsilon - (0.91(48\mu^2 - 15.5\mu^3)V_m^{-0.5} + \varepsilon_1 + \varepsilon_2 + \varepsilon_3) + s^+_{\varepsilon,2} - s^-_{\varepsilon,2} = 0 \tag{5.26}$$

$$s^+_{\varepsilon,1}, s^-_{\varepsilon,1}, s^+_{\varepsilon,2}, s^-_{\varepsilon,2} \geq 0. \tag{5.27}$$

Binary variables, y_{ε_1}, y_{ε_2} and y_{ε_3}, are also introduced for the calculation of ε_1, ε_2 and ε_3

$$y_{\epsilon_1} = \begin{cases} 0 & \text{if } \sum_{i \in G_1} n_i \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{5.28}$$

$$y_{\epsilon_2} = \begin{cases} 0 & \text{if } n_{COOH} \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{5.29}$$

$$y_{\epsilon_3} = \begin{cases} 0 & \text{if } \sum_{i \in G_3} n_i \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{5.30}$$

equivalently

$$y_{\varepsilon_1} - \sum_{i \in G_1} n_i \leq 0 \tag{5.31}$$

$$\frac{\sum_{i \in G_1} n_i}{\sum_{i \in G_1} n_i^U} - y_{\varepsilon_1} \leq 0 \tag{5.32}$$

$$n_{COOH} - y_{\varepsilon_2} n^U_{COOH} \leq 0 \tag{5.33}$$

$$y_{\varepsilon_2} - n_{COOH} \leq 0 \tag{5.34}$$

$$y_{\varepsilon_3} - \sum_{i \in G_3} n_i \leq 0 \tag{5.35}$$

$$\frac{\sum_{i \in G_3} n_i}{\sum_{i \in G_3} n_i^U} - y_{\varepsilon_3} \leq 0. \tag{5.36}$$

The following four slack variables $s^+_{\varepsilon_1,1}$, $s^-_{\varepsilon_1,1}$, $s^+_{\varepsilon_1,2}$, $s^-_{\varepsilon_1,2}$ are introduced for the calculation of ε_1 and ε_2

$$s^+_{\varepsilon_1,1} + s^-_{\varepsilon_1,1} - M(1 - y_{\varepsilon_1}) \leq 0 \tag{5.37}$$

$$s^+_{\varepsilon_1,2} + s^-_{\varepsilon_1,2} - M y_{\varepsilon_1} \leq 0 \tag{5.38}$$

$$\varepsilon_1 - 70 \left(\sum_{i \in G_{1-9}} n_i + 4.5 \right)^{-1} s^+_{\varepsilon_1,1} - s^-_{\varepsilon_1,1} = 0 \tag{5.39}$$

$$\varepsilon_1 + s^+_{\varepsilon_1,2} - s^-_{\varepsilon_1,2} = 0 \tag{5.40}$$

$$s^+_{\varepsilon_1,1}, s^-_{\varepsilon_1,1}, s^+_{\varepsilon_1,2}, s^-_{\varepsilon_1,2} \geq 0 \tag{5.41}$$

and

$$s^+_{\varepsilon_2,1} + s^-_{\varepsilon_2,1} - M(1 - y_{\varepsilon_2}) \leq 0 \tag{5.42}$$

$$s^+_{\varepsilon_2,2} + s^-_{\varepsilon_2,2} - M y_{\varepsilon_2} \leq 0 \tag{5.43}$$

$$\varepsilon_2 - 16 \left(\sum_{i \in G_{1-9}} n_i + 3 \right)^{-1} s^+_{\varepsilon_2,1} - s^-_{\varepsilon_2,1} = 0 \tag{5.44}$$

$$\varepsilon_2 + s^+_{\varepsilon_2,2} - s^-_{\varepsilon_2,2} = 0 \tag{5.45}$$

$$s^+_{\varepsilon_2,1}, s^-_{\varepsilon_2,1}, s^+_{\varepsilon_2,2}, s^-_{\varepsilon_2,2} \geq 0. \tag{5.46}$$

Finally ε_3 is given by

$$\varepsilon_3 - 2.5 y_{\varepsilon_3} = 0. \tag{5.47}$$

Dipole moment, μ

The dipole moment according to Sheldon et al. [2005] can be calculated from the following equation

$$\mu = \begin{cases} 0.11 \left(\sum_{i \in G} n_i d_i \right)^{0.29} V_m^{-0.16} & \text{if } \sum_{i \in G/G_{HC}} n_i \geq 1 \text{ and if } \sum_{i \in G} n_i d - i \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{5.48}$$

where d_i is the contribution of group i to the dipole moment and
$G_{HC} = \{$ CH_3, CH_2, CH, C, CH_2=CH, CH=CH, CH_2=C, CH=C, C=C, aCH, aC, $aCCH_3$, $aCCH_2$, aCCH $\}$. Three binary variables are considered, y_{μ_1}, y_{μ_2} and y_μ

$$y_{\mu_1} = \begin{cases} 1 & \text{if } \sum_{i \in G \setminus G_{HC}} n_i \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.49}$$

$$y_{\mu_2} = \begin{cases} 1 & \text{if } \sum_{i \in G_1} n_i d_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.50}$$

$$y_\mu = \begin{cases} 1 & \text{if } y_{\mu_1} = 1 \text{ and } y_{\mu_2} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.51}$$

These binary variables can be expressed equivalently as follows

$$y_{\mu_1} - \sum_{i \in G \setminus G_{HC}} n_i \leq 0 \tag{5.52}$$

$$\frac{\sum_{i \in G \setminus G_{HC}} n_i}{\sum_{i \in G \setminus G_{HC}} n_i^U} - y_{\mu_1} \leq 0 \tag{5.53}$$

$$y_{\mu_2} - 1 - \frac{\sum_{i \in G} n_i d_i}{\sum_{i \in G} n_i^U d_i} \leq 0 \tag{5.54}$$

$$\frac{\sum_{i \in G} n_i d_i}{\sum_{i \in G} n_i^U d_i} - y_{\mu_2} \leq 0 \tag{5.55}$$

$$y_\mu - 0.4(y_{\mu_1} + y_{\mu_2}) - 0.5 \leq 0 \tag{5.56}$$

$$y_{\mu_1} + y_{\mu_2} - 1.5 - y_\mu \leq 0. \tag{5.57}$$

Two positive slack variables s_μ^+ and s_μ^- are defined and the dipole moment, μ is calculated from

$$s_\mu^+ + s_\mu^- - M(1 - y_\mu) \leq 0 \tag{5.58}$$

$$\mu - 0.11 \left(\sum_{i \in G} n_i d_i \right)^{0.29} V_m^{-0.16} s_\mu^+ - s_\mu^- = 0 \tag{5.59}$$

$$\mu - M y_\mu \leq 0 \tag{5.60}$$

$$s_\mu^+, s_\mu^-, \mu \geq 0. \tag{5.61}$$

Abraham's hydrogen bond acidity, α

The hydrogen bond acidity is predicted by

$$\alpha = \begin{cases} 0 & \text{if } \sum_{i \in G} n_i A_i + 0.010641 \leq 0.029 \\ \sum_{i \in G} n_i A_i + 0.010641 & \text{otherwise,} \end{cases} \tag{5.62}$$

where A_i is the contribution of group i in HB acidity.

The binary variable, y_A is introduced and determined by

$$y_A = \begin{cases} 1 & \text{if } 0.010641 + \sum_{i \in G} n_i A_i > 0.05 \\ 0 & \text{otherwise} \end{cases} \tag{5.63}$$

alternatively,

$$\sum_{i \in G} n_i A_i - M y_A - 0.018359 \leq 0 \tag{5.64}$$

$$M(y_A - 1) \sum_{i \in G} n_i A_i + 0.018359 \leq 0. \tag{5.65}$$

Then, α is determined by the constraints:

$$-\alpha + \sum_{i \in G} n_i A_i + y_A - 0.989359 \leq 0 \tag{5.66}$$

$$\alpha - \sum_{i \in G} n_i A_i - 0.010641 \leq 0 \tag{5.67}$$

$$0 \leq \alpha \leq M y_A. \tag{5.68}$$

Abraham's hydrogen bond basicity, β

Hydrogen bond basicity is given by:

$$\beta = \begin{cases} 0 & \text{if } \sum_{i \in G} n_i B_i + 0.12371 \leq 0.124 \\ \sum_{i \in G} n_i B_i + 0.123731 & \text{otherwise,} \end{cases} \tag{5.69}$$

where B_i is the contribution of group i in HB basicity.

Binary variable y_B is determined by:

$$y_B = \begin{cases} 1 & \text{if } 0.123731 + \sum_{i \in G} n_i B_i > 0.124 \\ 0 & \text{otherwise} \end{cases} \tag{5.70}$$

equivalently,

$$\sum_{i \in G} n_i B_i - M y_B - 0.00029 \leq 0 \tag{5.71}$$

$$M(y_B - 1) \sum_{i \in G} n_i B_i + 0.00029 \leq 0. \tag{5.72}$$

Then, β is determined by the constraints:

$$-\beta + \sum_{i \in G} n_i B_i + y_B - 0.87629 \leq 0 \tag{5.73}$$

$$\beta - \sum_{i \in G} n_i B_i - 0.123731 \leq 0 \tag{5.74}$$

$$0 \leq \beta \leq M y_B. \tag{5.75}$$

Aromaticity, φ

The aromaticity is predicted by

$$\varphi \sum_{i \in G} n_i \lambda_i - \sum_{i \in G} n_i \xi_i = 0, \tag{5.76}$$

where λ_i is the number of non-hydrogen atoms in group i and ξ_i is the number of aromatic carbon atoms in group i.

Electronegative halogenicity, ψ

The electronegative halogenicity is given by

$$\psi \sum_{i \in G} n_i \lambda_i - \sum_{i \in G} n_i \varsigma_i = 0, \tag{5.77}$$

where ς_i is the number of electronegative halogen atoms in group i.

Refractive index, n_D

The refractive index is given by

$$n_D = \begin{cases} \frac{1}{7.26}((0.48872\delta)^{0.36} + 8.15) & \text{if } \sum_{i \in G_{n_D}} n_i \geq 1 \\ \frac{1}{14.95}(0.48872\delta + 13.47) & \text{otherwise,} \end{cases} \tag{5.78}$$

where G_{n_D} = { CH_3O, CH_2O, CH-O, CH_2Cl, CHCl. $CHCl_2$, I, Br, aCCl, aCF }
The binary variable y_{n_D} is introduced

$$y_{n_D} = \begin{cases} 1 & \text{if } \sum_{i \in G_{n_D}} n_i \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.79}$$

or equivalently,

$$y_{n_D} - \sum_{i \in G_{n_D}} \leq 0 \tag{5.80}$$

$$\frac{\sum_{i \in G_{n_D}} n_i}{\sum_{i \in G_{n_D}} n_i^U} - y_{n_D} \leq 0. \tag{5.81}$$

Four positive slack variables $s^+_{n_D,1}$, $s^-_{n_D,1}$, $s^+_{n_D,2}$ and $s^-_{n_D,2}$ are defined and the refractive index, n_D, is predicted by

$$s^+_{n_D,1} + s^-_{n_D,1} - M(1 - y_{n_D}) \leq 0 \tag{5.82}$$

$$s^+_{n_D,2} + s^-_{n_D,2} - My_{n_D} \leq 0 \tag{5.83}$$

$$n_D - \frac{1}{7.26}((0.48872\delta)^{0.36} + 8.15) + s^+_{n_D,1} - s^-_{n_D,1} = 0 \tag{5.84}$$

$$n_D - \frac{1}{14.95}(0.48872\delta + 13.47) + s^+_{n_D,2} - s^-_{n_D,2} = 0 \tag{5.85}$$

$$s^+_{n_D,1}, s^-_{n_D,1}, s^+_{n_D,2}, s^-_{n_D,2} \geq 0. \tag{5.86}$$

Enthalpy of vaporisation at 298 K, ΔH_{vap}

The enthalpy of vaporisation in units of J mol^{-1} is calculated using the approach of Marrero and Gani [2001]

$$\Delta H_{vap} = 1000 \left(\sum_{i \in G} n_i H_{vap,i} + 11.733 \right), \tag{5.87}$$

where $H_{vap,i}$ is theheat of vaporisation contribution of group i in kJ mol^{-1}.

Liquid molar volume at 298 K , V_m

The liquid molar volume in units of cm^3 mol^{-1} is predicted with the approach proposed by Constantinou et al. [1995] and V_m is the molar volume

$$V_m = 1000 \left(\sum_{i \in G} n_i V_{m,i} + 0.01211 \right), \tag{5.88}$$

where $V_{m,i}$ is the liquid molar volume contribution of group i in m^3 kmol^{-1}.

Hildebrand solubility parameter, δ_H

The Hildebrand solubility parameter, δ_H, in MPa$^{0.5}$ is calculated as follows:

$$\delta_H = \left[\frac{\Delta H_{vap} - RT}{V_m} \right]^{0.5}, \tag{5.89}$$

where ΔH_{vap} is in J mol^{-1}, V_m is in cm^3 mol^{-1}, R is the gas constant in J mol^{-1} K^{-1} and T is 298 K. Equation (5.89) has been linearised by Folić et al. [2006].

Macroscopic surface tension, γ

The macroscopic surface tension, γ, in calories mol^{-1} Å^{-2} is given as a function of the Hildebrand solubility parameter, δ_H and the molar volume, V_m

$$\gamma = \begin{cases} 1.43932 \times 0.01707\delta^2 V_m^{1/3} & \text{if } n_{\text{OH}} + n_{\text{COOH}} + n_{\text{aCOH}} = 0 \\ 1.43932 \times 0.0068\delta^2 V_m^{0.45} & \text{otherwise.} \end{cases} \tag{5.90}$$

The binary variable, y_γ, is introduced:

$$y_\gamma = \begin{cases} 1 & \text{if } n_{\text{OH}} + n_{\text{COOH}} + n_{\text{aCOH}} \geq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.91}$$

which algebraically it can be expressed:

$$\frac{n_{OH} + n_{\text{COOH}} + n_{\text{aCOH}}}{N_{max}} - y_\gamma \leq 0 \tag{5.92}$$

$$(y_\gamma - 1)\frac{n_{\text{OH}} + n_{\text{COOH}} + n_{\text{aCOH}} - 0.5}{N_{max}} \leq 0. \tag{5.93}$$

The macroscopic surface tension is calculated with the use of the slack variables $s^+_{n_\gamma,1}$, $s^-_{n_\gamma,1}$, $s^+_{n_\gamma,2}$ and $s^-_{n_\gamma,2}$ as follows

$$s^+_{n_\gamma,1} + s^-_{n_\gamma,1} - M(1 - y_{n_\gamma}) \leq 0 \tag{5.94}$$

$$s^+_{n_\gamma,2} + s^-_{n_\gamma,2} - My_{n_\gamma} \leq 0 \tag{5.95}$$

$$\gamma - 1.43932 \times 0.01707\delta^2 V_m^{1/3} + s^+_{n_\gamma,1} - s^-_{n_\gamma,1} = 0 \tag{5.96}$$

$$\gamma - 1.43932 \times 0.0068\delta^2 V_m^{0.45} + s^+_{n_\gamma,2} - s^-_{n_\gamma,2} = 0. \tag{5.97}$$

Melting and boiling temperatures

The first-order group-contribution methods proposed by Marrero and Gani [2001] are used for the prediction of dimensionless equivalent melting and boiling temperatures

$$T_{m,e} = \exp\left(\frac{T_m}{T_{m,0}}\right) \sum_{i \in G} n_i T_{m_i}, \tag{5.98}$$

where T_m is the melting point temperature, $T_{m,0}$ is the a reference melting temperature (147.450 K) and T_{m_i} is the contribution of group i to $T_{m,e}$.

Similarly for $T_{b,e}$

$$T_{b,e} = \sum_{i \in G} n_i T_{b_i}, \tag{5.99}$$

where T_b is the boiling point temperature, $T_{b,0}$ is the a reference melting temperature (222.543 K) and T_{b_i} is the contribution of group i to $T_{b,e}$.

5.2.4 Design constraints

The designed solvents, in order to be suitable, need to be in the liquid phase at the reaction temperature. This limitation can be expressed in terms of the melting and the boiling temperatures; an upper bound of 317 K is set for the melting point and a lower bound of 292 K is set for the boiling point. Therefore, the dimensionless equivalent melting point temperature constraint is

$$T_{m,e} \leq 8.6, \tag{5.100}$$

and the dimensionless equivalent boiling point temperature constraint is

$$T_{b,e} \geq 3.7. \tag{5.101}$$

5.2.5 Chemical feasibility constraints

When designing a solvent, chemical feasibility constraints are required to ensure solvent candidates are feasible. Chemical feasibility constraints of the form $h_2(\mathbf{n}, \mathbf{y}) = \mathbf{0}$ and $g_2(\mathbf{n}, \mathbf{y}) = \mathbf{0}$, also used by Struebing et al. [2010, 2013], are presented in this section.

The design of three types of molecular structures are considered using structural groups; acyclic, bicyclic and monocyclic molecular structures. These types are de-

noted by the binary variables, y_1, y_2 and y_3

$$y_1 = \begin{cases} 1 & \text{if the molecule is acyclic} \\ 0 & \text{otherwise} \end{cases} \tag{5.102}$$

$$y_2 = \begin{cases} 1 & \text{if the molecule is bicyclic} \\ 0 & \text{otherwise} \end{cases} \tag{5.103}$$

$$y_3 = \begin{cases} 1 & \text{if the molecule is monocyclic} \\ 0 & \text{otherwise.} \end{cases} \tag{5.104}$$

Then, it has to be ensured that only one type is designed at a time

$$y_1 + y_2 + y_3 = 1. \tag{5.105}$$

A continuous variable m, corresponding to the binary variables introduced above is given by Odele and Macchietto [1993]

$$m = \begin{cases} 1 & \text{for an acyclic molecule} \\ 0 & \text{for a monocyclic molecule} \\ -1 & \text{for a bicyclic molecule} \end{cases} \tag{5.106}$$

or equivalently

$$m - (y_1 - y_2) = 0. \tag{5.107}$$

The use of a continuous variable allows the type of molecule (either acyclic, monocyclic or bicyclic) designed to be represented by one variable, and its use will become apparent in other chemical feasibility constraints, such as the octet rule, equation (5.109), [Odele and Macchietto, 1993] and modified bonding rule, equation (5.110), [Buxton et al., 1999].

In this work a limited set of cyclic molecules in considered. Cyclic molecules are constrained to molecules containing one ring or two fused rings. A monocyclic molecule, denoted by $y_7 = 1$, consists of six aromatic structural groups and a bicyclic molecule, denoted by $y_6 = 1$, consists of ten aromatic structural groups. This is

expressed by the following constraint

$$\sum_{i \in G_A} n_i - 6y_3 - 10y_2 = 0, \tag{5.108}$$

where G_A is a subset of G that contains the aromatic groups (Table A.2).

The octet rule is implemented to ensure that a molecule is structurally feasible and that each valency in an structural group is satisfied with a covalent bond [Odele and Macchietto, 1993]

$$\sum_{i \in G} (2 - v_i)\, n_i - 2m = 0, \tag{5.109}$$

where v_i is the valency of group i (Table A.4).

The bonding rule modified by Buxton et al. [1999] is included to ensure that only one covalent bond is formed between two adjacent structural groups

$$n_j(v_j - 1) + 2m - \sum_{i \in G} n_i \leq 0, \qquad \forall j \in G. \tag{5.110}$$

Finally, the following constraint is used to limit variable n_i, the number of groups of type i in the molecule, to take only integer values

$$\sum_{k=1}^{K} 2^{k-1} y_{i,k} - n_i = 0, \quad \forall i \in G, \tag{5.111}$$

where $y_{i,k}$ are binary variables for groups i over the index of summation k and parameter K can be chosen to ensure that no more than a certain number of identical groups appear in the molecule. In this work $K = 3$, allowing a maximum of seven groups of type i in the same molecule.

5.2.6 Molecular complexity constraints

Molecular complexity constraints determine the size of the molecules, the type of functional groups and the bond types in the molecules. The molecular complexity constraints applied by Struebing et al. [2010, 2013] are also used here. However, single-group molecules introduced by Struebing et al. [2010] are not included in this formulation. In order to include them, group contributions for molar volume and enthalpy of vaporisation, required for the prediction of surface tension, refractive

index and dielectric constant, have to be calculated. This was not necessary in the CAMD formulation of Struebing et al. [2010], because the aforementioned solvent properties were not included in the optimisation problem. Therefore, for simplicity reasons single-group molecules are not considered here.

The size of the molecule is controlled by the following constraints

$$n_{G,min} - \sum_{i \in G} n_i \leq 0 \tag{5.112}$$

$$\sum_{i \in G} n_i \leq n_{G,max}, \tag{5.113}$$

where $n_{G,min}$ is the minimum number of groups allowed in the designed molecule and $n_{G,max}$ is the maximum number of groups. The values chosen for $n_{G,min}$ and $n_{G,max}$ are presented in Table 5.2.

Table 5.2: Minimum and maximum number of groups allowed in a designed molecular candidate.

Parameter	**Description**	**Value**
$n_{G,min}$	Minimum number of groups allowed in a molecule	2
$n_{G,max}$	Maximum number of groups allowed in a molecule	7

The number of groups of type i in a molecule is restricted by an upper bound n_i^U

$$n_i \leq n_i^U, \quad \forall i \in G \tag{5.114}$$

the values of n_i^U associated with n_i are presented in Table A.1.

Main groups should appear in acyclic molecules at most $n_{G,max}$ times and in monocyclic molecules at most twice

$$\sum_{i \in G_M} n_i \leq 2y_3 + n_{G,max} y_1, \tag{5.115}$$

where G_M is a set of main groups and a subset of G, consisting of groups that only contain C and H atoms (Table A.2).

Non-aromatic functional groups are limited to only monocyclic and acyclic molecules by

$$\sum_{i \in G_F} \frac{n_i}{n_i^U} \leq y_1 + y_3, \tag{5.116}$$

where G_F is a set of functional groups and a further subset of G, consisting of groups that contain atoms other than C and H atoms (Table A.2).

Only one carbon-carbon double bond can occur in the designed solvent

$$n_{\mathrm{CH_2{=}CH}} + n_{\mathrm{CH{=}CH}} + n_{\mathrm{CH_2{=}C}} + n_{\mathrm{CH{=}C}} + n_{\mathrm{C{=}C}} \leq 1. \tag{5.117}$$

In monocyclic aromatic molecules, side-chains that consist of at most two non-aromatic groups are allowed. For this purpose, three binary variables are introduced, y_{aC}, y_{aCCH} and $y_{\mathrm{aCCH_2}}$

$$y_{\mathrm{j}} = \begin{cases} 1 & \text{if group } j \text{ is present in the molecule} \\ 0 & \text{otherwise,} \end{cases} \tag{5.118}$$

where $j =$ aC, aCCH, $\mathrm{aCCH_2}$. These binary variables are determined by the following constraints. y_{aC} is determined by

$$n_{\mathrm{aC}} - 0.9 - M y_{\mathrm{aC}} \leq 0 \tag{5.119}$$

and

$$M(y_{\mathrm{aC}} - 1) - n_{\mathrm{aC}} + 1 \leq 0. \tag{5.120}$$

y_{aCCH} is determined by

$$n_{\mathrm{aCCH}} - 0.9 - M y_{\mathrm{aCCH}} \leq 0 \tag{5.121}$$

and

$$M(y_{\mathrm{aCCH}} - 1) - n_{\mathrm{aCCH}} + 1 \leq 0. \tag{5.122}$$

$y_{\mathrm{aCCH_2}}$ is determined by

$$n_{\mathrm{aCCH_2}} - 0.9 - M y_{\mathrm{aCCH_2}} \leq 0 \tag{5.123}$$

and

$$M(y_{\mathrm{aCCH_2}} - 1) - n_{\mathrm{aCCH_2}} + 1 \leq 0. \tag{5.124}$$

A new binary variable y_M is introduced for monocyclic molecules so that

$$y_M = \begin{cases} 1 & \text{if } y_{\mathrm{aC}} + y_3 = 2 \\ 0 & \text{otherwise.} \end{cases} \tag{5.125}$$

y_M is obtained by two constraints:

$$y_3 + y_{\mathrm{aC}} - 1 - y_M \leq 0 \tag{5.126}$$

and

$$M(y_M - 1) - y_3 - y_{\mathrm{aC}} + 2 \leq 0. \tag{5.127}$$

The aC group can appear at most once in monocyclic and twice in bicyclic molecules

$$2y_6 + y_M - n_{\mathrm{aC}} = 0. \tag{5.128}$$

The complexity of the molecules designed is limited by allowing at most one aC, aCCH or $\mathrm{aCCH_2}$ group to appear in a monocyclic molecule by

$$y_M + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}} \leq 1. \tag{5.129}$$

Two types of groups in side chains are defined; the chain-ending groups in set G_{CE} and non-chain-ending groups in set G_{NCE} (Table A.2). Non-chain-ending groups are attached directly to the aromatic molecule, provided a chain-ending group is attached afterwards, whereas a chain-ending group can be attached directly to the aromatic group or to a non-chain-ending group. An aCCH group forms two side chains, of which one is restricted to a $\mathrm{CH_3}$ group by

$$y_{\mathrm{aCCH}} \leq n_{\mathrm{CH_3}}. \tag{5.130}$$

Chain-ending groups are limited to appear at most three times in aliphatic molecules and once in aromatic molecules with active y_M, y_{aCCH} or $y_{\mathrm{aCCH_2}}$ binary variables by

$$\sum_{i \in G_{CE}} n_i \leq 3y_5 + y_M + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}}, \tag{5.131}$$

whereas non-chain-ending groups can only appear up to three times in aliphatic molecules and once in aromatic molecules with active y_M or $y_{\mathrm{aCCH_2}}$ binary variables by

$$\sum_{i \in G_{NCE}} n_i \leq 3y_5 + y_M + y_{\mathrm{aCCH_2}}. \tag{5.132}$$

The overall MINLP problem described by equations 5.1 - 5.9 consists of 300 variables, of which 173 are binary, and 1004 constraints, of which 19 are non-convex. The objective function, that is the rate constant, is a black box function, as it requires QM calculations. Because of these QM calculations, the resulting problem is a bilevel optimisation problem (equations (5.13) and (5.14)). Therefore, if one considers the size and complexity of the problem as well as the multiple minima encountered, it is not a trivial problem to solve.

5.3 Fixed geometry assumption

In order to simplify the problem and limit the computational cost, the geometry of all species is assumed to remain constant when moving from the gas phase to the liquid phase. Thus, in all QM calculations during the MINLP optimisation the optimised geometries in the gas phase are used for reactants and transition state structure, and only single-point energy calculations are performed. For constant geometry $\mathbf{r}_i = \mathbf{r}_i^{*,IG}$, where $\mathbf{r}_i^{*,IG}$ are the optimised geometries in the gas phase, the Gibbs free energy of solvation depends only on the solvent properties. In this way the problem from bilevel turns into a single level problem, and equations (5.13) and (5.14) become

$$\underline{E}_i^{ele,L}(\mathbf{r}_i^*, \varepsilon, \alpha) = \underline{E}_i^{ele,L}(\mathbf{r}_i^{*,IG}, \varepsilon, \alpha), \tag{5.133}$$

and

$$\mathbf{r}_i^* = \mathbf{r}_i^{*,IG} = \arg\min_{r_i} \underline{E}_i^{ele,IG}(\mathbf{r}_i, \varepsilon, \alpha), \quad i = TS, A, B. \tag{5.134}$$

In the rest of this section the derivatives of the solvation free energy with respect to the solvent properties are discussed, as they are needed in the implementation of the QM-CAMD algorithm (see section 5.4).

5.3.1 Derivatives of the free energy of solvation

The electronic part of the solvation free energy, $\underline{E}^{ele,L}$, is a function of the solute geometry $\mathbf{r}^*$, the dielectric constant of the solvent ε, and the hydrogen bond acidity α (equation (5.12)). Constant geometry is assumed, so $\underline{E}^{ele,L}$ depends only on ε and α. Since there are no analytical expressions connecting the electronic energy with the solvent properties, the derivatives are calculated numerically with central differences

$$\frac{\partial \underline{E}^{ele,L}(p)}{\partial p} = \frac{\underline{E}^{ele,L}(p+\frac{h}{2}) - \underline{E}^{ele,L}(p-\frac{h}{2})}{h} \tag{5.135}$$

where $p = \varepsilon, \alpha$ and h is a small number, which, in order to be reasonable for all magnitudes, is calculated as follows

$$h = \begin{cases} 0.002 & \text{if } |p| < 10^{-15} \\ 0.002|p| & \text{otherwise.} \end{cases} \tag{5.136}$$

The $\underline{G}^{CDS,L}$ term of the Gibbs free energy of solvation is a function of six solvent properties; hydrogen bond acidity α, hydrogen bond basicity β, macroscopic surface tension γ, refractive index n_D, aromaticity φ, and halogenicity ψ (equation (5.15)). The analytical expressions are given from equations (4.77)- (4.80)

$$\underline{G}_i^{CDS,L} = \sum_{j=1}^{N_A} \sigma_j A_j + \sigma^{[M]} \sum_{j=1}^{N_A} A_j$$

where

$$\sigma_j = \tilde{\sigma}_{z_j} + \sum_{j'=1}^{N_A} \tilde{\sigma}_{z_j z_{j'}} T_j$$

$$\tilde{\sigma}_\theta = \tilde{\sigma}_\theta^{[n_D]} n_D + \tilde{\sigma}_\theta^{[\alpha]} \alpha + \tilde{\sigma}_\theta^{[\beta]} \beta$$

and

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]}(\gamma/\gamma_0) + \tilde{\sigma}^{[\varphi^2]}\varphi^2 + \tilde{\sigma}^{[\psi^2]}\psi^2 + \tilde{\sigma}^{[\beta^2]}\beta^2.$$

The derivatives of the $\underline{G}^{CDS,L}$ term were also derived by Sheldon et al. [2006] for the SM5.42R model [Zhu et al., 1998], another model of the SM_x family. The derivatives of $\underline{G}^{CDS,L}$ over the solvent properties of β, γ, n_D, φ and ψ are calculated analytically. For hydrogen bond acidity α an analytical expression can also be derived, but it is not used here as the electrostatic term of the energy, $\underline{E}^{ele,L}$, depends on it as well thus equation (5.135) is applied. The first derivative of $\underline{G}^{CDS,L}$ with respect to β is a linear function of β

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\sum_{j=1}^{N_A}\sigma_j A_j + \sigma^{[M]}\sum_{j=1}^{N_A} A_j\right) \tag{5.137}$$

$$= \sum_{j=1}^{N_A}\left(\sigma_j^{[\beta]} + \sum_{j=1}^{N_A}\sigma_{jj'}^{[\beta]} T_{jj'}\right) A_j + 2\tilde{\sigma}^{[\beta^2]}\beta\sum_{j=1}^{N_A} A_j. \tag{5.138}$$

$T_{jj'}$ is calculated from the following equation [Marenich et al., 2009]

$$T_{jj'} = \begin{cases} \exp\left(\frac{\Delta r_{Z_j Z_{j'}}}{R_{jj'} - \Delta r_{Z_j Z_{j'}} - r_{Z_j Z_{j'}}}\right) & \text{if } R_{jj'} < \Delta r_{Z_j Z_{j'}} + r_{Z_j Z_{j'}} \\ 0 & \text{otherwise,} \end{cases} \tag{5.139}$$

where $R_{jj'}$ is the interatomic distance between atoms j and j' and $r_{Z_j Z_{j'}}$, $\Delta r_{Z_j Z_{j'}}$ are atomic number specific parameters that can be found in the supporting information (Part 3) in Marenich et al. [2009].

The first derivative of $\underline{G}^{CDS,L}$ with respect to n_D is constant

$$\frac{\partial \underline{G}^{CDS,L}}{\partial n_D} = \frac{\partial}{\partial n_D}\left(\sum_{j=1}^{N_A}\sigma_j A_j\right) \tag{5.140}$$

$$= \sum_{j=1}^{N_A}\left(\sigma_j^{[n_D]} + \sum_{j'=1}^{N_A}\sigma_{jj'}^{[n_D]} T_{jj'}\right) A_j. \tag{5.141}$$

Similarly, the first derivative of $G^{CDS,L}$ with respect to γ is also constant

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \gamma} = \frac{\partial}{\partial \gamma}\left(\sigma^{[M]}\sum_{j=1}^{N_A} A_j\right) \tag{5.142}$$

$$= \tilde{\sigma}^{[\gamma]}\sum_{j=1}^{N_A} A_j. \tag{5.143}$$

Finally, the first derivatives of $G^{CDS,L}$ with respect to φ and ψ are linear functions of φ and ψ, respectively

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \varphi} = \frac{\partial}{\partial \varphi}\left(\sigma^{[M]} \sum_{j=1}^{N_A} A_j\right) \tag{5.144}$$

$$= 2\tilde{\sigma}^{[\varphi^2]}\varphi \sum_{j=1}^{N_A} A_j \tag{5.145}$$

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \psi} = \frac{\partial}{\partial \psi}\left(\sigma^{[M]} \sum_{j=1}^{N_A} A_j\right) \tag{5.146}$$

$$= 2\tilde{\sigma}^{[\psi^2]}\psi \sum_{j=1}^{N_A} A_j. \tag{5.147}$$

The solvent accessible surface areas A_j for atoms j are given in the Gaussian output, whereas the atomic surface tension parameters $\tilde{\sigma}$ can be found in Marenich et al. [2009].

5.4 The proposed algorithm

The CAMD framework is integrated in the SMIN-αBB algorithm developed by Adjiman et al. [1997, 2000]. SMIN-αBB is a deterministic global branch-and-bound optimisation algorithm for nonconvex MINLP problems with general nonconvexities in the continuous variables and linear and bi-linear participation of the binary variables. It is based on the αBB global optimisation algorithm for twice continuously differentiable nonlinear programming (NLP) problems [Androulakis et al., 1995, Adjiman et al., 1998]. As discussed before, the solvent design problem is a nonconvex problem with a black box objective function and thus its solution, and especially the global solution, is indeed challenging.

The integrated algorithm is summarised in figure 5.1. Initially quantum mechanical calculations with geometry optimisation in the gas phase are performed for the reactants and the transition state structure, in order to obtain the geometries, energies and partition functions. Then a solvent is chosen to initialise the optimisation.

Note that this is not obligatory, as the program can initialise with a random initial guess, however, a good starting point is important for a fast convergence.

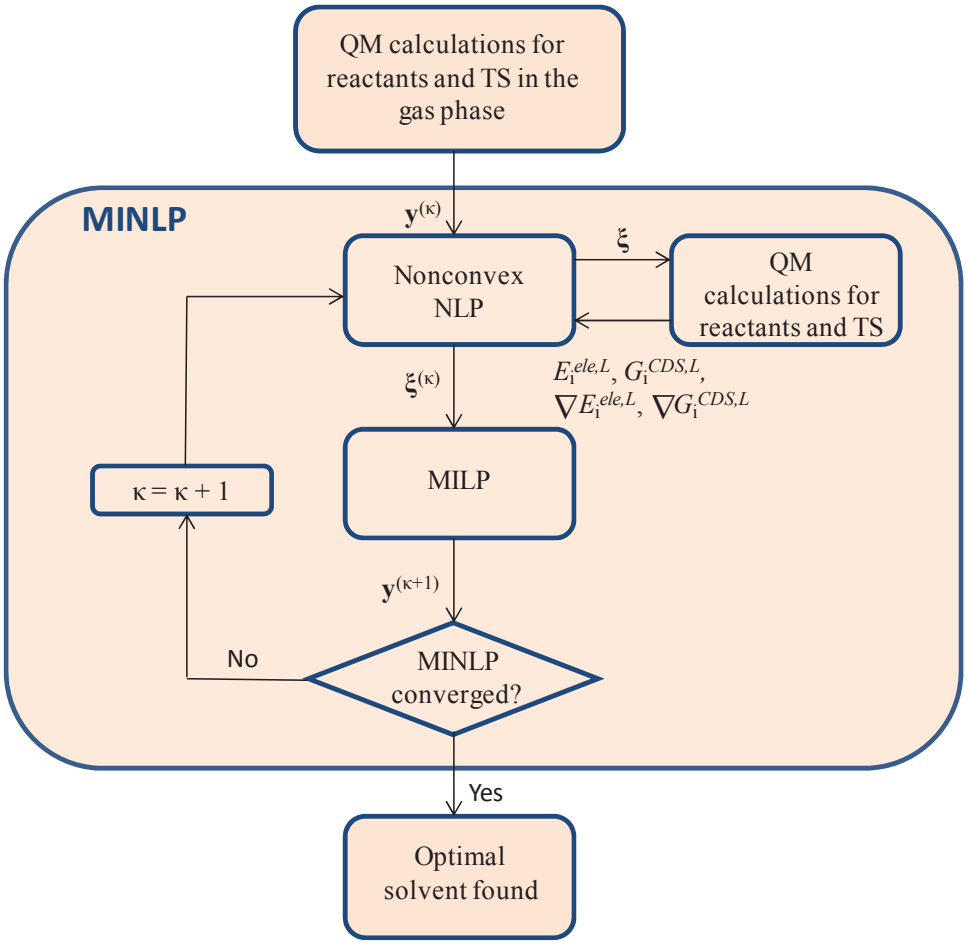


Figure 5.1: The full QM-CAMD solvent design algorithm.

SMIN-αBB applies an outer approximation algorithm [Duran and Grossmann, 1986], that decomposes the MINLP problem into two sub-problems: a NLP primal problem and a MILP master problem. Firstly, the binary variables ($\mathbf{y}$ in figure 5.1) are fixed to a $\mathbf{y}^{(\kappa)}$ particular combination (i.e. the solvent groups are set) and the resulting NLP problem is solved. The solution of the NLP provides the first set of continuous variables, that is the solvent properties $\boldsymbol{\xi}^{(\kappa)}$. For these solvent properties the MILP problem is solved to provide the second set of binary variables $\mathbf{y}^{(\kappa+\mathbf{1})}$, that is the second candidate solvent, and the algorithm iterates until the solutions of the NLP and the MILP problems converge.

During the NLP solution, calculations of objective function and the objective function's derivatives are required. The objective function is treated as a black-box function and is calculated with on-the-fly QM calculations. All terms in the expression of the rate constant (equation (5.11) other than the solvation free energy, $\Delta^{\ddagger}\underline{E}^{ele,L}$ + $\Delta^{\ddagger}\underline{G}^{CDS,L}$, are independent of the solvent, thus constant. The derivatives of the solvation free energy have been presented in the previous section. QM calculations required for the derivatives are also performed on-the-fly.

When the algorithm converges, the optimal solvent has been designed. To summarise, the required input to the algorithm are therefore:

- the optimised geometries of the reactants and transition state in the gas phase
- the derivatives of the non-electrostatic term of the solvation free energy, G^{CDS}, which can be calculated analytically as was shown in the previous section
- a solvent to initialise the MINLP.

5.5 Implementation

The QM calculations required for the calculation of the rate constant, that is, the initial calculations in the gas phase and the calculations in the liquid phase during optimisation, are performed using Gaussian 09 [Frisch et al., 2009] software. The optimisation problem is implemented within the global optimisation algorithm SMIN-αBB developed by Adjiman et al. [1997, 2000]. When a calculation of the objective function is required Gaussian is called automatically from αBB. Since the expression of the objective function is a black box, it is not possible for the αBB pre-processor to obtain analytical expressions for its derivatives via automatic differentiation of intrinsic functions as it would do in the case of explicit functions given. Thus, the derivatives of the objective function are provided in the implementation. When numerical derivatives of the objective function are required during optimisation Gaussian is automatically called from αBB. In every Gaussian call, the energies of the reactants and transition structure are calculated and the rate constant is returned

to the program. In order to realise the aforementioned calculations, two intrinsic functions were included in the SMIN-αBB implementation; one to call Gaussian 09 and return the energy value for the objective function, and one for calculation of the first order derivatives of the objective function. In addition, a number of scripts is created to make the connection between αBB and Gaussian and transfer the required information.

The interatomic distance $R_{jj'}$ required in the calculation of $T_{jj'}$ (equation (5.139)) in the non-electrostatic term $\underline{G}^{CDS,L}$ can be found in the Gaussian 09 output. The SASA, A_j, can be also found in the Gaussian 09 output when using the keyword "IOp(3/33=5)" for the debug print.

5.6 Application to a Menschutkin reaction

The solvent design methodology is applied to the Menschutkin reaction of phenacyl bromide with pyridine (figure 4.3). The analytical derivatives of $G^{CDS,L}$ are calculated for this reaction and the results of the application, as well as the limitations of the approach, are discussed.

5.6.1 Derivatives of $G^{CDS,L}$

The SMD atomic surface parameters $\tilde{\sigma}$ for the atoms in the reactants are given in Tables 5.3 and 5.4. The atomic number specific parameters $r_{Z_j Z_{j'}}$ and $\Delta r_{Z_j Z_{j'}}$, required for the calculation of $T_{jj'}$ (equation (5.139)), used here are given in Table 5.5.

The resulting first derivatives of $\underline{G}^{CDS,L}$ in equations (5.137) - (5.146) are

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \beta} = -2151.69 \text{ J mol}^{-1} \tag{5.148}$$

$$\frac{\partial \underline{G}^{CDS,L}}{\partial n_D} = 7173.76 \text{ J mol}^{-1} \tag{5.149}$$

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \gamma} = -14.24 \text{ J mol}^{-1} \tag{5.150}$$

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \varphi} = 341.01 \cdot \varphi \text{ J mol}^{-1} \tag{5.151}$$

$$\frac{\partial \underline{G}^{CDS,L}}{\partial \psi} = 543.66 \cdot \psi \text{ J mol}^{-1} \tag{5.152}$$

θ	$\tilde{\sigma}_\theta^{[n_D]}$	$\tilde{\sigma}_\theta^{[\alpha]}$	$\tilde{\sigma}_\theta^{[\beta]}$
H			
C	58.10	48.10	32.87
H, C	-36.37		
C, C	-62.05		
O	-17.56	193.06	-43.79
H, O	-19.39		
O, C	-15.70	95.99	
O, O			-128.16
N	32.62		
C, N	-99.76	152.20	
N, C		-41.00	
O, N			79.13
Br	-35.42		

Table 5.3: Atomic surface tension parameters (in units of cal mol^{-1} Å^{-2}), that depend on the atomic numbers, for the SMD model. Blank entries denote a zero value for the parameter. Any combinations that do not appear in the Table are set equal to zero.

$\tilde{\sigma}^{[\gamma]}$	0.35
$\tilde{\sigma}^{[\varphi^2]}$	-4.19
$\tilde{\sigma}^{[\psi^2]}$	-6.68
$\tilde{\sigma}^{[\beta^2]}$	0.00

Table 5.4: Atomic surface tension parameters (in units of cal mol^{-1} Å^{-2}), that do not depend on the atomic numbers, for the SMD model.

jj'	$r_{jj'}$	$\Delta r_{jj'}$
H, C	1.55	0.3
H, O	1.55	0.3
C, H	1.55	0.3
C, C	1.84	0.3
C, N	1.84	0.3
C, O	1.84	0.3
C, Br	2.3	0.3
N, C	1.84	0.3
O, C	1.33	0.1
O, N	1.5	0.3
O, O	1.8	0.3

Table 5.5: Values of parameters $r_{jj'}$ and $\Delta r_{jj'}$ (in units of Å), used in equation (5.139). Any combinations that do not appear in the table are set equal to zero.

5.6.2 Local solution for the Menschutkin reaction

Validity of fixed geometry assumption

The assumption of fixed geometry for reactants and transition state structure is initially validated. The rate constants of the Menschutkin reaction in five organic solvents, for which experimental values exist [Ganase et al., 2013], are calculated by

performing single-point energy calculations at the B3LYP/6-31+G(d) level of theory using the gas-phase geometry. The organic solvents tested are toluene, chloroform, tetrahydrofuran, acetone and acetonitrile, which were also studied in section 4.6. The predicted rate constants are shown in figure 5.2. There are deviations between predicted and experimental values and the percent average absolute deviation (% AAD) is equal to 241.5. However, the general trend is captured. Naturally, since the geometries of the species are not optimal, the resulting rate constants were not expected to agree quantitatively with experimental values. However, here that only the rate constant of a single reaction is considered as the performance measure, it is more important to capture the ranking of candidate solvents, rather than predict the exact values of the rate constants. Thus, a qualitative agreement is satisfactory.

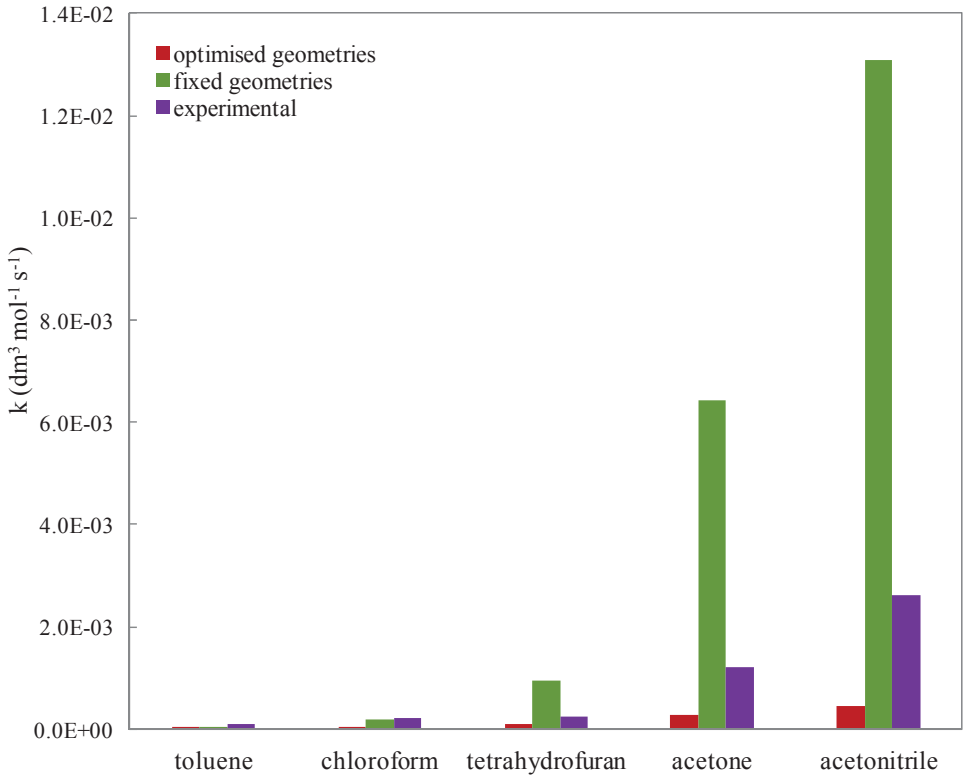


Figure 5.2: Predicted rate constants for the Menschutkin reaction of phenacyl bromide and pyridine in different organic solvents with single point energy calculations (fixed geometries) and geometry optimisations. All calculations are performed at the B3LYP/6-31+G(d). Experimental values are from Ganase et al. [2013].

Input to the algorithm

Quantum mechanical calculations with geometry optimisation of the reactants and the transition state structure in the gas phase are performed at the B3LYP/6-31+G(d) level of theory. The energies of the reactants and the TS, the imaginary frequency and partition functions in vacuum are shown in Table 5.6. For the frequency calculations a scale factor equal to 0.9614 is used, following Foresman and Frisch [1996]. The first order derivatives are provided from equations (5.148) - (5.152). The program is initialised with acetic acid, for which the rate constant obtained with single point energy calculation is $\bar{k}^{TST} = 1.763 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}.

	$\nu^{\ddagger}$	$\underline{E}^{ele,IG}$	$ZPVE$	$q^{'},^{IG}$
	[cm^{-1}]	[a.u. Particle^{-1}]	[J mol^{-1}]	-
Transition state	-424.424	-3204.29672	555490.3	0.14805E-76
Phenacyl bromide	-	-2956.02967	326730.9	0.36359E-40
Pyridine	-	-248.29579	224334.9	0.21541E-26

Table 5.6: Predicted values for the imaginary frequency $\nu^{\ddagger}$, the electronic energy $\underline{E}^{ele,IG}$, the zero-point vibrational energy $ZPVE$, and the total partition function $q^{'},^{IG}$, in the gas phase for the Menschutkin reaction of phenacyl bromide and pyridine. All calculations are performed at the B3LYP/6-31+G(d) level of theory.

Progress of the algorithm

Since this is a local solution, SMIN-αBB algorithm becomes the same as the outer approximation (OA) algorithm by Duran and Grossmann [1986]. The algorithm converges after two OA iterations, thus one NLP solution and one MILP solution are required for each OA iteration. SNOPT [Gill et al., 2005] and CPLEX [IBM] solvers are used for the NLP and MILP solutions, respectively. Details of the two NLP solutions are given in Table 5.7. The objective function is calculated 12 times in the first OA iteration (11 times for the NLP and once for the MILP) and 25 times for the second OA iteration (24 times for the NLP and once for the MILP). Every time the

objective function is called, its derivatives are also required. Thus, for every calculation of the objective function 5 × 3 Gaussian calls are needed: (1 call for the objective function + 2 calls for the derivative with respect to α + 2 calls for the derivative with respect to ε) × (reactant A + reactant B + TS). The derivatives with respect to the other solvent properties are analytical. Therefore, a total of 185 Gaussian calls are required during optimisation (37 objective function calls × 5).

	1^{st} NLP	2^{nd} NLP
Major iterations	6	8
Minor iterations	380	406
Objective function calls	11	24

Table 5.7: Details of the NLP solutions from SNOPT of the two outer approximation iterations during the local solution for the Menschutkin reaction of phenacyl bromide and pyridine.

The CPU time required for this local run of the MINLP problem is approximately 3 days in total; 15 hours per processor on a six-processor (Intel(R) Xeon(R) X5690, 3.47GHz) node. It is important to mention that the CPU time for a single-point calculation is a few minutes (1 - 5 minutes for this reaction, depending on the species), while the CPU time for a geometry optimisation can vary from minutes to days, and more commonly is in the order of hours for this reaction. Therefore, if geometry optimisations were performed instead of single-point calculations, the local solution would be computationally very expensive if not prohibitive.

Designed solvent

The local solution found is: 5 × ACH, AC, CHO, that is benzaldehyde. The reaction rate constant in benzaldehyde from QM is $\bar{k}^{TST} = 7.298 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1} with single point energy calculation. The energetics in benzaldehyde are given in Table 5.8. All calculations are performed at the B3LYP/6-31+G(d) level of theory using the SMD solvation model. The energies of the reactants and the transition state

structure for optimised geometries in benzaldehyde, with B3LYP/6-31+G(d) and SMD, are presented in Table 5.9. The rate constant in benzaldehyde with optimised geometries is $k^{TST} = 3.104 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1} and the corresponding value for acetic acid is $k^{TST} = 1.763 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}. Therefore, a better solvent has been indeed found. Unfortunately, there is no experimental rate constant value for the reaction in benzaldehyde to compare with. Although a better solvent than acetic acid has been designed, this is not the best solvent possible, since the solution is local. The optimal solvent designed by Struebing et al. [2013] for the same reaction was nitromethane. Note that nitromethane is considered a single-group molecule and is not included in the current design space here. The rate constant in nitromethane is 6.085 $\cdot 10^{-4}$ dm^3 mol^{-1} s^{-1} using SMD at the B3LYP/6-31+G(d) level of theory with optimised geometries, that is almost double the rate constant in benzaldehyde. Thus, a more thorough investigation of the design space is needed.

	$\underline{E}^{ele,L} + \underline{G}^{CDS,L}$ [a.u. Particle^{-1}]	$\underline{G}^{CDS,L}$ [kcal mol^{-1}]
Transition state	-3204.34116	-6.83
Phenacyl bromide	-2956.04869	-5.19
Pyridine	-248.30697	-1.25

Table 5.8: Predicted values for the electronic part, $\underline{E}^{ele,L}$, and the CDS part, $\underline{G}^{CDS,L}$, of the free energy for the Menschutkin reaction of phenacyl bromide and pyridine in benzaldehyde with single point energy calculations. All calculations are performed at the B3LYP/6-31+G(d) level of theory.

5.6.3 Limitations

When solving the MINLP solvent design problem locally, promising candidate solvents are obtained. However, a limited area of the design space is taken into account, which depends on the solvent used as a starting point. Therefore, it would be ideal to solve the problem globally, investigating the whole design space and obtain the

	$\nu^{\ddagger}$ [cm^{-1}]	$\underline{E}^{ele,L} + \underline{G}^{CDS,L}$ [a.u. Particle^{-1}]	$\underline{G}^{CDS,L}$ [kcal mol^{-1}]
Transition state	-455.700	-3204.33859	-6.94
Phenacyl bromide	-	-2956.04907	-5.29
Pyridine	-	-248.30700	-1.24

Table 5.9: Predicted values for the imaginary frequency, $\nu^{\ddagger}$, the electronic part, $\underline{E}^{ele,L}$, and the CDS part, $\underline{G}^{CDS,L}$, of the free energy for the Menschutkin reaction of phenacyl bromide and pyridine in benzaldehyde. The geometries of all species are optimised both in the gas and the liquid phase. All calculations are performed at the B3LYP/6-31+G(d) level of theory.

global optimal solvent for a reaction. The methodology and its implementation are general and can be used in a global optimisation algorithm, although finding the global solution cannot be guaranteed because of the black box function. Nevertheless, the computational time required for a global solution is prohibitive. Particularly, the same MINLP problem has been used for a global test search with the SMIN-αBB algorithm, considering constant values of α for the underestimators [Adjiman et al., 1998]. According to Adjiman et al. [1998], the α value has to satisfy the following condition

$$\alpha \geq \{0, -\frac{1}{2}\lambda_{min}([\mathbf{H}_f])\}, \tag{5.153}$$

where $\lambda_{min}([\mathbf{H}_f])$ is the minimum eigenvalue of the interval Hessian matrix. The value of α indicates the nonconvexity of the function, that is the higher the α value, the more nonconvex the function. The values need to be chosen so as not to be too large to increase the convergence time, but either too small not to underestimate the function. Therefore, the values of 5 and 10 have been used in the test case. Constant values of α in the same order of magnitude have been used also in other works, e.g. by Klepeis and Floudas [1999] and Klepeis et al. [2003]. The αBB algorithm [Adjiman et al., 1998] applies a branch-and-bound method [Floudas, 1995] for global optimisation. The computational time required for only one node (one nonconvex MINLP) in the branch-

and bound tree is approximately a week, using a six-processor (Intel(R) Xeon(R) X5690, 3.47GHz) node. Thus, the presented full QM-CAMD approach for solvent design can be successfully applied to identify promising solvents in a small region of the design space, but in order to broaden the search, approximate methods that require low computational time need to be used to represent the kinetic model.

5.7 Conclusions

A new methodology for the design of solvents that maximise the performance of a reaction has been presented. The solvent design problem is formulated as a computer-aided molecular design problem where a performance measure (objective function) is maximised, subject to structure and property constraints, chemical feasibility and molecular complexity constraints, and design constraints. The performance measure considered here is the reaction rate constant, which is calculated with on-the-fly quantum mechanical calculations. The integration of quantum mechanical calculations in the CAMD framework results to a bilevel MINLP formulation with black-box functions. The problem is simplified and turned into a single-level optimisation problem by assuming fixed geometry when moving from the gas phase to the liquid phase. The derivatives of the free energy of solvation are required during the solution of the MINLP and their derivation for fixed geometry has been presented.

The proposed approach has been implemented within the global optimisation algorithm SMIN-αBB and applied to a Menschutkin reaction providing a promising solvent. The main limitation of the approach is the fixed geometry assumption which introduces significant uncertainty in the design, as well as the fact that it is constrained to local solutions, due to its high computational cost. A possible solution to this problem is the use of surrogate models that would allow the exploration of a larger area of the design space at reduced computational cost. Such an approach is explored in the next chapter.

Chapter 6

QM-Kriging CAMD methodology for reactions

6.1 Introduction

In the previous chapter it was shown that a full-QM approach to solvent design is rather limiting due to high computational requirements. Here, quantum mechanical calculations are coupled with a surrogate model to reduce the computational cost and thus make the solvent design more efficient. A QM-Kriging-based approach to computer-aided molecular design of solvents for reactions has been developed and it is presented in this chapter. The work is based on the methodologies proposed by Folić et al. [2007] and Struebing et al. [2010] for the design of solvents that maximise the reaction rate constant. In a series of papers Folić et al. [2004, 2005, 2006, 2007, 2008a] developed an iterative methodology where a small numbers of experimental rate constants in organic solvents is used to regress the solvatochromic equation [Kamlet and Taft, 1976], [Taft and Kamlet, 1976], which is then included in a CAMD framework to design a candidate solvent. In principle, the approach allowed for the use of experimental information by measuring the rate constant in the new candidate solvent identified and adding the measured value to the first set of solvents, rebuilding the solvatochromic equation, and so on until convergence criterion is satisfied. Struebing et al. [2010] further extended the methodology by using quantum-mechanically

derived rate constants for input, instead of experimental data, introducing the first *ab initio* method (no required experimental data of the reaction) for the design of solvents for reactions [Struebing et al., 2013]. The solvatochromic equation, as described in section 3.4.1, was used as a surrogate model, to reduce the computational cost of computer-aided molecular design. It is a linear free energy relationship that connects the reaction rate constant with a small number of solvent properties. However, in reality, the relation of the rate constant with the solvent properties is more complex than just linear [Buncel et al., 2003]. The underlying quantum mechanical model, used to generate rate constant values for the solvatochromic equation, is in fact nonlinear in several solvent properties. Therefore, a more sophisticated surrogate model would be able to capture this behaviour better. In this work, the solvent design approach of Struebing et al. [2013] is further developed by implementing Kriging, a response surface method, to derive the surrogate model.

In section 6.2 an introduction to surrogate models is given. Special emphasis is given to Kriging and the basic concepts are described in section 6.2.3. The proposed design methodology is then described in section 6.3, and applied to a Menschutkin reaction in section 6.6 and a Cope elimination reaction in section 6.7.

6.2 An introduction to surrogate models

Engineered and physical systems can be really complex, so the models that simulate them can be complex as well and thus challenging to solve and very time-consuming. In order to reduce the size of the problem and the computational cost, surrogate models are often used, not only for simulation but also for optimisation. The choice of a suitable surrogate and then the way to integrate it within the optimisation formulation are issues that have to be addressed when adopting a surrogate-based approach. There are two very interesting reviews on this matter by Jones [2001] and Forrester and Keane [2009]. In this section, some of the most widely used surrogates are briefly described, while special emphasis is given to Kriging which has been chosen for application in this work.

6.2.1 Polynomials

Polynomial response surface models (RSM) are still one of the most widely used type of surrogate models in engineering [Box and Draper, 1987, Forrester and Keane, 2009]. The method of RSM was first introduced by Box and Wilson [1951] and the basic concept is to use a small number of function evaluations (also referred to as observations) and a specific functional form in order to build the surrogate model and obtain an optimal response. A polynomial approximation of order m of a function $f(x)$ of 1 dimension can be written as

$$f(x) \approx \widehat{y}(x; m, \mathbf{a}) = \mathrm{a}_0 + \mathrm{a}_1 x + \mathrm{a}_2 x^2 + ... + \mathrm{a}_m x^m = \sum_{i=0}^{m} \mathrm{a}_i x^{(i)}. \tag{6.1}$$

The above equation can be written in a matrix form as $\mathbf{Ua} = \mathbf{y}$, where $\mathbf{y}$ is a vector of the real function values and U is the Vandermonde matrix

$$\mathbf{U} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{pmatrix}.$$

The vector of estimated polynomial regression coefficients $\mathbf{a}$, using least squares, is

$$\mathbf{a} = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{y}. \tag{6.2}$$

There are various methods to calculate the order of the polynomial, m, e.g. the null hypothesis method or by minimising the cross-validation error [Forrester and Keane, 2009]. Naturally, the higher the order of the polynomial, m, the better the accuracy of the polynomial model. However, a high order also implies high data and computational requirements in order to achieve a statistically significant model. Polynomial surrogate models are relatively simple and fast and this is why they are very popular. Nevertheless, in high-dimensional problems it may not be possible to obtain the data required to calculate the multiple terms in a high order polynomial, resulting only in low-order polynomials, that are not be very accurate.

More sophisticated versions of polynomial methods are the weighted least-squares [Aitken, 1935] and the moving least-squares [Aitken, 1935]. In the weighted least-squares method each observation is considered as having a different contribution to the estimation of the polynomial coefficients and weights are applied to the observations according their relative importance. The moving least-squares improves on the weighted least-squares method by varying the weights according to the distance between the point to be predicted and each observed data point. Since the weights are functions of distance, they have to be calculated at every prediction. Thus the method is more accurate but more computationally expensive.

6.2.2 Radial Basis Functions

Radial basis functions (RBF) methods [Broomhead and Loewe, 1988] are based on the interpolation of a set of data by using a linear combination of polynomials and "basis functions". Simple functions are used as "basis" in order to build complicated surfaces. If we assume $\mathbf{x}_j$, $j = 1, \ldots, n$ to be a vector of n sampled data points, each of dimension d (d is the number of degrees of freedom), and the values of the function at these points to be $y_j = y(\mathbf{x}_j)$, the predictor at a new point $\mathbf{x}^*$ will be of the form

$$\widehat{y}(\mathbf{x}^*) = \sum_{k=1}^{m} a_k \pi_k(\mathbf{x}^*) + \sum_{j=1}^{n} b_j \varphi(\mathbf{x}^* - \mathbf{x}_j) \tag{6.3}$$

where π_k are the polynomial terms and φ are the basis functions. The coefficients a_k and b_j are the weights, that is the parameters that need to be optimised.

Radial basis functions, $\varphi(\mathbf{z})$, where $\mathbf{z} = \mathbf{x}^* - \mathbf{x}_j$, can be either fixed or parametric. Examples of fixed basis functions are

$$\text{Linear: } \varphi(\mathbf{z}) = \|\mathbf{z}\| \tag{6.4}$$

$$\text{Cubic: } \varphi(\mathbf{z}) = \|\mathbf{z}\|^3 \tag{6.5}$$

$$\text{Thin plate spline: } \varphi(\mathbf{z}) = \|\mathbf{z}\|^3 \log(\|\mathbf{z}\|), \tag{6.6}$$

where $\|\mathbf{z}\|$ is the Euclidean norm. Parametric radial basis functions offer more gener-

ality and flexibility, but, naturally, they also introduce complexity. Examples include

$$\text{Gaussian: } \varphi(\mathbf{z}) = \exp(\frac{-\|\mathbf{z}\|^2}{2\gamma^2}) \tag{6.7}$$

$$\text{Multiquadratic: } \varphi(\mathbf{z}) = \sqrt{\|\mathbf{z}\|^2 + \gamma^2} \tag{6.8}$$

$$\text{Kriging: } \varphi(\mathbf{z}) = \exp\left(-\sum_{l=1}^{d} \theta_l |\mathbf{z}|^{p_l}\right), \tag{6.9}$$

where γ, θ, p are positive adjustable parameters.

In order to interpolate the data the following condition has to be fulfilled

$$y_i = \sum_{k=1}^{m} a_k \pi_k(\mathbf{x}_i) + \sum_{j=1}^{n} b_j \varphi(\mathbf{x}_i - \mathbf{x}_j), \quad \text{for } i = 1, \ldots, n. \tag{6.10}$$

The unknowns are the coefficients a_k and b_j, of which there are $n+m$. Thus m extra equations are required. It can be shown [Jones, 2001] that the following constraints should be added

$$\sum_{j=1}^{n} b_j \pi_k(\mathbf{x}_j) = 0, \quad \text{for } k = 1, \ldots, m. \tag{6.11}$$

Except for the case of Kriging, which is considered as a category of its own (see next section), the use of radial basis functions in the chemical engineering community is not common. Two recent examples of applications of radial basis function methods in modelling chemical processes are the work of Singh et al. [2013] on predicting the adsorptive removal of chlorophenol from aqueous solution with various process variables using five non-linear models, where RBF with Gaussian functions proved to give one of the two best predictions, and the work of Chang and Chen [2011], who applied a cubic basis function model in their optimisation methodology for the etherification of glycerol with tert-butyl alcohol.

6.2.3 Kriging

A more extensive description of Kriging is given in this section, as it is applied in this work. The description provided here corresponds to the so-called "ordinary" Kriging, the most popular type of Kriging. Kriging, originally developed in geostatistics, was introduced by Krige [1951] and further developed by Matheron [1963]. A very extensive review of the method and its applications can be found in Cressie [1993].

A common feature of all surrogate models is that assumptions are made concerning the nature of the landscape. Kriging is one of the least biased methods in terms of the range of surfaces that can emulate. The parameters in the basis function (equation (6.9)) may increase the complexity of the method, but they also increase its efficiency. An example of the good performance of Kriging is shown in figure 6.1, where it has been applied to predict the Branin function,

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} - 6 \right)^2 + 6(1 - \frac{1}{8\pi}) \cos(x_1) + 10, \tag{6.12}$$

with the use of 21 sample points. It can be seen that the prediction of Kriging is very good. Over most of the surface the absolute error varies between 0 and 2 (figures 6.1c,d).

The derivation of the Kriging method presented by Jones [2001] is adopted here. We suppose that we want to predict the value of a function $f(\mathbf{x})$ at some point $\mathbf{x}$. Before sampling any points of the surface, the prediction will be uncertain. We assume that this uncertainty is represented by a random variable $Y(\mathbf{x})$ that is normally distributed with mean value μ and variance σ^2. If we consider two points $\mathbf{x}_i$ and $\mathbf{x}_j$, with random variables $Y(\mathbf{x}_i)$ and $Y(\mathbf{x}_j)$, it is assumed that the correlation between these two random variables is of the form

$$\text{Corr}\left[Y(\mathbf{x}_i), Y(\mathbf{x}_j)\right] = \exp\left(- \sum_{l=1}^{d} \theta_l |\mathbf{x}_{il} - \mathbf{x}_{jl}|^{p_l} \right) \tag{6.13}$$

where d is the dimensionality of the problem, and θ_l, p_l are parameters that have to satisfy the constraints $\theta_l \geq 0$ and $0 < p_l \leq 2$ respectively. From the correlation function it can be seen that when $\mathbf{x}_i = \mathbf{x}_j$, the correlation is 1, while as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow \infty$, the correlation tends to zero. Parameters θ_l and p_l have a "physical" meaning. θ_l determines how much the lth variable affects the function; large values indicate highly active variables. p_l determines the smoothness of the function in the lth direction; values near 2 indicate smooth functions, while values near 0 indicate non-differentiable, rough functions.

The likelihood of the observed data can be given by the following expression

$$\frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp \frac{-(\mathbf{y} - \mathbf{1}\mu)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2} \tag{6.14}$$

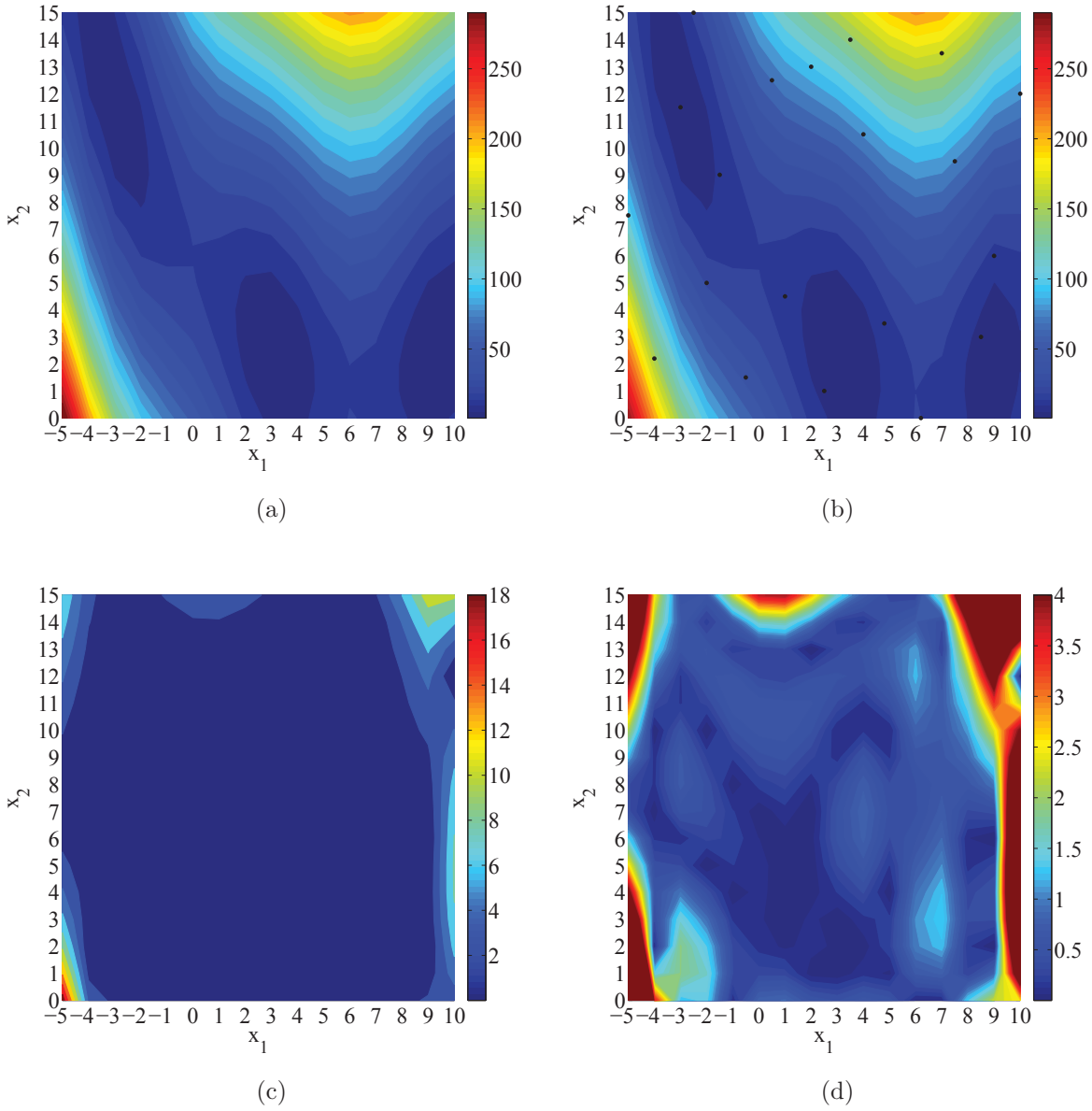


Figure 6.1: a: The real Branin function evaluated at 256 points. b: Representation of the Branin function with Kriging based on 21 sample points, shown as black dots. c: Absolute difference between the real Branin function and the Kriging prediction. d: Absolute difference between the real Branin function and the Kriging prediction, focusing only on absolute errors up to 4.

where n is the number of observations, $\mathbf{R}$ is a $n \times n$ matrix (referred to from now on as the correlation matrix) with (i, j) element given by equation (6.13), $\mathbf{1}$ is a vector of n elements, all equal to one, and μ, σ^2 are the mean and the variance of the predictions as mentioned previously. The vector $\mathbf{y}$ denotes the observed function values

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \tag{6.15}$$

Parameters μ, σ^2, θ_l and p_l are estimated by maximising the likelihood function (equation (6.14)) or more conveniently the log of the likelihood function, which, ignoring constant terms, is

$$-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log(|\mathbf{R}|) - \frac{(\mathbf{y} - \mathbf{1}\mu)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}. \tag{6.16}$$

Setting the derivatives of this function with respect to μ and σ^2 equal to zero, the optimal values of μ and σ^2, $\widehat{\mu}$ and $\widehat{\sigma}$ respectively, can be expressed as functions of the correlation matrix, $\mathbf{R}$

$$\widehat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \tag{6.17}$$

$$\widehat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\widehat{\mu})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\widehat{\mu})}{n}. \tag{6.18}$$

Substituting equations (6.17), (6.18) into equation (6.16), the so-called "concentrated log-likelihood" expression is derived. Again ignoring the constant terms, the concentrated log-likelihood is

$$-\frac{n}{2}\log(\widehat{\sigma}^2) - \frac{1}{2}\log(|\mathbf{R}|). \tag{6.19}$$

The concentrated log-likelihood function depends only on the correlation matrix, $\mathbf{R}$, or correspondingly on the correlation parameters θ_l and p_l. This is the function that is maximised with decision variables θ_l and p_l. Once the correlation parameters are known, $\widehat{\mu}$ and $\widehat{\sigma}^2$ are calculated from equations (6.17) and (6.18).

In order to make a prediction at some new point $\mathbf{x}^*$, the new point ($\mathbf{x}^*$, y^*) is added as a new observation to the observed data and the "augmented" likelihood function is calculated by using the parameters values from the maximisation of the

likelihood function. The Kriging predictor is then the value of y^* that maximises the "augmented" likelihood function. It can be shown [Jones, 2001] that the Kriging predictor, $\widehat{y}(\mathbf{x}^*)$ is given by the following expression

$$\widehat{y}(\mathbf{x}^*) = \widehat{\mu} + \mathbf{r}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\widehat{\mu}) \tag{6.20}$$

where $\mathbf{r}$ is the vector of correlations of $Y(\mathbf{x}^*)$ with $Y(\mathbf{x}_i)$, for $i = 1, \ldots, n$, and is of the following form

$$\mathbf{r} = \begin{pmatrix} \text{Corr}\left[Y(\mathbf{x}^*), Y(\mathbf{x}_1)\right] \\ \vdots \\ \text{Corr}\left[Y(\mathbf{x}^*), Y(\mathbf{x}_n)\right] \end{pmatrix} \tag{6.21}$$

or using equation (6.13)

$$\mathbf{r} = \begin{pmatrix} \exp\left(-\sum_{l=1}^{d} \theta_l |\mathbf{x}_l^* - \mathbf{x}_{1l}|^{p_l}\right) \\ \vdots \\ \exp\left(-\sum_{l=1}^{d} \theta_l |\mathbf{x}_l^* - \mathbf{x}_{nl}|^{p_l}\right) \end{pmatrix}. \tag{6.22}$$

Making the connection to the previous section, if $\varphi(\mathbf{z})$ is the Kriging basis function, then the i^{th} element of $\mathbf{r}$ is $\varphi(\mathbf{x}^* - \mathbf{x}_i)$. Defining also the i^{th} element of $\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\widehat{\mu})$ as b_i and $\widehat{\mu} = a$, then the Kriging predictor takes the form

$$\widehat{y}(\mathbf{x}^*) = a + \sum_{i=1}^{n} b_i \varphi(\mathbf{x}^* - \mathbf{x}_i), \tag{6.23}$$

where it can be seen that, as argued in the previous section, it is a linear function of polynomials (here only a constant, a) and basis functions.

An important advantage of Kriging is that it provides an estimate of the error of a prediction. The mean-squared error of the predictor is

$$s^2(\mathbf{x}^*) = \widehat{\sigma}^2 \left[1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} + \frac{(\mathbf{r}^T\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}\right]. \tag{6.24}$$

There are also other variations that can be found in the literature; universal, blind and co- Kriging. In universal Kriging [Cressie, 1993] the mean term is no longer constant, but a function of $\mathbf{x}$

$$\widehat{\mu} = \widehat{\mu}(\mathbf{x}) = \sum_{i=0}^{m} \mu_i \nu_i(\mathbf{x}) \tag{6.25}$$

where ν_i's are some known functions and μ_i's are unknown parameters. Although universal Kriging will give a better description of the function when the trends in the data are known (i.e. the ν_i's), it is rare that this information exists. Blind Kriging [Joseph et al., 2008] overcomes this limitation by identifying ν_i's from data-analytic procedures. The mean term now becomes

$$\widehat{\mu} = \widehat{\mu}(\mathbf{x}) = \boldsymbol{\nu}(\mathbf{x})^T \boldsymbol{\mu}_m \tag{6.26}$$

where $\boldsymbol{\nu}(\mathbf{x})^T = (1, \nu_1, \ldots, \nu_m)$, $\boldsymbol{\mu}_m = (\mu_0, \mu_1, \ldots, \mu_m)^T$ and m are unknown. The blind Kriging predictor is

$$\widehat{y}(\mathbf{x}^*) = \boldsymbol{\nu}(\mathbf{x})^T \widehat{\boldsymbol{\mu}}_m + \mathbf{r}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{V}_m \widehat{\boldsymbol{\mu}}_m) \tag{6.27}$$

where $\mathbf{V}_m = (\boldsymbol{\nu}(\mathbf{x}_1), \ldots, \boldsymbol{\nu}(\mathbf{x}_n))^T$ and $\widehat{\boldsymbol{\mu}}_m = (\mathbf{V}_m^T \mathbf{R}^{-1} \mathbf{V}_m)^{-1} \mathbf{V}_m^T \mathbf{R}^{-1} \mathbf{y}$. The challenge here is not only to apply a sophisticated model for the mean, but also to use the suitable variables in the development of the mean model, as unnecessary variables (i.e. variables that have little effect on the response surface) deteriorate the overall performance. Joseph et al. [2008] use a Bayesian forward selection technique for the calculation of $\boldsymbol{\nu}_i$ functions and show that with that technique suitable variables can be successfully identified, improving notably the prediction of blind Kriging over ordinary Kriging.

Couckuyt et al. [2012] implemented blind Kriging in Matlab and tested the method on different examples. They claim that its performance is as good as or better than that of ordinary Kriging, in the cases where a clear trend exists, i.e. the data cover existing non-linearities in the surface. If this is not the case, ordinary Kriging is more suitable. Furthermore, blind Kriging requires double the computational cost of ordinary Kriging.

There are cases in which "cheap" data can be acquired that may not be very accurate, but it can be combined with expensive, accurate data to provide a good prediction. This is the basic idea of multi-fidelity surrogate-based methods; a greater quantity of cheap data is combined with a small number of expensive data to improve the performance of the expensive method. Co-Kriging [Cressie, 1993], [Forrester

and Keane, 2009] is a multi-fidelity method where a "cheap" simulation method is combined with Kriging.

Kriging has recently attracted the attention of the chemical and process engineering community. Zerpa et al. [2005] combined field scale numerical simulation with multiple surrogates, including Kriging, to optimise alkaline-surfactant-polymer flooding processes. The optimal operating conditions obtained provided a 30% greater objective value than the mean value of the initial sample used to build the surrogates. Hawe et al. [2010] developed partial least squares (PLS) and Kriging models using electron density properties of neutral and protonated pyridines, in order to predict their basicities in the gas phase and in water. Kriging proved to give better predictions than PLS. Recent applications of Kriging in process engineering related topics include the modelling and optimisation of CO_2 capture technologies. Hasan et al. [2012] have simulated and optimised a pressure swing adsorption process and vacuum swing adsorption process by solving the resulting nonlinear algebraic and partial differential equation systems using a Kriging model. The use of Kriging, enabled the solution of the complex problem by commercial solvers. Another work on the design of pressure swing adsorption systems is that of Beck et al. [2012], where surrogate based optimisation (SBO) methods with Kriging are used to solve the model. The authors showed that, concerning both computational time and quality of solution, the performance of SBO methods with Kriging was outstanding compared to a genetic algorithm. Kriging has been also used for the solution of mixed-integer nonlinear programs that contain black-box functions with noisy and/or incomplete data. Davis and Ierapetritou [2007, 2009] developed a branch and bound method where Kriging is used to build global models for the black-box functions and a NLP subproblem objective at each node of the tree. The best Kriging solutions are further refined by optimisation of sequential response surfaces. The method was applied to two process synthesis problems and was shown to be successful at finding the global optimum 90 and 86% of the time. In similar context, Boukouvala et al. [2010] used Kriging to produce predictive models for noisy and incomplete data sets in unit process operations of a pharmaceutical table productions processes.

6.3 QM-CAMD-Kriging algorithm

A new methodology is proposed for the design of optimal solvents for reactions. The methodology combines a CAMD framework with QM calculations and the Kriging surface approximate model. The overall methodology is shown in figure 6.2.

Step 1. Definition of the initial set of solvents.

An initial set of solvents is chosen, preferably with diverse dielectric constants. This initial set of solvents will be used to build the first Kriging surface. Here, two options, consisting of ten and fourteen initial solvents, are investigated.

Step 2. Rate constant by QM for a specific number of solvent(s).

The rate constant of the reaction in the required solvent or solvents is calculated using conventional transition state theory and the SMD solvation model (the reader is referred to section 4.5.4 for details). Equation (4.81) is used for the QM calculations. The solvent properties required in the SMD model are calculated from the group-contribution methods by Sheldon et al. [2005] and Folić et al. [2007].

Step 3. Building the Kriging surface.

A Kriging surface is built based on the information obtained from Step 2. The use of a surrogate model allows to limit the number of QM calculations. The choice of the surrogate, however, is challenging as it affects two important aspects of the design; accuracy of results and convergence with the detailed model, that is, agreement between the predictions of the QM model and the predictions of the simpler, surrogate model for the reaction rate constant. Kriging, although simpler than the QM model, is sophisticated enough, as discussed in section 6.2.3, to meet both requirements of accuracy and convergence.

Step 4. Design of a candidate optimal solvent.

The Kriging expression for the rate constant from Step 3 is included in the CAMD formulation (chapter 5). Details of the implementation of Kriging into the CAMD framework and the equations that are replaced are discussed in section 6.4. The CAMD problem becomes a standard MINLP problem. The solvent that maximises the rate constant of the reaction, based on the current approximate Kriging model,

subject to a number of constraints is thus considered a candidate optimal solvent.

Step 5. Convergence check.

If a new solvent is designed in Step 4, the algorithm returns to Step 2 in order to calculate the rate constant of the designed solvent with the QM model, before it is included in the initial set of data, and the Kriging surface is rebuilt. The algorithm terminates when no better solvent is found.

Step 6. Suitability check / Experimental validation.

The candidate solvent is checked against any criteria that were not included in the design problem formulation, such as chemical stability or safety. If these criteria are also met, the suitability of the designed solvent is validated via kinetic experiments. This last part is not performed in this thesis.

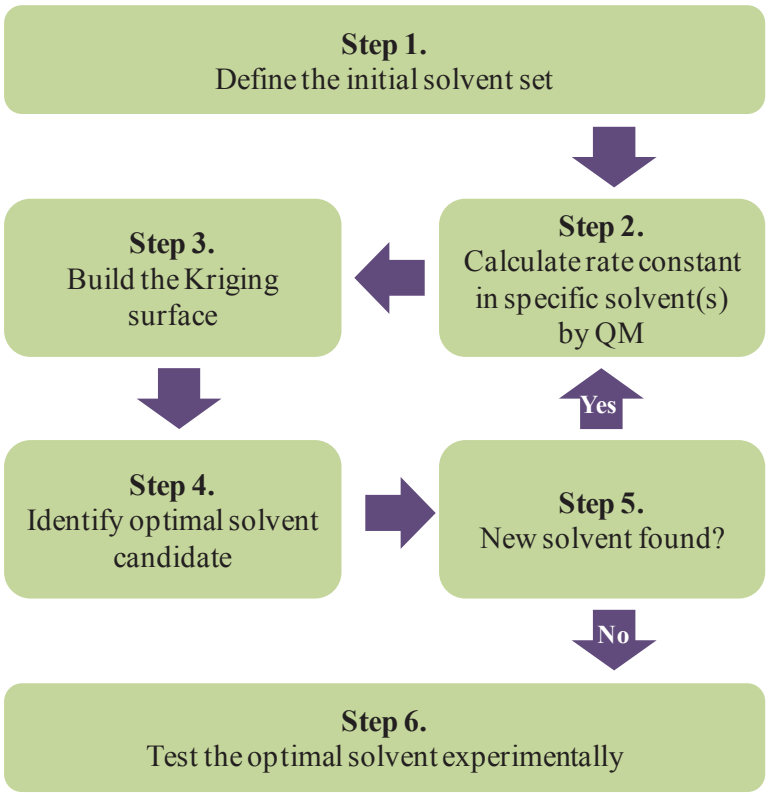


Figure 6.2: The proposed QM-CAMD-Kriging algorithm.

When Kriging is included in the mathematical formulation of the computer-aided molecular design problem, given from equations (5.1) - (5.9), the MINLP problem

becomes

$$\max_{\boldsymbol{\xi},\mathbf{n},\mathbf{y}} \; k^{KR}(\boldsymbol{\xi}) \tag{6.28}$$

subject to

$$h_1(\boldsymbol{\xi},\mathbf{n},\mathbf{y}) = \mathbf{0} \tag{6.29}$$

$$g_1(\boldsymbol{\xi},\mathbf{n},\mathbf{y}) \leq \mathbf{0} \tag{6.30}$$

$$h_2(\mathbf{n},\mathbf{y}) = \mathbf{0} \tag{6.31}$$

$$g_2(\mathbf{n},\mathbf{y}) \leq \mathbf{0} \tag{6.32}$$

$$d(\boldsymbol{\xi},\mathbf{n},\mathbf{y}) \leq \mathbf{0} \tag{6.33}$$

$$\boldsymbol{\xi} \in \mathbb{R}^m \tag{6.34}$$

$$\mathbf{n} \in \mathbb{R}^{\mathbf{q}} \tag{6.35}$$

$$y_i \in \{0,1\}^u \quad i = 1,...,q \tag{6.36}$$

where the objective function, equation (6.28), is now an analytical expression of the rate constant derived from Kriging k^{KR}, h_1 is a set of of structure-property equality constraints, g_1 is a set of structure-property inequality constraints, h_2 is a set of chemical feasibility and molecular complexity equality constraints, g_2 is a set of chemical feasibility and molecular complexity inequality constraints and d is a set of design constraints. $\boldsymbol{\xi}$ is a m-dimensional vector of variables denoting physical properties, $\mathbf{n}$ is a q-dimensional vector of variables denoting the number of groups in a molecule and $\mathbf{y}$ is a $q \times u$ matrix of binary variables used to activate groups and constrain continuous variables to integer values. The constraints are as described in chapter 5, except for constraints h_1, where the Kriging equations are now also included, as discussed in the next section.

6.4 The reaction rate constant from Kriging

The solvent in the SMD solvation model [Marenich et al., 2009] is defined by seven properties: dielectric constant ε, refractive index n_D, hydrogen bond acidity α, hy-

drogen bond basicity β, surface tension γ, aromaticity φ, and halogenicity ψ. These seven solvent properties are used to build a Kriging surface of the rate constant. The objective function (6.28) is the reaction rate constant from Kriging, replacing equation (5.11) in the case of the full problem, where QM calculations were included in the problem formulation. The Kriging equations are included in the structure and property constraints $h_1(\boldsymbol{\xi}, \mathbf{n}, \mathbf{y}) = \mathbf{0}$ (equation (6.29)) and are reformulated below to express the reaction rate constant as function of the solvent properties. The resulting MINLP problem consists of $465 + 7n$ variables, of which $149 + 7n$ are binary, where n is the number of data points used in the Kriging model (e.g., for 10 data points, the total number of variables is 695, of which 219 are binary).

The Kriging equation (6.20) for the rate constant can be written as

$$k^{KR} = \hat{k}^{TST} + \mathbf{r}^T(\boldsymbol{\xi}_s)\mathbf{R}^{-1}\left(\mathbf{k}^{TST} - \mathbf{1}\hat{k}^{TST}\right) \tag{6.37}$$

where k^{KR} is the prediction of Kriging for the rate constant at a set of solvent properties $\boldsymbol{\xi}_s$ in the surface and $\boldsymbol{\xi}_s = (\varepsilon, n_D, \alpha, \beta, \gamma, \varphi, \psi)$. $\mathbf{k}^{TST}$ is a vector of n rate constants calculated by QM, and $\hat{k}^{TST}$ is the mean value (equation (6.17)) of $\mathbf{k}^{TST}$. The $(i,\ j)$ element of the correlation matrix $\mathbf{R}$ is

$$R_{ij} = \exp\left(-\sum_{l=1}^{7}\theta_l|\boldsymbol{\xi}_{s,il} - \boldsymbol{\xi}_{s,jl}|^{p_l}\right) \tag{6.38}$$

where number 7 corresponds to the number of solvent properties. The vector $\mathbf{r}$ at a new set of properties $\boldsymbol{\xi}^*$ is

$$\mathbf{r}(\boldsymbol{\xi}_s) = \begin{pmatrix} \exp\left(-\sum_{l=1}^{7}\theta_l|\boldsymbol{\xi}^*_{s,l} - \boldsymbol{\xi}_{s,1l}|^{p_l}\right) \\ \vdots \\ \exp\left(-\sum_{l=1}^{7}\theta_l|\boldsymbol{\xi}^*_{s,l} - \boldsymbol{\xi}_{s,nl}|^{p_l}\right) \end{pmatrix} \tag{6.39}$$

The resulting Kriging surface has seven dimensions (variables) and fourteen adjustable parameters; seven θ and seven p. As has been mentioned before, for the general case of Kriging, $\theta_l \geq 0$ and $0 < p_l \leq 2$. In this work to simplify the problem, parameter p is set equal to 2 for all variables, as is commonly done [Zerpa et al., 2005], [Caballero and Grossmann, 2008], [Beck et al., 2012], and the following bounds

are set for parameter θ (similarly to Beck et al. [2012])

$$0.000001 \leq \theta_l \leq 10. \tag{6.40}$$

6.5 Implementation

Quantum mechanical calculations

For the quantum mechanical calculations Gaussian 09 [Frisch et al., 2009] is used. All calculations are performed at the M05-2X/6-31G(d) level of theory. The geometries of all species are optimised both in the gas and the liquid phases.

Regression of the Kriging model

For the implementation of the Kriging equations a code has been developed in Fortran 98. The input to the code consists of the solvent data points, that is, the seven properties for each solvent in the data set and the rate constants in these solvents calculated with the QM model. The concentrated log-likelihood (equation (6.19)) is maximised and the output includes parameters θ, the mean value $\widehat{\mu}$, the variance $\widehat{\sigma}^2$, the inverse correlation matrix $\mathbf{R}^{-1}$ and the augmented vector $\mathbf{r}$. Instead of maximisation, minimisation of the negative of the concentrated log-likelihood function is performed using routine E04JYF from the NAG Library (see www.nag.co.uk). Fortran routines ludcmp and gaussj from numerical recipes [Press et al., 1986] are used to calculate the determinant and the inverse of the correlation matrix, respectively.

Optimisation CAMD

The Kriging information after the regression is used in the CAMD formulation to design the optimal solvent. The CAMD formulation has been implemented in GAMS [GAMS Development Corporation, 2011] and it is solved with the global solver BARON [Tawarmalani and Sahinidis, 2005]. CONOPT [ARKI Consulting and Development] and CPLEX [IBM] have been chosen to be the NLP and MILP solvers, respectively, used within BARON.

6.6 Application to a Menschutkin reaction

The proposed solvent design methodology is applied to the Menschutkin reaction of phenacyl bromide and pyridine (figure 4.3). The M05-2X/6-31G(d) level of theory is used in this chapter for the quantum mechanical calculations, since it proved to provide the best quantitative predictions compared to other theories tested (see section 4.6). The main issue that is addressed here is how the best solution of the general problem described in chapter 5 can be ensured. The quality of the solution significantly depends on the extent of the search of the design space; the more thorough the search, the closer the reported solution can be to the global optimum. For this reason, the effect of the number of initial set of solvents on the algorithm is investigated and two cases are considered; in the first case ten initial solvents are included, while in the second case the first Kriging surface is built with information from fourteen solvents. A final case study is presented, where the OH group is removed from the design space, as it has been noticed [Struebing, 2011] that the SMD model tends to overestimate the rate constant in alcohols. Data for the properties of the solvents that are considered in the design are given whenever are available.

6.6.1 Case 1. Ten initial solvents

Seven Kriging parameters, the θ parameters, are regressed when building a Kriging surface. Thus ten initial data points are considered to be reasonable for the initialisation of the algorithm. The criterion for the choice of solvents used here is the dielectric constant. Solvents that cover a wide range of dielectric constants will enable the representation of a large design space.

Iteration 1: CH_2, Br, OH

Step 1. Ten organic solvents with diverse dielectric constants are chosen; toluene, chlorobenzene, ethyl acetate, dimethoxyethane, tetrahydrofuran, 1,2-dichloroethane, butanone, acetone, ethanol and acetonitrile. Experimental values of the dielectric constants of the initial set of solvents are given in Table 6.1.

Solvent	ε	Source
toluene	2.38	Buep and Baron [1988]
chlorobenzene	5.61	Fialkov et al. [1966]
ethyl acetate	6.06	Shirke et al. [2001]
dimethoxyethane	7.15	Salomon [1989]
tetrahydrofuran	7.39	Bhattacharyya et al. [1965]
1,2-dichloroethane	10.36	Fialkov et al. [1966]
butanone	18.39	Gilani et al. [2011]
acetone	20.56	Schiavo and Scrosati [1976]
ethanol	24.55	Ritzoulis et al. [2000]
acetonitrile	35.96	Salomon [1993]

Table 6.1: Experimental dielectric constants at 298.15 K of the initial set of solvents used in Case 1 for the Menschutkin reaction.

Step 2. The rate constants of the initial set of solvents are calculated using quantum mechanics and the SMD solvation model. The solvent properties required in the SMD model are calculated with the group-contribution methods by Sheldon et al. [2005] and Folić et al. [2007]. For acetonitrile and tetrahydrofuran, which are treated as single-group molecules, experimental values are used [Struebing, 2011]. The properties of the solvents and the rate constants calculated from QM are given in Table 6.2.

Step 3. The first Kriging surface is built based on this information and the resulting θ parameters can be seen in Table 6.3 (Iteration 1). All variables, except for basicity β, are active, that is, the corresponding parameters θ is non-zero. The variables that are the most active are the refractive index n_D and acidity α. Their θ values are equal to ten, which is the chosen upper bound. The mean value (equation (6.17)) of this first surface is $\widehat{\mu} = 1.597 \cdot 10^{-3}$ and the variance (equation (6.18)) is $\widehat{\sigma}^2 = 1.474 \cdot 10^{-6}$, as can also be seen also in Table 6.4 (Iteration 1).

Step 4. The MINLP problem is solved and the candidate solvent designed in this

first iteration has three groups: CH_2, Br, OH; it corresponds to bromomethanol. The predicted rate constant from Kriging is $2.678 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The properties of the designed solvent are given in Table 6.5 (Iteration 1). As was expected from the θ values, it has high values for refractive index n_D and acidity α, compared to the initial set of solvents, but also a very high value for the surface tension γ.

Step 5. A new solvent has been designed, thus the algorithm goes back to Step 2 and the second iteration begins.

Iteration 2: 2 × CH_3, C=C, $CHCl_2$, OH

Step 2. The rate constant in bromomethanol, is calculated with the QM model and the resulting value is $4.588 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. Thus a better solvent, compared to the initial set, has been designed. The absolute deviation between the predictions of Kriging and the QM model is $1.910 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}.

Step 3. The Kriging surface is now built with eleven points and the results of the regression can be found in Table 6.3 (Iteration 2). The mean value is $\widehat{\mu} = 1.834 \cdot 10^{-3}$ and the variance is $\widehat{\sigma}^2 = 5.138 \cdot 10^{-6}$ (Table 6.4, Iteration 2). The values of the θ parameters are similar to the previous iteration for most properties, but the parameter for the surface tension γ has been increased and halogenicity ψ is no longer active.

Step 4. When implementing the new Kriging surface in the CAMD formulation a new solvent is designed and consists of five groups: 2 × CH_3, C=C, $CHCl_2$, OH. The designed molecule can be either 2-methyl-3-hydroxyl-4-dichloro-but-2-ene (or similar), or 1,1-dichloro-2-methyl-3-hydroxyl-but-2-ene (or similar). The rate constant that is predicted by Kriging is $4.747 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. All properties of the new solvent are given in Table 6.5. The designed solvent is again an alcohol.

Step 5. Since a new solvent has been designed, the algorithm continues with iteration 3.

Initial set of solvents									
Solvent	Solvent properties							k^{TST}	k^{EXP}
	n_D	α	β	γ	ε	φ	ψ		
toluene	1.50	0.00	0.16	39.23	2.35	0.86	0.00	1.632E-05	1.145E-04
chlorobenzene	1.43	0.00	0.09	43.74	5.46	0.86	0.14	2.822E-04	
ethyl acetate	1.50	0.00	0.48	38.55	6.10	0.00	0.00	3.473E-04	
dimethoxyethane	1.46	0.00	0.62	32.18	4.92	0.00	0.00	2.349E-04	
tetrahydrofuran	1.40	0.00	0.36	39.44	7.43	0.00	0.00	5.810E-04	2.585E-04
1,2-dichloroethane	1.44	0.00	0.15	43.53	9.40	0.00	0.50	6.830E-04	
butanone	1.49	0.00	0.49	35.86	17.06	0.00	0.00	1.302E-03	
acetone	1.49	0.00	0.49	34.14	19.53	0.00	0.00	1.544E-03	1.228E-03
ethanol	1.73	0.33	0.49	39.68	17.60	0.00	0.00	2.576E-03	3.000E-04
acetonitrile	1.34	0.06	0.20	41.25	35.69	0.00	0.00	2.342E-03	2.613E-03

Table 6.2: The initial set of solvents that is used in Case 1 for the Menschutkin reaction and their predicted properties from the GC methods by Sheldon et al. [2005], Folić et al. [2007]. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constant k^{TST} is in dm^3 mol^{-1} s^{-1}. The experimental rate constants k^{EXP} in dm^3 mol^{-1} s^{-1} are from Ganase et al. [2013].

Iteration 3: 2 × CH_2, $CHCl_2$, OH

The rate constant of the reaction in 2 × CH_3, C=C, $CHCl_2$, OH is calculated with the QM model and it is found to be equal to $5.509 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, which is higher than the value of the solvent designed in the first iteration. The absolute deviation of the Kriging prediction is now reduced to $7.620 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}. A new Kriging surface is now regressed with 12 data and the resulting parameters are shown in Tables 6.3 and 6.4 (Iteration 3).

The solution of the MINLP CAMD leads to a new solvent: $CHCl_2CH_2CH_2OH$, that is 1,1-dichloropropanol. The predicted rate constant from the Kriging predictor is $5.585 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1} (Table 6.5). The groups $CHCl_2$ and OH, already present in the solvent designed in Iteration 2, reappear in the new candidate.

Iteration 4: CH_2, $CHCl_2$, OH

The rate constant of 1,1-dichloropropanol from the QM model is equal to $5.592 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, with only an absolute deviation of $1.530 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1} from the Kriging prediction. Thus in this iteration a better solvent has also been designed. The solvent is included in the Kriging data and the new surface is built (Tables 6.3, 6.4, Iteration 4). The resulting MINLP is solved and $CHCl_2CH_2OH$, that is, 1,1-dichloroethanol, is designed. The rate constant from Kriging is $5.738 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1} (Table 6.5). The new solvent is very similar to the solvent designed in Iteration 3.

Iteration 5: $CHCl_2$, OH

The rate constant in 1,1-dichloroethanol from the QM model is $6.103 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, which is still higher than the other solvents designed, and the absolute deviation now increased to $3.650 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}. 1,1-dichloroethanol is included in the Kriging data set and the parameters of the new Kriging surface are shown in Tables 6.3 and 6.4 (Iteration 5). The optimal solvent designed in this iteration is $CHCl_2OH$, dichloromethanol. The predicted rate constant from the Kriging surface is $6.785 \cdot$

10^{-3} dm^3 mol^{-1} s^{-1}. The properties of the designed solvent are given is Table 6.5.

Iteration 6: $CHCl_2$, OH

The QM predicted rate constant for dichloromethanol is $6.832 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The absolute deviation of Kriging is only $4.700 \cdot 10^{-5}$ dm^3 mol^{-1} s^{-1}. The Kriging surface is regressed and it is very similar to the surface in Iteration 5 (Table 6.3, Iteration 6). The mean Kriging value has been increased from $2.508 \cdot 10^{-3}$ to $2.729 \cdot 10^{-3}$, but the variance is decreased from $9.024 \cdot 10^{-6}$ to $8.705 \cdot 10^{-6}$. In this iteration dichloromethanol is designed for the second time with a predicted rate constant of $6.833 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, with a $1.000 \cdot 10^{-6}$ dm^3 mol^{-1} s^{-1} deviation from the QM prediction. Since no new solvent is designed, the algorithm terminates and the best solvent found is dichloromethanol.

Solvent properties	Kriging parameters θ					
	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Iteration 6
n_D	10.00	10.00	10.00	10.00	2.035	2.158
α	10.00	10.00	10.00	10.00	4.057	6.009
β	1E-06	1E-06	0.004	1E-06	1E-06	1E-06
γ	0.903	3.312	1E-06	2.119	2.741	2.482
ε	2.204	1.164	0.213	0.779	1.104	0.920
φ	0.015	0.004	0.001	0.001	0.002	0.003
ψ	0.003	1E-06	0.001	0.391	1E-06	1E-06

Table 6.3: The Kriging parameters θ in Case 1 for the Menschutkin reaction.

Kriging parameters	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Iteration 6
$\widehat{\mu}$	1.597E-03	1.834E-03	4.379E-03	2.967E-03	2.508E-03	2.729E-03
$\widehat{\sigma}^2$	1.474E-06	5.138E-06	1.968E-05	4.189E-06	9.024E-06	8.705E-06

Table 6.4: The Kriging mean values, $\widehat{\mu}$, and variances, $\widehat{\sigma}^2$, in Case 1 for the Menschutkin reaction.

Overall, six iterations were required for the algorithm to converge. QM calculations for fifteen solvents have been performed; ten initial solvents and five designed candidate solvents. When compared to the best solvent in the initial set, ethanol ($k^{TST} = 2.576 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}), the designed solvent, dichloromethanol ($k^{TST} = 6.832 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}), provides a 165% increase in the rate constant. The optimal solvent designed by Struebing et al. [2013] for the same reaction was nitromethane, CH_3NO_2. Note that nitromethane is considered a single-group molecule and thus is not included in the design space here. The rate constant in nitromethane is $3.018 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1} at the M05-2X/6-31G(d) level of theory. Thus a 126% increase in the predicted rate constant is achieved here with dichloromethanol. Therefore, this methodology proved successful in providing a better solution to the problem and, therefore, identifying a better solvent, based on the underlying computational model. Unfortunately, there are no experimental data to verify the predictions. Dichloromethanol, although not a popular solvent, has been used as a solvent in Cowan [1999].

It should be noted that dichloromethanol is an improved solvent, but it cannot be argued that is the optimal solvent. In order to find the optimal solvent, an exhaustive search of the design space is required as well as the use of global optimisation techniques (for example in the regression of the Kriging surface a local NLP solver is used). The search of the design space depends on the data points used to build the Kriging model and this is what is addressed in the next section. The number of solvents in the initial set is increased from ten to fourteen to investigate how the choice of the number of initial data points affects the solution of the problem.

Iteration	Designed solvent	Solvent properties							k^{TST}	k^{KR}
		n_D	α	β	γ	ε	φ	ψ		
1	CH_2, Br, OH	1.483	0.345	0.490	51.89	20.83	0.000	0.333	4.588E-03	2.678E-03
2	$2 \times CH_3$, C=C, $CHCl_2$, OH	1.459	0.494	0.486	50.05	17.13	0.000	0.250	5.509E-03	4.747E-03
3	$2 \times CH_2$, $CHCl_2$, OH	1.462	0.492	0.486	47.79	19.38	0.000	0.333	5.592E-03	5.585E-03
4	CH_2, $CHCl_2$, OH	1.468	0.493	0.486	48.61	21.98	0.000	0.400	6.103E-03	5.738E-03
5	$CHCl_2$, OH	1.476	0.494	0.486	50.21	25.68	0.000	0.500	6.832E-03	6.785E-03
6	$CHCl_2$, OH	1.476	0.494	0.486	50.21	25.68	0.000	0.500	6.832E-03	6.833E-03

Table 6.5: The predicted properties and rate constants from TST and Kriging of the designed solvents in Case 1 for the Menschutkin reaction. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constants k^{TST} and k^{KR} are given in dm^3 mol^{-1} s^{-1}.

6.6.2 Case 2. Fourteen initial solvents

It has been reported that the choice of the initial data affects the predictability of the Kriging model [Jones, 2001]. The initial points on the Kriging surface, if not chosen correctly, could be misleading and give a poor description of the total surface. Naturally, it is not possible to know beforehand which are the "right" solvents that would give the best description of the rate constant surface, and of course the "right" solvents would differ from reaction to reaction. Nevertheless, although the choice of solvents may be challenging, it has also been shown [Jones, 2001] that the larger the number of the initial data, the better the description of the surface from Kriging. This is what will be investigated in this section. The number of the initial solvents is increased, so as to observe the effect of it on the solution of the problem. Data from fourteen solvents are included in the initial data set, in Step 1 of the algorithm, instead of ten as in Case 1 in section 6.6.1. Four additional solvents have been added to the first initial set in Case 1: hexane, acetic acid, 2-butanol, 1,2-dichlorobenzene (Table 6.6).

Iteration 1: CH_2Cl, OH

Step 1. The initial set of solvents consists of fourteen solvents with diverse dielectric constants: hexane, toluene, chlorobenzene, ethyl acetate, acetic acid, dimethoxyethane, tetrahydrofuran, 1,2-dichlorobenzene, 1,2-dichloroethane, 2-butanol, butanone, acetone, ethanol and acetonitrile. Experimental values of their dielectric constants are shown in Table 6.6.

Step 2. The rate constant in the initial set of solvents is calculated with QM at the M05-2X/6-31G(d) level of theory with the SMD model. The solvent properties calculated with the group-contribution methods from Sheldon et al. [2005] and Folić et al. [2007] and the predicted rate constants are given in Table 6.7.

Step 3. The first Kriging surface is built based on the initial fourteen points and the resulting parameters for θ are shown in Table 6.8 (Iteration 1). The refractive index n_D is the only non-active variable. The highest value of θ corresponds to acidity

Solvent	ε	Source
hexane	1.89	Iglesias et al. [2000]
toluene	2.38	Buep and Baron [1988]
chlorobenzene	5.61	Fialkov et al. [1966]
ethyl acetate	6.06	Shirke et al. [2001]
acetic acid	6.13	Hanna [1984]
dimethoxyethane	7.15	Salomon [1989]
tetrahydrofuran	7.39	Bhattacharyya et al. [1965]
1,2-dichlorobenzene	9.93	Riddick et al. [1986]
1,2-dichloroethane	10.36	Fialkov et al. [1966]
2-butanol	15.95	Sastry and Patel [2000]
butanone	18.39	Gilani et al. [2011]
acetone	20.56	Schiavo and Scrosati [1976]
ethanol	24.55	Ritzoulis et al. [2000]
acetonitrile	35.96	Salomon [1993]

Table 6.6: Experimental dielectric constants at 298.15 K of the initial set of solvents that is used in Case 2 for the Menschutkin reaction.

α, followed by the dielectric constant ε. The mean value $\widehat{\mu}$ is now equal to $3.841 \cdot 10^{-4}$, which is lower than the corresponding value for 14 solvents in Case 1 (Iteration 5), and the variance $\widehat{\sigma}^2$ is $3.219 \cdot 10^{-6}$, also lower than the corresponding value in Case 1 (Iteration 5). The values are given in Table 6.9 (Iteration 1).

Step 4. The candidate solvent designed in this first iteration has two groups: CH_2Cl, OH, thus chloromethanol. The predicted rate constant from Kriging for chloromethanol is $3.067 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The properties of the designed solvent are given in Table 6.10. Chloromethanol is very similar to the optimal solvent designed in Case 1, dichloromethanol.

Step 5. Since a new solvent has been designed, chloromethanol is added to the first set of solvents and the algorithm proceeds to the second iteration.

Initial set of solvents									
Solvent	Solvent properties							k^{TST}	k^{EXP}
	n_D	α	β	γ	ε	φ	ψ		
hexane	1.39	0.00	0.12	28.07	3.10	0.00	0.00	7.726E-04	
toluene	1.50	0.00	0.16	39.23	2.35	0.86	0.00	1.632E-05	1.145E-04
chlorobenzene	1.43	0.00	0.09	43.74	5.46	0.86	0.14	2.822E-04	
ethyl acetate	1.50	0.00	0.48	38.55	6.10	0.00	0.00	3.473E-04	
acetic acid	1.58	0.61	0.43	27.12	3.15	0.00	0.00	1.136E-04	
dimethoxyethane	1.46	0.00	0.62	32.18	4.92	0.00	0.00	2.349E-04	
tetrahydrofuran	1.40	0.00	0.36	39.44	7.43	0.00	0.00	5.810E-04	2.585E-04
1,2-dichlorobenzene	1.44	0.00	0.02	48.79	4.81	0.75	0.25	2.098E-04	
1,2-dichloroethane	1.44	0.00	0.15	43.53	9.40	0.00	0.50	6.830E-04	
2-butanol	1.63	0.33	0.49	37.75	13.78	0.00	0.00	2.075E-03	
butanone	1.49	0.00	0.49	35.86	17.06	0.00	0.00	1.302E-03	
acetone	1.49	0.00	0.49	34.14	19.53	0.00	0.00	1.544E-03	1.228E-03
ethanol	1.73	0.33	0.49	39.68	17.60	0.00	0.00	2.576E-03	3.000E-04
acetonitrile	1.34	0.06	0.20	41.25	35.69	0.00	0.00	2.342E-03	2.613E-03

Table 6.7: The initial set of solvents used in Case 2 for the Menschutkin reaction and their predicted properties from the GC methods by Sheldon et al. [2005], Folić et al. [2007]. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constant k^{TST} is in dm^3 mol^{-1} s^{-1}. The experimental rate constants k^{EXP} in dm^3 mol^{-1} s^{-1} are from Ganase et al. [2013].

Iteration 2: $CHCl_2$, OH

Step 2. The rate constant in chloromethanol is calculated with the QM model and the resulting value is $4.912 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The absolute deviation between the predictions of Kriging and TST is $1.845 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}.

Step 3. The Kriging surface is built with fifteen solvents and the resulting parameters for θ are shown in Table 6.8 (Iteration 2). All variables are now active. An increase in the parameter value for halogenicity ψ is noted. The remaining Kriging parameters have the same trend as in the previous iteration. Similar values are also obtained for $\widehat{\mu}$ and $\widehat{\sigma}^2$ parameters (Table 6.9, Iteration 2).

Step 4. The solvent designed in the second iteration is $CHCl_2$, OH, i.e. dichloromethanol, the optimal solvent designed in Case 1. The predicted rate constant from Kriging is $5.208 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}.

Step 5. A new candidate solvent has been designed. Dichloromethanol is included in the regression data and the algorithm returns to Step 2 for the third iteration.

Iteration 3: $CHCl_2$, OH

The rate constant for dichloromethanol from the QM model is $6.832 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, thus the absolute deviation of Kriging is $1.624 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The Kriging surface is regressed and it is quite different from the surface in Iteration 2 (Table 6.8, Iteration 3). The θ value for the surface tension γ is now the highest, followed by acidity α. In the contrary, the parameters for aromaticity φ and halogenicity ψ decrease. Both the mean Kriging value and the variance increase (Table 6.9, Iteration 3). In this iteration dichloromethanol is designed for the second time with a predicted rate constant of $6.833 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, with a $1.000 \cdot 10^{-6}$ dm^3 mol^{-1} s^{-1} absolute deviation from the QM prediction. No new solvent has been designed, thus the algorithm terminates. The best solvent found is dichloromethanol.

In this case study, 3 iterations were required for the algorithm to converge. QM calculations for sixteen solvents were performed. In the next section the outcome of the two case studies is briefly discussed.

Solvent properties	Kriging parameters θ		
	Iteration 1	Iteration 2	Iteration 3
n_D	1E-06	0.118	0.223
α	2.434	2.665	1.246
β	0.119	0.101	0.041
γ	0.552	0.297	4.084
ε	1.728	1.368	0.136
φ	0.046	0.020	3E-04
ψ	0.033	1.090	1E-06

Table 6.8: The Kriging parameters θ in Case 2 for the Menschutkin reaction.

Kriging parameters	Iteration 1	Iteration 2	Iteration 3
$\widehat{\mu}$	3.841E-04	2.273E-04	1.072E-03
$\widehat{\sigma}^2$	3.219E-06	5.073E-06	3.320E-05

Table 6.9: The Kriging mean values, $\widehat{\mu}$, and variances, $\widehat{\sigma}^2$, in Case 2 for the Menschutkin reaction.

Discussion

The number of initial solvents does not affect the result in this particular application. Both case studies converged to the same solvent. The number of required QM calculations in the case studies is similar; 15×3 in Case 1 and 16×3 in Case 2. Thus both approaches require almost the same number of data points to build a good predictive Kriging surface. It is also worth mentioning that the most active variables of the final surfaces in the two cases are the same: the refractive index, acidity, surface tension and dielectric. Naturally, an exact match of the Kriging parameters is not expected, as the data used for their regression are different.

Iteration	Solvent	Solvent properties							k^{TST}	k^{KR}
		n_D	α	β	γ	ε	φ	ψ		
1	CH_2Cl, OH	1.484	0.328	0.498	50.74	26.68	0.000	0.333	4.912E-03	3.067E-03
2	$CHCl_2$, OH	1.476	0.494	0.486	50.21	25.68	0.000	0.500	6.832E-03	5.208E-03
3	$CHCl_2$, OH	1.476	0.494	0.486	50.21	25.68	0.000	0.500	6.832E-03	6.833E-03

Table 6.10: The predicted properties and rate constants from TST and Kriging of the designed solvents in Case 2 for the Menschutkin reaction. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constants k are given in dm^3 mol^{-1} s^{-1}.

The approximate CPU time required for the solution of the MINLP problems in the two case studies is shown in Table 6.11. The calculations are performed in GAMS [GAMS Development Corporation, 2011] with the global solver BARON [Tawarmalani and Sahinidis, 2005]. The convergence tolerance considered is 10^{-2} for the absolute difference of upper and lower bound. The final Kriging surface in Case 1 is built with one solvent less and the CPU time is lower. This is probably because the solvents that were added during the iterations were promising solvents that improved the surface and not randomly chosen, like the initial solvents in Case 2. Therefore, it could be argued that the best approach is to start the algorithm with a relatively small number of solvents and progressively improve the Kriging surface with promising candidate solvents. However, with this approach there is always the risk of being limited to one part of the design space, thereby improving the Kriging surface in one part, but leaving other parts unexplored. On the other hand, when more initial points are included, the computational time is higher, but the risk of restricting the exploration of the design space is reduced.

Iteration	Case 1	Case 2
1	1	37
2	2	200
3	1	250
4	52	-
5	75	-
6	75	-
Total	206	487

Table 6.11: Approximate CPU (one-processor Intel(R) Xeon(R) X5690, 3.47GH) time in hours for the MINLP solutions in Cases 1 and 2 for the Menschutkin reaction.

All the candidate solvents designed in these two cases are alcohols, including the optimal solvent. Struebing [2011] showed that the SMD solvation model overestimates the rate constants in alcohols. In the next section, a third case study is performed,

where alcohols are not allowed to be designed. To achieve this, the OH group is removed from the design space.

6.6.3 Restricted design space: OH group removed

The OH group is removed from the functional groups in the CAMD formulation, in order to avoid possible bias in the solution, due to the overestimation of the rate constants of solvents containing the OH group from the SMD model. Fourteen solvents are used in the initial data set.

Iteration 1: CH_2Cl, CH_2NO_2

In this case study, the initial set of solvents should not contain alcohols. Therefore, the initial solvents used in Case 2 are used but ethanol and 2-butanol are replaced by benzaldehyde and nitroethane. Thus the initial set contains the following solvents: hexane, toluene, chlorobenzene, ethyl acetate, acetic acid, dimethoxyethane, tetrahydrofuran, 1,2-dichlorobenzene, 1,2-dichloroethane, benzaldehyde, butanone, acetone, nitroethane and acetonitrile (Table 6.12).

The Kriging surface is regressed with these data and the resulting θ parameters are shown in Table 6.14. It is noted that the most active variables are the dielectric constant ε and the basicity β. This last property was not generally included in the highly active variables in the previous cases. The Kriging mean value and the variance are given in Table 6.15. The Kriging predictor is included in the CAMD framework and the designed candidate solvent consists of two groups: CH_2Cl, CH_2NO_2. The predicted rate constant from Kriging is $2.232 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. Since a new solvent has been designed, the algorithm proceeds to Iteration 2.

Iteration 2: CH_2Cl, CH_2NO_2

The rate constant of the designed solvent is calculated with the QM model and it is $3.239 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The absolute deviation of the Kriging prediction is thus $1.007 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. CH_2Cl, CH_2NO_2 is included in the data set and

Solvent	ε	Source
hexane	1.89	Iglesias et al. [2000]
toluene	2.38	Buep and Baron [1988]
chlorobenzene	5.61	Fialkov et al. [1966]
ethyl acetate	6.06	Shirke et al. [2001]
acetic acid	6.13	Hanna [1984]
dimethoxyethane	7.15	Salomon [1989]
tetrahydrofuran	7.39	Bhattacharyya et al. [1965]
1,2-dichlorobenzene	9.93	Riddick et al. [1986]
1,2-dichloroethane	10.36	Fialkov et al. [1966]
benzaldehyde	17.40	Borovikov et al. [1977]
butanone	18.39	Gilani et al. [2011]
acetone	20.56	Schiavo and Scrosati [1976]
nitroethane	28.00[a]	Jehlicka [1990]
acetonitrile	35.96	Salomon [1993]

Table 6.12: Experimental dielectric constants at 298.15 K of the initial set of solvents that is used in the case study with removed OH group for the Menschutkin reaction. [a]Measured at 303.15 K.

the Kriging surface is re-built. The Kriging parameters are shown in Tables 6.14 and 6.15. The highly active variables are now the halogenicity ψ and the dielectric constant ε. The candidate solvent designed in this iteration is CH_2Cl, CH_2NO_2 for the second time. The Kriging predicted rate constant is now $3.239 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, as also predicted from the QM model. Therefore, the algorithm converged and the optimal solvent found is CH_2Cl, CH_2NO_2, i.e. 1-chloro-2-nitroethane.

The group CH_2Cl has reappeared in the previous case studies and is similar to group $CHCl_2$ that appears in the optimal solvent designed before. Group CH_2NO_2 appears for the first time here, but it was found to be in candidate solvents in Struebing et al. [2013]. Therefore, although the designed solvent, 1-chloro-2-nitroethane,

Initial set of solvents									
Solvent	Solvent properties							k^{TST}	k^{EXP}
	n_D	α	β	γ	ε	φ	ψ		
hexane	1.39	0.00	0.12	28.07	3.10	0.00	0.00	7.726E-04	
toluene	1.50	0.00	0.16	39.23	2.35	0.86	0.00	1.632E-05	1.145E-04
chlorobenzene	1.43	0.00	0.09	43.74	5.46	0.86	0.14	2.822E-04	
ethyl acetate	1.50	0.00	0.48	38.55	6.10	0.00	0.00	3.473E-04	
acetic acid	1.58	0.61	0.43	27.12	3.15	0.00	0.00	1.136E-04	
dimethoxyethane	1.46	0.00	0.62	32.18	4.92	0.00	0.00	2.349E-04	
tetrahydrofuran	1.40	0.00	0.36	39.44	7.43	0.00	0.00	5.810E-04	2.585E-04
1,2-dichlorobenzene	1.44	0.00	0.02	48.79	4.81	0.75	0.25	2.098E-04	
1,2-dichloroethane	1.44	0.00	0.15	43.53	9.40	0.00	0.50	6.830E-04	
benzaldehyde	1.60	0.00	0.46	52.40	21.27	0.75	0.00	1.808E-03	
butanone	1.49	0.00	0.49	35.86	17.06	0.00	0.00	1.302E-03	
acetone	1.49	0.00	0.49	34.14	19.53	0.00	0.00	1.544E-03	1.228E-03
nitroethane	1.66	0.00	0.30	55.52	20.03	0.00	0.00	2.017E-03	
acetonitrile	1.34	0.06	0.20	41.25	35.69	0.00	0.00	2.342E-03	2.613E-03

Table 6.13: The initial set of solvents that is used in the case study with removed OH group for the Menschutkin reaction and their predicted properties from the GC methods by Sheldon et al. [2005], Folić et al. [2007]. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constant k^{TST} is in dm^3 mol^{-1} s^{-1}. The experimental rate constants k^{EXP} in dm^3 mol^{-1} s^{-1} are from Ganase et al. [2013].

Solvent properties	Kriging parameters θ	
	Iteration 1	Iteration 2
n_D	1E-06	1E-06
α	0.042	0.016
β	1.188	1.185
γ	1E-06	0.132
ε	3.278	2.669
φ	0.583	0.113
ψ	0.383	4.131

Table 6.14: The Kriging parameters θ in the case study with removed OH group for the Menschutkin reaction.

Kriging parameters	Iteration 1	Iteration 2
$\widehat{\mu}$	1.296E-03	1.610E-03
$\widehat{\sigma}^2$	8.962E-07	1.459E-06

Table 6.15: The Kriging mean values, $\widehat{\mu}$, and variances, $\widehat{\sigma}^2$, in the case study with removed OH group for the Menschutkin reaction.

is not a commercial solvent, the design shows that solvents that include these groups are prominent solvents for the Menschutkin reaction. Note that the rate constant for the designed solvent is again higher than the rate constant in nitromethane ($3.018 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}) designed by Struebing et al. [2013]. The final Kriging surface for this case study is different from the resulting surfaces in Cases 1 and 2. Now, the most active variables are halogenicity ψ, dielectric constant ε and basicity β.

Iteration	Solvent	Solvent properties							k^{TST}	k^{KR}
		n_D	α	β	γ	ε	φ	ψ		
1	CH_2Cl, CH_2NO_2	1.463	0.000	0.314	67.26	24.71	0.000	0.167	3.239E-03	2.232E-03
2	CH_2Cl, CH_2NO_2	1.463	0.000	0.314	67.26	24.71	0.000	0.167	3.239E-03	3.239E-03

Table 6.16: The predicted properties and rate constants from TST and Kriging for the designed solvents in the case study with removed OH group for the Menschutkin reaction. The surface tension γ is in cal mol^{-1} $Å^{-2}$. The rate constants k are given in dm^3 mol^{-1} s^{-1}

6.7 Application to a Cope elimination reaction

The developed methodology for solvent design for reactions has been also applied to a Cope elimination reaction. Cope elimination reactions are reversible reactions where an amine oxide decomposes into an olefin and a hydroxylamine [Cope et al., 1949]. Solvent effects on Cope elimination reactions have been studied both experimentally [Sahyun and Cram, 1963, Kwart et al., 1978, Kwart and Brechbiel, 1981] and computationally [Komaromi and Tronchet, 1997, Acevedo and Jorgensen, 2006] and it has been shown that the rate of these reactions as well as their equilibrium are highly influenced by the solvent. It has been observed that the forward reaction (elimination) is favoured in aprotic solvents, whereas the reverse elimination reaction is favoured in protic solvents [Ciganek et al., 1995]. The rate of a Cope elimination reaction can vary up to six orders of magnitude when going from protic to aprotic solvents [Sahyun and Cram, 1963]. Even among aprotic solvents, the rate is significantly increased with decreasing polarity [Sahyun and Cram, 1963]. The elimination reaction of methylamine oxide, shown in figure 6.3, is investigated here. This reaction has been studied theoretically using quantum mechanics by Komaromi and Tronchet [1997]. The mechanism and energetics of the reaction were studied, mainly in the

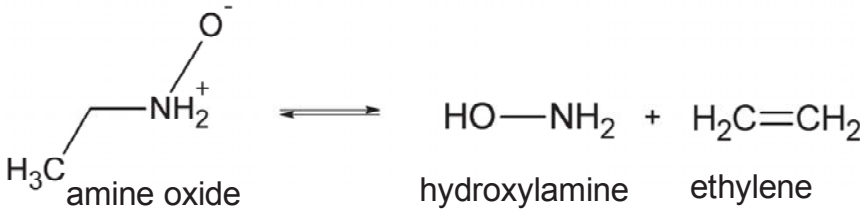


Figure 6.3: The Cope elimination reaction.

gas phase using several quantum mechanical models. The authors also investigated the solvent effects on the rate and equilibrium using a simple continuum solvation model, by varying the dielectric constant. To the best of my knowledge, experimental studies on this particular Cope elimination reaction have not been carried out. Only the forward reaction (elimination reaction) is considered here. Since ten initial solvents proved to be sufficient in the investigation of the Menschutkin reaction (section

6.6), this is also the number of the initial data points that is considered for the Cope elimination example. The progress of the algorithm and the results are presented in the rest of this section.

6.7.1 Results

Iteration 1: 3 × CH_3, 2 × CH_2, CH

Step 1. Ten organic solvents with diverse dielectric constants are chosen here to be included in the first set of data; hexane, chlorobenzene, ethyl acetate, acetic acid, tetrahydrofuran, 1,2-dichloroethane, butanone, acetone, ethanol and acetonitrile. Experimental values of the dielectric constants of the solvents in the initial set are given in Table 6.17.

Solvent	ε	Source
hexane	1.89	Iglesias et al. [2000]
chlorobenzene	5.61	Fialkov et al. [1966]
ethyl acetate	6.06	Shirke et al. [2001]
acetic acid	6.13	Hanna [1984]
tetrahydrofuran	7.39	Bhattacharyya et al. [1965]
1,2-dichloroethane	10.36	Fialkov et al. [1966]
butanone	18.39	Gilani et al. [2011]
acetone	20.56	Schiavo and Scrosati [1976]
ethanol	24.55	Ritzoulis et al. [2000]
acetonitrile	35.96	Salomon [1993]

Table 6.17: Experimental dielectric constants at 298.15 K of the solvents in the initial set used for the Cope elimination reaction.

Step 2. The rate constants for the initial set of solvents are calculated using quantum mechanics and the SMD solvation model. The properties of the solvents using the GC methods by Sheldon et al. [2005] and Folić et al. [2007] and the predicted

rate constants are given in Table 6.2. In the case of acetonitrile and tetrahydrofuran, which are treated as single-group molecules, experimental values are used [Struebing, 2011].

Step 3. The first Kriging surface is built from the initial set of data and the resulting parameters are shown in Table 6.19. The mean value is $\widehat{\mu} = 1.300 \cdot 10^{-5}$ and the variance is $\widehat{\sigma}^2 = 2.855 \cdot 10^{-10}$. All variables, except for the surface tension are active, with bacisity, aromaticity and halogenicity having θ values equal to ten and thus being the most important variables in the design.

Step 4. The MINLP problem is solved and the candidate solvent designed in this first iteration consists of six groups: $3 \times CH_3$, $2 \times CH_2$, CH , that is 2-methylpentane or 3-methylpentane. The predicted rate constant from Kriging is $1.188 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}. The properties of methylpentane are shown in Table 6.21.

Step 5. Since a new solvent has been designed, the algorithm goes back to Step 2 for the second iteration.

Iteration 2: CH_3, CH_2, CH_2Cl

Step 2. The rate constant in methylpentane from the QM model is $1.072 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}. Thus indeed a much better solvent, compared to the initial set, has been designed. The absolute deviation between the predictions of Kriging and the QM model is small and equal to $1.164 \cdot 10^{-5}$ dm^3 mol^{-1} s^{-1}.

Step 3. Methylpentane is included in the data set for building the Kriging surface and the results of the regression are given in Table 6.19 (Iteration 2). The mean value is $\widehat{\mu} = 1.217 \cdot 10^{-4}$ and the variance is $\widehat{\sigma}^2 = 3.457 \cdot 10^{-8}$ (Table 6.20, Iteration 2). Both values have increased. The values of the θ parameters are very different from the previous iteration. Basicity has the highest θ value, while the rest of the properties have small values, close to zero. This means that basicity is the most active variable and has the highest impact on the rate constant.

Step 4. The new Kriging surface is implemented in the CAMD formulation and the resulting MINLP is solved. A new solvent is designed and consists of three groups: CH_3, CH_2, CH_2Cl, that is 1-chloropropane. The rate constant that is predicted by

Initial set of solvents								
Solvent	Solvent properties							k^{TST}
	n_D	α	β	γ	ε	φ	ψ	
hexane	1.39	0.00	0.12	28.07	3.10	0.00	0.00	2.507E-05
chlorobenzene	1.43	0.00	0.09	43.74	5.46	0.86	0.14	6.271E-06
ethyl acetate	1.50	0.00	0.48	38.55	6.10	0.00	0.00	5.565E-06
acetic acid	1.58	0.61	0.43	27.12	3.15	0.00	0.00	4.762E-08
tetrahydrofuran	1.40	0.00	0.36	39.44	7.43	0.00	0.00	3.911E-06
1,2-dichloroethane	1.44	0.00	0.15	43.53	9.40	0.00	0.50	3.284E-06
butanone	1.49	0.00	0.49	35.86	17.06	0.00	0.00	1.694E-06
ethanol	1.73	0.33	0.49	39.68	17.60	0.00	0.00	9.652E-11
acetone	1.49	0.00	0.49	34.14	19.53	0.00	0.00	9.045E-07
acetonitrile	1.34	0.06	0.20	41.25	35.69	0.00	0.00	4.452E-07

Table 6.18: The initial set of solvents that is used in Case 1 for the Cope reaction and their predicted properties from the GC methods by Sheldon et al. [2005], Folić et al. [2007]. The surface tension γ is in cal mol^{-1} Å^{-2}. The rate constant k^{TST} is in dm^3 mol^{-1} s^{-1}.

Kriging for $CH_3CH_2CH_2Cl$ is $5.215 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}. The properties of the designed solvent are given in Table 6.21.

Step 5. Since a new solvent has been designed, the algorithm continues with iteration 3.

Iteration 3: 3 × CH_3, 2 × CH_2, CH

The reaction rate constant in 1-chloropropane is calculated with the QM model and is found to be equal to $3.603 \cdot 10^{-5}$ dm^3 mol^{-1} s^{-1}. The absolute deviation of the QM model from the Kriging model is $5.012 \cdot 10^{-3}$ dm^3 mol^{-1} s^{-1}, higher than the deviation at Iteration 1. 1-chloropropane is included in the data set and a new Kriging model is built. The Kriging parameters are shown in Tables 6.19 and 6.20 (Iteration

3). The mean and variance values are similar to the previous iteration, while the highest θ value corresponds now to surface tension, followed by basicity. The Kriging model is implemented in the CAMD formulation and the optimal solvent designed is $3 \times CH_3$, $2 \times CH_2$, CH, methylpentane, which was also designed in Iteration 1. The predicted rate constant from Kriging is $7.267 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}, with an absolute deviation of $6.195 \cdot 10^{-4}$ dm^3 mol^{-1} s^{-1} from the QM prediction. The overestimation of the rate constant from the Kriging model is probably due to the value of $\widehat{\mu}$, which is higher than all the rate constants of the solvents that are considered in the design. Since no new solvent has been designed in this iteration, the algorithm terminates here.

Solvent properties	Kriging parameters θ		
	Iteration 1	Iteration 2	Iteration 3
n_D	2.554	1E-06	0.759
α	0.285	0.003	0.003
β	10.00	7.726	5.017
γ	1E-06	0.020	10.00
ε	0.623	0.001	1E-06
φ	10.00	0.089	1E-06
ψ	10.00	0.029	1E-06

Table 6.19: The Kriging parameters θ for the Cope elimination reaction.

The algorithm converged after 3 iterations and 12 k^{TST} calculations. The candidate solvent that has been designed for the Cope elimination reaction is methylpentane, with a rate constant equal to 1.072E-04 $\cdot 10^{-4}$ dm^3 mol^{-1} s^{-1}, which compared to the best solvent in the initial data set, hexane ($k = 2.51 \cdot 10^{-5}$ dm^3 mol^{-1} s^{-1}), provides a 326% increase in the rate constant. It should be noted that, although there are no experimental data of the reaction rate constant in 2-methylpentane or 3-methylpentane to verify the predictions, experimental studies claim that Cope elim-

Kriging parameters	Iteration 1	Iteration 2	Iteration 3
$\widehat{\mu}$	1.300E-05	1.217E-04	1.374E-04
$\widehat{\sigma}^2$	2.855E-10	3.457E-08	1.455E-08

Table 6.20: The Kriging mean values, $\widehat{\mu}$, and variances, $\widehat{\sigma}^2$, for the Cope elimination reaction.

ination reactions are accelerated in aprotic, apolar solvents (e.g. dimethyl sulfoxide (DMSO) and tetrahydrofuran (THF)) [Sahyun and Cram, 1963], which is consistent with the results in this work. Interestingly, hexane and methylpentane are very similar molecules, nevertheless, the predicted difference in the rate constants is significant. This is likely due to the lower dielectric constant of methylpentane, since polarity highly affects the rate constant of the reaction and the rest of the properties are similar for the two solvents. There are experimental data for some properties of 2-methylpentane and 3-methylpentane and these are given in Table 6.22. The properties of the designed solvent are in good agreement with the experimental data and, particularly, they are closer to the properties of 3-methylpentane. Both 2-methylpentane and 3-methylpentane are used as commercial solvents (e.g., 2-methylpentane is used in Kalinovskaya and Vij [1999], Power et al. [2011] and 3-methylpentane is used in Hiratsuka et al. [1996], Kozlowski et al. [1999], Huck and Leigh [2011]).

In this example, the convergence of the Kriging model and the QM model is not as good as for the Menschutkin application. The most probable causes are the assumptions for the Kriging parameters, θ and p (set bounds for θ and fixed value for p). The Kriging parameters are restricted to take particular values, in order to simplify the problem (see section 6.4). These assumptions did not cause any obvious problem for the Menschutkin example, but the effect for the Cope elimination example is evident; the parameters often hit bounds during the estimation and the mean values for the Kriging models are high (Table 6.20), leading to overestimated Kriging predictions for the rate constants. Nevertheless, despite this overestimation,

the algorithm identifies a much improved solvent. In addition to widening the bounds of the θ and p parameters, the quality of the predictions could probably be improved by using global optimisation for the parameter estimation to ensure that the best possible parameters are used. Increasing the number of the initial data points could also be a possible solution to increase the reliability of the model.

Iteration	Designed solvent	Solvent properties							k^{TST}	k^{KR}
		n_D	α	β	γ	ε	φ	ψ		
1	3 × CH_3, 2 × CH_2, CH	1.377	0.000	0.000	26.45	1.996	0.000	0.000	1.072E-04	1.188E-04
2	2 CH_3, CH_2, CH_2Cl	1.459	0.000	0.1360	32.37	8.760	0.000	0.250	3.603E-06	5.215E-03
3	3 × CH_3, 2 × CH_2, CH	1.377	0.000	0.000	26.45	1.996	0.000	0.000	1.072E-04	1.188E-04

Table 6.21: The predicted properties and rate constants from TST and Kriging of the designed solvents for the Cope elimination reaction. The surface tension γ is in cal mol^{-1} $Å^{-2}$. The rate constants k^{TST} and k^{KR} are given in dm^3 mol^{-1} s^{-1}.

Solvent	Solvent properties		
	n_D	γ	ε
2-methylpentane	1.369[a]	24.18[b]	1.873[c]
3-methylpentane	1.374[a]	25.33[d]	1.887[c]

Table 6.22: Experimental data for the properties of 2-methylpentane and 3-methylpentane. The surface tension γ is in cal mol^{-1} Å^{-2}. [a]Aucejo et al. [1995], [b]Liessmann et al. [1995], [c]Ruzicka [1935], [d]Dewan and Mehta [1990]

6.8 Conclusions

Popular surrogate methods have been briefly reviewed and a description of the Kriging method has been presented. A new methodology for the optimal design of solvents for reactions has been presented. The approach combines a Kriging surface, built using quantum mechanically calculated rate constants, with a MINLP CAMD formulation. Initially a Kriging surface is built with QM data from a small number of solvents. The resulting Kriging predictor is included in the CAMD framework and a candidate solvent is designed. The designed solvent is included in the regression data, the Kriging surface is rebuilt and a new candidate solvent is designed. The algorithm iterates until no new solvent is designed. The new methodology has been applied successfully to a Menschutkin reaction and a Cope elimination reaction. In both applications, the proposed QM-CAMD-Kriging approach leads to the design of very promising solvents and, in the case of the Menschutkin reaction, the use of the Kriging model is found to be more successful than the solvatochromic surrogate model.

Chapter 7

Conclusions and Perspectives

7.1 Summary

The effects of solvents on reaction rates are widely recognised and have been studied for decades. Despite this fact, solvents are still chosen in academia and industry based on experience. Over the last 25 years there has been a growing interest in developing systematic methods for solvent design. These methods are known as computer-aided molecular design (CAMD) methods, and many notable approaches have been proposed for the design of pure and mixed solvents. A brief review of key CAMD methods and applications was presented in Chapter 2. The main advantages of CAMD methods are that they enable the screening of large number of molecules space and that multiple performance objectives and constraints can be considered.

In particular in problems where the impact of a solvent is important, it has been shown by numerous researchers that CAMD can be used to identify those solvents that improve the performance of a process. However, only a few works have been reported for the design of solvents for reactions; this is the topic of this thesis.

The design of a mixed solvent, in particular a gas-expanded liquid for a reactive process, was first considered (Chapter 3). Gas-expanded liquids have attracted a lot of attention because in addition to having the advantages of mixed solvents (composition as variable), they are also environmentally friendly, as an amount of the organic solvent is replaced by CO_2. A methodology for the integrated design

of a CO_2-expanded liquid in a reactive system has been developed here to identify the optimal co-solvent and operating conditions that minimise the cost of the process. In order to model this system, a number of aspects were taken into account: kinetics, phase equilibrium and process requirements. An empirical model, the solvatochromic equation, was used to describe the kinetics of the reaction, together with a preferential solvation model to correlate the reaction kinetics to the composition of the mixed solvent. The solid-liquid-vapour phase equilibrium in the reactor was predicted with the group-contribution VTPR equation of state, which was shown to be successful in predicting the phase equilibrium of CO_2-expanded liquids. The developed method was applied to the Diels-Alder reaction between anthracene and 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) and three organic co-solvents were considered; acetonitrile, acetone and methanol. Pure acetone resulted in the lowest cost, and both acetone and acetonitrile proved to be suitable co-solvents, as they offer good performance for a large range of CO_2 concentrations. It was also shown that with a small increase in the cost, the required amount of organic co-solvent is significantly reduced, leading to an environmentally benign solvent mixture. Naturally, a detailed environmental analysis would be necessary to make precise predictions. In conclusion, it was found that it is important to consider multiple process performance criteria in order to design the optimal GXL. The approach proposed is general and can be applied also to other processes that include GXLs, but relies on an empirical kinetic model which, although simple, requires kinetic experimental data.

In the search for more predictive approaches to solvent design in reactive systems, the effective use of quantum mechanical calculations in solvent design was investigated in the remainder of the thesis. The most widely used *ab initio* kinetic theory, conventional transition state theory, was chosen to be used here for the prediction of the reaction rate constant. An expression for the liquid-phase reaction rate constant was derived in Chapter 4 following to CTST, where the rate constant is a function of the activation Gibbs free energy of solvation. In considering the prediction of reaction rate constant in solution, the solvation free energy of the transition state and the reactants are also needed. Various models for predicting the solvation free

energy have been proposed in the literature, but SMD, a continuum solvation model, where the solute is described at the quantum level and the solvent is considered as a continuum medium described by several bulk properties, is chosen to be used in this thesis. The SMD solvation model is considered to be a good compromise between accuracy and computational time. The derived expression for the rate constant with the SMD model was used to predict the rate constant of the Menschutkin reaction of phenacyl bromide and pyridine in five organic solvents, where experimental data are available. Several levels of theory and basis sets were tested and it was shown that the QM models can, in general, represent satisfactorily the experimental data. The M05-2X/6-31G(d) method proved to be the most accurate among them, which can in part be explained by the fact that it was included in the parametrisation of the SMD model.

A computer-aided molecular design formulation for the design of optimal solvents for reactions was presented in Chapter 5. It incorporates the quantum mechanical expression for the rate constant, from CTST and SMD, as the objective function that is maximised subject to structure and property, chemical feasibility, and design constraints. The rate constant is calculated with on-the-fly quantum mechanical calculations. The resulting formulation is a bilevel MINLP problem with black-box functions. In order to simplify this non-trivial problem, the geometry of the reactants and the transition state were assumed not to change when passing from the gas phase to the liquid phase, and thus the problem was turned into a single level optimisation problem. The proposed approach was implemented within the global optimisation SMIN-αBB algorithm and applied to the Menschutkin reaction of phenacyl bromide and pyridine. The problem was solved locally with acetic acid as the initial guess and benzaldehyde was designed as the candidate solvent, with an improved rate constant. Ideally, the problem should be solved globally in order to design the optimal solvent, but the computational time required is prohibitive at this point in time.

In order to limit the computational requirements of the solvent design problem, but at the same time retain the accuracy provided by quantum mechanics, the use of a surrogate model was proposed in Chapter 6. Surrogate models have been widely

used in engineering applications to model functions that are computationally expensive to evaluate, like in the case of black-box functions. However, a simple surrogate model, such as the solvatochromic equation used in Chapter 3, would have an inherent mismatch with the detailed QM model, thus a sophisticated response surface model, Kriging, was chosen in this work to ensure the convergence between the detailed and the surrogate models. A new *ab initio* methodology has been developed for the design of solvents for reactions, which combines quantum mechanical calculations, the Kriging model and a computer-aided molecular design formulation. As a first step, an initial set of solvents is chosen and the rate constants in these solvents is calculated with the QM model (CTST and SMD). This information is then used to build the Kriging model of the rate constant. The Kriging expression is included in the CAMD formulation and the optimal solvent, given by the current model, is designed. If a new solvent is designed, its rate constant is calculated with the QM model, it is included in the data set for the Kriging regression, the Kriging model is rebuilt, and so on. The algorithm iterates until no new solvent is designed. The new methodology was successfully applied to two chemical reactions: the Menschutkin reaction of phenacyl bromide and pyridine and the Cope elimination reaction of methylamine oxide, and very promising solvents were identified. The choice of the number of solvents in the initial set was also investigated by choosing ten or fourteen solvents and it was shown not to affect the solution of the algorithm. The candidate solvent designed for the Menschutkin reaction was dichloromethanol with a 165% increase in the rate constant when compared to the best solvent of the initial data set. More importantly, dichloromethanol is also an improved solvent (a 239% increase) in comparison with the solvent designed when the solvatochromic equation was used as the surrogate model. In the case of the Cope elimination, the candidate solvent designed was methylpentane, with a 336% increase in the rate constant, compared to the best solvent included in the initial data set. Therefore, the developed QM-CAMD-Kriging methodology proved to be successful in identifying better solvents for reactions, without the need of any experimental data.

7.2 Main contributions

- A methodology for the integrated design of a CO_2-expanded solvent and reactive system has been developed, where kinetics, phase equilibrium and process requirements are taken into account. The proposed methodology has been applied to the Diels-Alder reaction of anthracene and PTAD. Acetone and acetonitrile were found to be suitable co-solvents, as they offer good performance for the reactive process over a range of CO_2 concentrations.

- A new *ab initio* methodology for the design of optimal solvents for reactions has been proposed. An optimisation computer-aided molecular design formulation is used and quantum mechanical calculations are combined with a Kriging surrogate model to predict the kinetics.

- A general mathematical formulation for the problem of solvent design for reactions is presented, where quantum mechanical calculations for the reaction rate constant are integrated in a computer-aided molecular design formulation. The formulation is based on the works of Sheldon et al. [2006], Folić et al. [2007] and Struebing et al. [2013]. The difference here is that the rate constant is calculated from conventional transition state theory and quantum mechanical calculations are integrated in the CAMD framework.

- The proposed QM-CAMD-Kriging methodology has been successfully applied to the Menschutkin reaction of phenacyl bromide and pyridine, and to the Cope elimination reaction of methylamine oxide. Improved solvents were identified for both reactions; dichloromethanol and methylpentane were designed for the Menschutkin and the Cope reactions, respectively. The use of the Kriging model proved to be significantly beneficial compared to the empirical solvatochromic equation.

- The rate constant of the Menschutkin reaction of phenacyl bromide and pyridine in five organic solvents has been calculated with CTST and the SMD continuum

solvation model, using several levels of theories and basis sets. The M052X/6-31G(d) model was found to provide the best agreement with experiments.

7.3 Future work

7.3.1 Solubility

Currently the methodology for the *ab initio* design of solvents for reactions is based on the maximization of the reaction rate constant. In principle one is interested in the solvent that maximizes the reaction rate. This is a function of the reaction rate constant and the solubility of reactants in the solvent (assuming the reaction is performed at the solubility limit). The solubility of reactants can be calculated in one of many ways. Group contribution methods such as UNIFAC can be used for that purpose [Buxton et al., 1999]. Another possibility is to use an equation of state or a group-contribution-based equation of state as was done in the case of the design of the CO_2-expanded solvent for the Diels-Alder reaction. The solubility can also be estimated using an advanced thermodynamic model such as the SAFT-γ equation of state [Lymperiadis et al., 2007].

7.3.2 The solvent design space

In this thesis, a reduced set of structural groups has been included in the design space. By including more structural groups, the solvent design space would be enhanced and the variation of the solvent candidates would be improved. Moreover, it would be desirable to include single-group molecules, introduced by Struebing et al. [2013], as several widely used commercial solvents belong to that category, such as acetonitrile, dimethyl sulfoxide (DMSO) and nitromethane.

7.3.3 The Kriging model

For simplicity, the values of the Kriging parameters θ and p of the correlation matrix were restricted in specific ranges in this thesis. θ parameters were allowed to take

values up to 10, while in the original Kriging model there is no upper bound. Parameters p were set to be equal to 2, while they should range between 0 and 2. These assumptions were made to simplify the Kriging model and limit the computational time required to solve the MINLP problem. However, enabling the parameters to take all possible values would lead to more accurate Kriging models and possibly to faster convergence of the algorithm, this would compensate for the higher CPU time for the MINLP solution. Furthermore, Kriging surface regression, and thus the estimation of the Kriging parameters, is currently performed using a local NLP solver. It is reasonable to assume that the efficiency and accuracy of the methodology could be improved significantly if this problem was solved with a global optimisation algorithm.

7.3.4 Solvation models

It would be interesting to investigate the predictive capabilities of other popular continuum solvation models, such as COSMO-RS [Klamt, 1995]. In addition, the use of explicit solvation models could improve the accuracy of the predictions, especially in the case of protic solvents where the reliability of continuum models is not very satisfactory, as the system will be described in more detail, albeit at the cost of increased computational complexity [Truhlar, 2013]. Explicit solvation models could also make possible the design of solvent mixtures, as is discussed later in this section.

7.3.5 Mixed solvents

The advantages of solvent mixtures are well known and have already been discussed in this thesis. In Chapter 2, where the design of CO_2-expanded liquids was considered, the organic co-solvents were chosen through enumeration. The methodology was restricted by the availability of kinetic data, because an empirical kinetic model was used. The developed QM-CAMD-Kriging methodology does not need any data, but it considers only pure solvents. Therefore, it would be very useful to be able to also design mixed solvents. Although CAMD methods for mixed solvents have been developed before (see section 2.3), in this approach, a way has to be found to correlate

the solvent properties, that are used in the SMD solvation model, with the solvent mixture structure and composition. Another approach to this problem would be the use of explicit solvation models, where the solvent molecules can be explicitly defined.

7.3.6 Integration with process design

The problem of the design of the CO_2-expanded solvent, presented in Chapter 3, demonstrated how the specific characteristics of the process of interest can affect the solvent design problem. In this example the solvent design space was limited by available experimental data to three. On the other hand, the *ab initio* solvent design methodology is capable of assessing the suitability of a large number of solvents ($>$ 1000). It would be very interesting to consider also process requirements together with an *ab initio* calculation of the rate constant to develop an integrated solvent and process optimisation. In order to realise this, additional constraints have to be included in the CAMD formulation to account for the process in consideration (see section 2.5). As discussed in Chapter 2, a detailed environmental analysis is also necessary for complete solvent and process design.

Publications

Siougkrou E., Galindo A. and Adjiman C.S., "On the optimal design of gas-expanded liquids based on process performance", Chemical Engineering Science (2014), http://dx.doi.org/10.1016/j.ces.2013.12.025i

Struebing H., Ganase Z., Karamertzanis P.G., Siougkrou E., Haycock P., Piccione P.M., Armstrong A., Galindo A., Adjiman C.S., "Accelerating reaction kinetics in solution: Computer-aided molecular design and experimental validation", Nature Chemistry 5, 952-957 (2013)

Siougkrou E., Galindo A. and Adjiman C.S., "Integrated design of a reactor and a gas-expanded solvent", In: Pistikopoulos E.N., Georgiadis M.C., Kokossis A.C.

Computer-Aided Chemical Engineering 29, 316-320 (2011)

Conference Presentations

Siougkrou E., Galindo A. and Adjiman C.S., "QM-CAMD: Advances in computer-aided molecular design of solvents for reactions", AIChE 2013, San Francisco, USA - oral presentation

Siougkrou E., Galindo A. and Adjiman C.S., "QM-CAMD: Advances in computer-aided molecular design of solvents for reactions", PPEPPD 2013, Iguazu Falls, Argentina - poster presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Ab-initio and computer-aided molecular design for the identification of optimal solvents for reactions", AIChE 2012, Pittsburgh, USA - oral presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Ab-initio and computer-aided molecular design for the identification of optimal solvents for reactions", ISCRE 2012, Maastricht, The Netherlands - poster presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Ab-initio and computer-aided molecular design for the identification of optimal solvents for reactions", FOMMS 2012, Portland, USA - poster presentation

Siougkrou E., Galindo A. and Adjiman C.S., "9Integrated solvent and reaction system design using a CO_2-expanded liquid", IChemE 2011 Rational Catalyst and Process Design, Oxford, UK - poster presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Integrated solvent and reaction system design using a CO_2-expanded liquid", Thermodynamics 2011, Athens, Greece - short

oral presentation/poster presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Integrated design of a reactor and a gas-expanded solvent", 21st European Symposium on Computer Aided Process Engineering (ESCAPE 21), Chalkidiki, Greece - oral presentation

Siougkrou E., Galindo A. and Adjiman C.S., "Integrated solvent and reaction system design using a CO_2-expanded liquid", CAPE 2011, Bradford, UK - oral presentation

Bibliography

M. H. Abraham. Scales of solute hydrogen-bonding: Their construction and application to physicochemical and biochemical processes. *Chemical Society Reviews*, 22: 73–83, 1993a.

M. H. Abraham. Application of solvation equations to chemical and biochemical processes. *Pure and Applied Chemistry*, 65:2503–2512, 1993b.

M. H. Abraham, R. M. Doherty, M. J. Kamlet, J. M. Harris, and R. W. Taft. Linear solvation energy relationships. Part 38. An analysis of the use of solvent parameters in the correlation of rate constants, with special reference to the solvolysis of t-butyl chloride. *Journal of Chemical Society Perkin Transactions 2*, pages 1097–1101, 1987.

M. H. Abraham, G. S. Whiting an R. M. Doherty, and W. J. Shuely. XVI. a new solute solvation parameter, π_2^H from gas chromatographic data. *Journal of Chromatography*, 587:213–228, 1991.

O. Acevedo and W. L. Jorgensen. Cope elimination: Elucidation of solvent effects from QM/MM simulations. *Journal of American Chemical Society*, 128:6141–6146, 2006.

O. Acevedo and W. L. Jorgensen. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymantic reactions. *Accounts of Chemical Research*, 43:142–151, 2010.

L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors. *Computer Aided*

Molecular Design: Theory and Practice. Elsevier, Amsterdam, The Netherlands, 2003.

C. S. Adjiman, I. P. Androulakis, and C. A. Floudas. Global optimization of MINLP problems in process synthesis and design. *Computers and Chemical Engineering*, 21:S445–S450, 1997.

C. S. Adjiman, S. Dallwig, C. A. Floudas, and A. Neumaier. A global optimaization method, αBB, for general twice-differentiable constrained NLPs - I. Theoritical advances. *Computers and Chemical Engineering*, 22:1137–1158, 1998.

C. S. Adjiman, I. P. Androulakis, and C. A. Floudas. Global optimization of mixed-integer nonlinear problems. *AIChE Journal*, 46:1769–1797, 2000.

C. S. Adjiman, A. J. Clarke, G. Cooper, and P. C. Taylor. Solvents for ring-closing metathesis reactions. *Chemical Communications*, pages 2806–2808, 2008.

T. Adrian and G. Maurer. Solubility of carbon dioxide in acetone and propionic acid at temperatures between 298k and 333k. *Journal of Chemical and Engineering Data*, 42:668–672, 1997.

J. Ahlers and J. Gmehling. Development of a universal group contribution equation of state 3. Prediction of vapor-liquid equilibria, excess enthalpies, and activity coefficients at infinite dilution with the VTPR model. *Industrial and Engineering Chemistry Research*, 41:5890–5899, 2002.

J. Ahlers, T. Yamaguchi, and J. Gmehling. Development of a universal group contribution equation of state. 5. Prediction of the solubility of high-boiling compounds in supercritical gases with the group contribution equation of state Volume-Translated Peng-Robinson. *Industrial and Engineering Chemistry Research*, 43:6569–6576, 2004.

A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1935.

G. R. Akien and M. Poliakoff. A critical look at reactions in class I and II gas-expanded liquids using CO_2 and other gases. *Green Chemistry*, 11:1083–1100, 2009.

P. Akula, P.-M. Kleniati, and C. S. Adjiman. On the design of optimal solvent mixtures using generalised disjunctive programming. In I. D. L. Bogle and M. Fairweather, editors, *Computer-Aided Chemical Engineering*, volume 30, 2012.

I. P. Androulakis, C. D. Maranas, and C. A. Floudas. αBB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization*, 7:337–363, 1995.

M. Anouti, M. Caillon-Caravanier, Y. Dridi, J. Jacquemin, C. Hardacre, and D. Lemordant. Liquid densities, heat capacities, refractive index and excess quantities for protic ionic liquids + water binary system. *Journal of Chemical Thermodynamics*, 41:799–808, 2009.

M. Antosik, M. Galka, and S. K. Malanowski. Vapor-liquid equilibrium in a ternary system cyclohexane + ethanol + water. *Journal of Chemical Engineering Data*, 49: 7–10, 2004.

A. Apostolakou and C. S. Adjiman. Refrigerant design case study. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 289–301. Elsevier, Amsterdam, The Netherlands, 2003a.

A. Apostolakou and C. S. Adjiman. Optimization methods in CAMD - II. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 63–93. Elsevier, Amsterdam, The Netherlands, 2003b.

S. Arrhenius. *Zeitschrift für Physikalische Chemie*, 4:226–248, 1889.

R. W. Ashcraft, S. Raman, and W. H. Green. Predicted reaction rates of $H_xN_yO_z$ intermediates in the oxidation of hydroxylamine by aqueous nitric acid. *The Journal of Physical Chemstry A*, 112:7577–7593, 2008.

A. Aucejo, M.C. Burguet, R. Munoz , and J. L. Marques. Densities, viscosities, and refractive indices of the binary liquid systems n-alkanes + isomers of hexane at 298.15 K. *Journal of Chemical Engineering Data*, 40:871–874, 1995.

M. H. Back and K. J. Laidler. *Selected readings in chemical kinetics*. Pergamon, Oxford, 1967.

J. Barbosa, V. Sanz-Nebot, and I. Toro. Solvatochromic parameter values and pH in acetonitrile-water mixtures optimization of mobile phase for the separation of peptides by high-performance liquid chromatography. *Journal of Chromatography A*, 725:249–260, 1996.

A. Bardow, K. Steur, and J. Gross. Continuous-molecular targeting for integrated solvent and process design. *Industrial and Engineering Chemistry Research*, 49: 2834–2840, 2010.

P. W. C. Barnard and B. V. Smith. The Menschutkin reaction: A group experiment in a kinetic study. *Journal of Chemical Education*, 58(3):282–285, 1981.

J. Beck, D. Fridrich, S. Brandani, S. Guillas, and E. S. Fraga. Surrogate based optimisation for design of pressure swing adsorption systems. In I. D. L. Bogle and M. Fairweather, editors, *Computer-Aided Chemical Engineering*, volume 30, pages 1217–1221, 2012.

T. Benighaus and W. Thiel. A general boundary potential for hybrid QM/MM simulations of solvated biomolecular systems. *Journal of Chemical Theory and Computation*, 5:3114–3128, 2009.

M. Berthelot and L. Pean de Saint-Gilles. Recherches sur les affinités. De la formation et de la décomposition des éthers. *Annales de Chimie et de Physique, 3^e série*, 65: 385–422, 1862.

D. N. Bhattacharyya, C. L. Lee, J. Smid, and M. Szwarc. Studies of ions and ion pairs in tetrahydrofuran solution. Alkali metal salts of tetraphenylboride. *Journal of Physical Chemistry*, 69:608–611, 1965.

M. Born. Volumen und hydratationswärme der ionen. *Zeitschrift für Physik A Hadrons and Nuclei*, 1:45–48, 1920.

M. Born. Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58:1486–1493, 1936.

Y. Y. Borovikov, M. I. Shman'ko, and L. I. Zolotun. Physicochemical investigation of the interaction of aldehydes with various proton donors. *Russian Journal of General Chemistry*, 47:1739–1744, 1977.

E. Bosch, F. Rived, and M. Rosés. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 4. Preferential solvation of solvatochromic indicators in mixtures of 2-methylpropan-2-ol with hexane, benzene, propan-2-ol, ethanol and methanol. *Journal of Chemical Society Perkin Transactions 2*, pages 2177–2184, 1996a.

E. Bosch, M. Rosés, K. Herodes, I. Koppel, and V. Taal. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 2. Effect of temperature on the E(t)-(30) polarity parameter of dipolar hydrogen bond acceptor-hydrogen bond donor mixtures. *Journal of Physical Organic Chem.*, 9:403–410, 1996b.

F. Boukouvala, F. J. Muzzio, and M. G. Ierapetritou. Predictive modeling of pharmaceutical processes with missing and noisy data. *AIChE Journal*, 56:2860–2872, 2010.

E. P. Box and N. R. Draper. *Empirical model building and response surfaces*. New York: McGraw-Hill, 1987.

G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society Series B*, 13:1–45, 1951.

G. Brancato, N. Rega, and V. Barone. A hybrid explictit/implicit solvation method for first-principle molecular dynamics simulations. *Journal of Chemical Physics*, 128:144501, 2008.

E. A. Brignole and M. Cismondi. Molecular design-generation and test methods. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 23–41. Elsevier, Amsterdam, The Netherlands, 2003.

E. A. Brignole, S. Bottini, and R. Gani. A strategy for the design and selection of solvents for separation processes. *Fluid Phase Equilibria*, 29:125–132, 1986.

D. S. Broomhead and D. Loewe. Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

A. Bruylants. *Chemical Abstracts*, 87:166770x, 1977.

A. H. Buep and M. Baron. Dielectric properties of binary systems. 7. Carbon tetrachloride with benzene, with toluene, and with p-xylene at 298.15 and 308.15 K. *Journal of Physical Chemistry*, 92:840–843, 1988.

U. Buhvestov, F. Rived, C. Ráfols, E. Bosch, and M. Rosés. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 7. Comparison of the enhancement of the water structure in alcohol-water mixtures measured by solvatochromic indicators. *Journal of Physical Organic Chemistry*, 11:185–192, 1998.

E. Buncel, R. Stairs, and H. Wilson. *The role of the solvent in chemical reactions.* Oxford University Press, 2003.

A. Buxton, A. G. Livingston, and E. N. Pistikopoulos. Optimal design of solvent blends for environmental impact minimization. *AIChE Journal*, 45:817–843, 1999.

J. A. Caballero and I. E. Grossmann. An algorithm for the use of surrogate models in modular flowshhet optimization. *AIChE Journal*, 54:2633–2650, 2008.

K. V. Camarda and C. D. Maranas. Optimization in polymer design using connectivity indices. *Industrial and Engineering Chemistry Research*, 38:1884–1892, 1999.

K. V. Camarda and P. Sunderesan. An optimization approach to the design of value-added soybean oil products. *Industrial and Engineering Chemistry Research*, 44: 4361–4367, 2005.

E. Cancès, B. Mennucci, and J. Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *Journal of Chemical Physics*, 107:3030–3041, 1997.

R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55:2471–2474, 1985.

R. Carlson. *Design and optimization in organic synthesis*. Elsevier, Amsterdam, The Netherlands, 1992.

C. Cativiela, J. I. García, J. Gil, R. M. Martínez, J. A. Mayoral, L. Salvatella, J. S. Urieta, A. M. Mainar, and M. H. Abraham. Solvent eVects on Diels-Alder reactions. The use of aqueous mixtures ox Xuorinated alcohols and the study of reactions of acrylonitrile. *Journal of Chemical Society Perkin Transactions 2*, pages 653–660, 1997.

E. A. Cepeda and M. Diaz. Solubility of anthracene and anthraquinone in acetonitrile, methyl ethyl ketone, isopropyl alcohol and their mixtures. *Fluid Phase Equilibria*, 121:267–272, 1996.

C. J. Chang and A. D. Randolph. Separation of β-carotene mixtures precipitated from liquid solvents with high-pressure CO_2. *Biotechnology Progress*, 7:275–278, 1991.

J.-S. Chang and D.-H. Chen. Optimization on the etherification of glycerol with tert-butyl alcohol. *Journal of the Taiwan Institute of Chemical Engineers*, 42:760–767, 2011.

J. Chen and J. Gmehling. Modification of PSRK mixing rules and some results for vapor-liquid equilibria, heats of mixing and activity coefficients at infinite dilution. *Fluid Phase Equilibria*, 200:411–429, 2002.

H. C. Cheng and F. S. Wang. Computer-aided biocompatible solvent design for an integrated extractive fermentation-separation process. *Chemical Engineering Journal*, 162:809–820, 2010.

J. R. Chipperfield, editor. *Non-aqueous solvents*. Oxford University Press, Oxford, 1999.

N. Churi and L. E. K. Achenie. Novel mathematical programming model for computer aided molecular design. *Industrial and Engineering Chemistry Research*, 35:3788–3794, 1996.

E. Ciganek, J. M. Read, and J. C. Calabrese. Reverse Cope elimination reactions. 1. Mechanism and scope. *Journal of Organic Chemistry*, 60:5795–5802, 1995.

M. Cismondi and E. A. Brignole. Molecular design of solvents: An efficient algorithm for branched molecules. *Industrial and Engineering Chemistry Research*, 43:784–790, 2004.

J. H. Clark and S. J. Tavener. Alternative solvents: Shades of green. *Organic Process Research and Development*, 11:149–155, 2007.

L. Constantinou and R. Gani. New group-contribution method for estimating properties of pure compounds. *AICHE Journal*, 40:1697–1710, 1994.

L. Constantinou, R. Gani, and J. P. OĆonnell. Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method. *Fluid Phase Equilibria*, 103:11–22, 1995.

E. Conte, A. Martinho, H. A. Matos, and R. Gani. Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity. *Industrial and Engineering Chemistry Research*, 47:7940–7654, 2008.

K. A. Cooper, M. L. Dhar, E. D. Hughes, C. N. Ingold, B. J. MacNulty, and L. I. Woolf. Mechanism of elimination reactions. Part VII. Solvent effects on rates and product-proportions in uni- and bi-molecular substitution and elimination reactions

of alkyl halides and sulphonium salts in hydroxylic solvents. *Journal of Chemical Society*, pages 2043–2049, 1948.

A. C. Cope, T. T. Foster, and P. H. Towle. Thermal decomposition of amine oxides to olefins and dialkylhydroxylamines. *Journal of Organic Chemistry*, 71:3929–3934, 1949.

J. L. Cordiner. CAMD for solvent selection in industry - II. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 229–246. Elsevier, Amsterdam, The Netherlands, 2003.

I. Couckuyt, A. Forrester, D. Gorissen, F. De Turck, and T. Dhaene. Blind kriging: Implementation and performance analysis. *Advances in Engineering Software*, 49: 1–13, 2012.

M. M. Cowan. Plant products as antimicrobial agents. *Clinical Microbiology Reviews*, 12:564–582, 1999.

C. J. Cramer. *Essentials of Computational Chemistry. Theory and Models*. WILEY, second edition, 2004.

C. J. Cramer and D. G. Truhlar. General parameterized SCF model for free energies of solvation in aqueous solution. *Journal of the American Chemical Society*, 13: 8305–8311, 1991.

C. J. Cramer and D. G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chemical Reviews*, 99:2161–2200, 1999.

C. J. Cramer and D. G. Truhlar. SMx continuum models for condensed phases. In G. Maroulis and T. E. Simos, editors, *Trends and Perspectives in Modern Computational Science*, volume 6, pages 112–140. Brill/VSP, Leiden, 2006.

C. J. Cramer and D. G. Truhlar. A universal approach to solvation modeling. *Accounts of Chemical Research*, 41:760–768, 2008.

N. A. C. Cressie, editor. *Statistics for spatial data*. WILEY Interscience, revised edition, 1993.

Q. Cui, M. Elstner, and M. Karplus. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *Journal of Physical Chemistry B*, 106:2721–2740, 2002.

C. Curutchet, M. Orozco, and F. J. Luque. Solvation in octanol: Parametrization of the continuum MST model. *Journal of Computational Chemistry*, 22:1180–1193, 2001.

A. D. Curzons, D. J. C. Constable, and V. L. Cunningham. Solvent selection quide: a quide to the integration of the environmental, health and safety criteria into the selection of solvents. *Clean Products Process*, 1:82–90, 1999.

M. J. Davila and J. P. M. Trusler. Thermodynamic properties of mixtures of n-methyl-2-pyrrolidinone and methanol at temperatures between 298.15 K and 343.15 K and pressures up to 60 MPa. *Journal of Chemical Thermodynamics*, 41:35–45, 2009.

E. Davis and M. Ierapetritou. A kriging method for the solution of nonlinear programs with black-box functions. *AIChE Journal*, 53:2001–2012, 2007.

E. Davis and M. Ierapetritou. A kriging based method for the solution of mixed-integer nonlinear programs containing black-box functions. *Journal of Global Optimization*, 43:191–205, 2009.

R. K. Dewan and S. K. Mehta. Correlation between topological features and surface tension of binary liquid mixtures. *Monatshefte für Chemie*, 121:593–600, 1990.

J. M. Douglas. *Conceptual Design of Chemical Processes*. McGraw-Hill, 1988.

M. A. Duran and I. E. Grossmann. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36:307–339, 1986.

M. R. Eden, S. B. Jørgensen, R. Gani, and M. M. El-Halwagi. A novel framework for simultaneous separation process and product design. *Chemical Engineering and Processing*, 43:595–608, 2004.

B. Eliasson, D. Johnels, S. Wold, U. Edlund, and M. Sjöström. On the correlation between solvent scales and solvent-induced C-13 NMR chemical-shifts of a planar lithium carbanion - A multivariate data-analysis using a principal component-multiple-regression-like formalism. *Acta Chemica Scandinavica Series B - Organic Chemistry and Biochemistry*, 36:155–164, 1982.

M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Transactions of the Faraday Society*, 31:875, 1935.

M. B. Ewing and J. C. Sanchez Ochoa. Vapor pressures of acetonitrile determined by comparative ebulliometry. *Journal of Chemical Engineering Data*, 49:486–491, 2004.

O. Exler and K. Schittkowski. A trust region SQP algorithm for mixed-integer nonlinear programming. *Optimization Letters*, 1:269–280, 2007.

H. Eyring. The activated complex in chemical reactions. *Journal of Chemical Physics*, 3:107–115, 1935.

J. Fages, H. Lochard, J. J. Letourneau, M. Sauceau, and E. Rodier. Particle generation for pharmaceutical applications using supercritical fluid technology. *Powder Technology*, 141:219–226, 2004.

J. Fang, H. Jin, T. Ruddy, K. Pennybaker, D. Faley, and B. Subramaniam. Economic and environmental impact analyses of catalytic olefin hydroformylation in CO_2-expanded liquid (CXL) media. *Industrial and Engineering Chemistry Research*, 46:8687–8692, 2007.

Y. Y. Fialkov, Y. A. Tarasenko, and Y. Y. Borovikov. Physico-chemical analysis

of amine acetate - indifferent solvent systems. *Zhurnal Organicheskoi Khimii*, 36: 981–987, 1966.

C. A. Floudas, editor. *Nonlinear and mixed-integer optimization: fundamentals and applications*. Oxford University Press, 1995.

M. Folić, C. S. Adjiman, and E. N. Pistikopoulos. The design of solvents for optimal reaction rates. In A. P. Barbosa Povoa and H. Matos, editors, *Computer-Aided Chemical Engineering*, volume 18, pages 175–180, 2004.

M. Folić, C. S. Adjiman, and E. N. Pistikopoulos. A computer-aided methodology for optimal solvent for reactions with experimental verification. In L. Puigjaner and A. Espuna, editors, *Computer-Aided Chemical Engineering*, volume 20a-20b, pages 1651–1656, 2005.

M. Folić, C. S. Adjiman, and E. N. Pistikopoulos. A computer-aided methodology with robust design criteria for selection of solvents for reactions. In W. Marquardt and C. Pantelides, editors, *Computer-Aided Chemical Engineering*, volume 21, pages 787–792, 2006.

M. Folić, C. S. Adjiman, and E. N. Pistikopoulos. Design of solvents for optimal reaction rate constants. *AIChE Journal*, 53:1240–1256, 2007.

M. Folić, C. S. Adjiman, and E. N. Pistikopoulos. Computer-aided solvent design for reactions: Maximizing product formation. *AIChE Journal*, 47:5190–5202, 2008a.

M. Folić, R. Gani, C. Jiménez-González, and D. J. C. Constable. Systematic selection of green solvents for organic solvents for organic reactive systems. *Chinese Journal of Chemical Engineering*, 16:376–383, 2008b.

J. W. Ford, M. E. Janakat, J. Lu, C. L. Liotta, and C. A. Eckert. Local polarity in co_2-expanded acetonitrile: A nucleophilic substitution reaction and solvatochromic probes. *American Chemical Society*, 73:364–3368, 2008a.

J. W. Ford, J. Lu, C. L. Liotta, and C. A. Eckert. Solvent effects on the kinetics of a Diels-Alder reaction in gas-expanded liquids. *Industrial and Engineering Chemistry Research*, 47:632–637, 2008b.

J. B. Foresman and Æ . Frisch, editors. *Exploring chemistry with electronic structure methods*. Gaussian Inc., Pittsburgh, PA, second edition, 1996.

A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45:50–79, 2009.

W Forster and R M Laird. The mechanism of alkylation reactions. Part 1. The effect of substituents on the reaction of phenacyl bromide with pyridine in methanol. *Journal of the Chemical Society, Perkin Transactions* 2, pages 135–138, 1982.

A. Fredenslund, R. L. Jones, and J. M. Prausnitz. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21:1086–1099, 1975.

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision E.01, 2004. Gaussian, Inc., Wallingford, CT.

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision A.1, 2009. Gaussian, Inc., Wallingford, CT.

A. Galindo, L. A. Davies, A. Gil-Villegas, and G. Jackson. The thermodynamics of mixtures and the corresponding mixing rules in the saft-vr approach for potentials of variable range. *Molecular Physics*, 93:241–252, 1998.

Z. Ganase, A. Armstrong, A. Galindo, and C. S. Adjiman. In preparation. 2013.

R. Gani. Chemical product design: challenges and opportunities. *Computers and Chemical Engineering*, 28:2441–2457, 2004.

R. Gani and E. A. Brignole. Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilibria*, 13:331–340, 1983.

R. Gani, C. Jiménez-González, and D. J. C. Constable. Method for selection of solvents for promotion of organic reactions. *Computers and Chemical Engineering*, 29:1661–1676, 2005.

R. Gani, P. A. Gómez, M. Folić, C. Jiménez-González, and D. J. C. Constable. Solvents in organic synthesis: Replacement and multi-step reaction systems. *Computers and Chemical Engineering*, 32:2420–2444, 2008.

B. C. Garrett and D. G. Truhlar. Criterion of minimum state density in the transition-state theory of bimolecular reactions. *Journal of Chemical Physics*, 70:1593–1598, 1979a.

B. C. Garrett and D. G. Truhlar. Generalized transition-state theory - Classical mechanical theory an applications to collinear reactions of hydrogen molecules. *Journal of Physical Chemistry*, 83:1052–1078, 1979b.

M. Ghanta, D. R. Faley, D. H. Buscch, and B. Subramaniam. Comparative economic and environmental assessments of H_2O_2-based and tertiary butyl hydroperoxide-based propylene oxide technologies. *ACS Sustainable Chemistry and Engineering*, 1:268–277, 2012a.

M. Ghanta, T. Ruddy, D. R. Faley, D. H. Buscch, and B. Subramaniam. Is the liquid-phase H_2O_2-based ethylene oxide process more economical and greener that the gas-phase O_2-based silver-catalyzed process? *Industrial and Engineering Chemistry Research*, 52:1–29, 2012b.

A. Gil-Villegas, A. Galindo, P. J. Whitehead, S. J. Mills, G. Jackson, and A. N. Burgess. Statistical associating fluid theory for chain molecules with attractive potentials of variable range. *Journal of Chemical Physics*, 106:4168–4186, 1997.

A. Ghanadzadeh Gilani, N. Paktinat, and M. Moghadam. Relative permittivity data of binary mixtures containing 2-butanol, 2-butanone, and cyclohexane. *Journal of Chemical Thermodynamics*, 43:569–575, 2011.

P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47:99–131, 2005.

A. Giovanoglou, J. Barlatier, C. S. Adjiman, E. N. Pistikopoulos, and J. L. Cordiner. Optimal solvent design for batch separation based on economic performance. *AIChE Journal*, 49:3095–3109, 2003.

J. Gmehling, R. Witting, J. Lohmann, and R. Joh. A modified UNIFAC (Dortmund)

model 4. Revision and extention. *Industrial and Engineering Chemistry Research*, 41:1678–1688, 2002.

K. Gong, S. Chafin, K. Pennybaker, D. Faley, and B. Subramaniam. Economic and environmental impact analyses of solid acid catalyzed isoparaffin/olefin alkylation in supercritical carbon dioxide. *Industrial and Engineering Chemistry Research*, 47:9072–9080, 2008.

A. Habibi-Yangjeh. A model for correlation of various solvatochromic parameters with composition in aqueous and organic binary solvent systems. *Bulletin of the Korean Chemical Society*, 25:1165–1170, 2004.

A. Halvorsen and J. Songstad. The reactivity of 2-bromo-1-phenylethanone (phenacyl bromide) toward nucleophilic species. *Journal of Chemical Society, Chemical Communications*, pages 327–328, 1978.

L. P. Hammett. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *Journal of American Chemical Society*, 59:96–103, 1937.

A. A. Hanna. Molar volume contraction for alcohols in acetic acid. *Journal of Chemical Engineering Data*, 29:75–78, 1984.

A. R. Harifi-Mood, A. Habibi-Yangjeh, and M. R. Gholami. Solvatochromic parameters for binary mixtures of 1-(1-butyl)-3-methylimidazolium tetrafluoroborate with some protic molecular solvents. *Journal of Physical Chemistry B*, 110:7073–7078, 2006.

A. R. Harifi-Mood, A. Habibi-Yangjeh, and M. R. Gholami. Solvent effects on kinetics of the reaction between 2-chloro-3,5-dinitropyridine and aniline in aqueous and alcoholic solutions of [bmim]BF4. *International Journal of Chemical Kinetics*, 39: 681–687, 2007.

P. M. Harper, M. Hostrup, and R. Gani. A hybrid CAMD method. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular*

Design: Theory and Practice, volume 12, pages 139–165. Elsevier, Amsterdam, The Netherlands, 2003.

M. M. Faruque Hasan, R. C. Baliban, J. A. Elia, and C. A. Floudas. Modeling, simulation, and optimization of postcombustion co2 capture for variable feed concentration and flow rate. 2. Pressure swing adsorption and vacuum swing adsorption processes. *Industrial and Engineering Chemistry Research*, 51:15665–15682, 2012.

G. I. Hawe, I. Alkorta, and P. L. A. Popelier. Prediction of the basicities of pyridines in the gas phase and in aqueous solution. *Journal of Chemical Information and Modeling*, 50:87–96, 2010.

N. E. Henriksen and F. Y. Hansen, editors. *Theories of molecular reaction dynamics.* Oxford University Press, first edition, 2008.

K. Herodes, I. Leito, I. Koppel, and M. Rosés. Solute-solvent and solvent-solvent interactions in binary solvent mixtures.8. The E-T(30) polarity of binary mixtures of formamides with hydroxylic solvents. *Journal of Physical Organic Chemistry*, 12:109–115, 1999.

M. Higashi, A. V. Marenich, R. M. Olson, A. C. Chamberlin, J. Pu, C. P. Kelly, J. D. Thompson, J. D. Xidos, J. Li, T. Zhu, G. D. Hawkins, Y. Y. Chuang, P. L. Fast, B. J. Lynch, D. A. Liotard, D. Rinaldi, J. Gao, C. J. Cramer, and D. G. Truhlar. *GAMESSPLUS – version* $2008 - 2$. University of Minnesota, Minneapolis, based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. 14 (1993) 1347, 2008.

T. L. Hill. *An introduction to statistical thermodynamic.* Dover Publications Inc., New York, 1986.

H. Hiratsuka, Y. Kadokura, H. Chida, M. Tanaka, S. Kobayashi, T. Okutsu, M. Obab, and K. Nishiyamab. Photochemical processes of benzyltrimethylsilane at 77 K.

Remarkable solvent effects and reaction mechanism. *Journal of Chemical Society, Faraday Transactions*, 92:3035–3041, 1996.

J. Ho, A. Klamt, and M. L. Coote. Comment on the correct use of continuum solvent models. *Journal of Physical Chemistry A*, 114:13442–13444, 2010.

D. Hopfe. Thermophysical data of pure substances. In *Data Compilation of FIZ CHEMIE Germany*, 1990.

M. Hostrup, R. Gani, Z. Kravanja, A. Sorsak, and I. Grossmann. Integration of thermodynamic insights and MINLP optimization for the synthesis, design and analysis of process flowsheets. *Computers and Chemical Engineering*, 25:73–83, 2001.

L. Hovarth. Molecular design. Chemical structure generation from the properties of pure organic compounds. In *Studies in Physical and Theoritical Chemistry Book Series*, volume 75. Elsevier, Amsterdam, The Netherlands, 1992.

L. A. Huck and W. J. Leigh. Kinetic and mechanistic studies of the reactions of diarylgermylenes and tetraaryldigermenes with carbon tetrachloride. *Canadian Journal of Chemistry*, 89:241–255, 2011.

E. D. Hughes and C. N. Ingold. Mechanism of substitution at a saturated carbon atom. Part IV. A discussion of constitutional and solvent effects on the mechanism, kinetics, velocity, and orientation of substitution. *Journal of Chemical Society*, pages 244–255, 1935.

A. S. Hukkerikar, S. Kalakul, B. Sarup, D. M. Young, G. Sin, and R. Gani. Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution$^+$ (GC^+) property models and uncertainty analysis. *Journal of Chemical Information and Modelling*, 52:2823–2839, 2012.

J. U. Hwang, J. J. Chung, S. D. Yoh, and J. G. Jee. Kinetics for the reaction of

phenacyl bromide with pyridine in acetone under high pressure. *Bulletin of the Korean Chemical Society*, 4:273–240, 1983.

R. J. Hwang and J. I. Ortiz. Mitigation of asphaltics deposition during CO_2 flood by enhancing CO_2 solvency with chemical modifiers. *Journal of Organic Geochemistry*, 31:1451–1462, 2000.

IBM. International Business Machines Corporation.

ICAS documentation. *Internal report PEC02-14*.

T. P. Iglesias, J. L. Legido, S. M. Pereira, B. de Cominges, and M. I. Paz Andrade. Relative permittivities and refractive indices on mixing for (n-hexane + 1-pentanol, or 1-hexanol, or 1-heptanol) at T=298.15 K. *Journal of Chemical Thermodynamics*, 32:923–930, 2000.

International Union of Pure and Applied Chemistry. *Quantities, units and symbols in physical chemistry*. Blackwell Science Ltd, 1993.

A. Jalan, R. W. Ashcraft, R. H. West, and W. H. Green. Predicting solvation energies for kinetic modeling. *Annu. Rep. Prog. Chem., Sec. C*, 106:211–258, 2010.

J. Jamali-Paghaleh, A. R. Harifi-Mood, and M. R. Gholami. Reaction kinetics investigation of 1-fluoro-2,4-dinitrobenzene with substituted anilines in ethyl acetate-methanol mixtures using linear and nonlinear free energy relationships. *Journal of Physical Organic Chemistry*, 24:1095–1100, 2011.

V. Jehlicka. *Liquid Permitivity Database*. Prague Institute of Chemical Technology, 1990.

F. Jensen. *Introduction to computational chemistry*. WILEY, second edition, 2007.

P. G. Jessop and B. Subramaniam. Gas-expanded liquids. *Chemical Reviews*, 107: 2666–2694, 2007.

C. Jiménez-González, A. D. Curzons, D. J. C. Constable, and V. L. Cunningham. Expanding GSK's solvent selection guide - application of the life cycle assessment to enhance solvent selections. *Clean Technologies and Environmental Policy*, 7: 42–50, 2005a.

C. Jiménez-González, A. D. Curzons, D. J. C. Constable, and V. L. Cunningham. Expanding GSK's solvent selection guide - application of life cycle assessment to enhance solvent selections. *Clean Technologies and Environmental Policy*, 7:42–50, 2005b.

C. Jiménez-González, P.Poechlauer, Q. B. Broxterman, B. Yang, D. Ende, J. Baird, C. Bertsch, R. E. Hannah, P. Dell'Orco, H. Noorman, S. Yee, R. Reintjens, A. Wells, V. Massonneau, and J. Manley. Key green engineering research areas for sustainable manufacturing: A perspective from pharmaceutical and fine chemicals manufacturers. *Organic Process Research and Development*, 15:900–911, 2011.

D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21:345–383, 2001.

V. R. Joseph, Y. Hung, and A. Sudjianto. Blind kriging: A new method for developing metamodels. *Journal of Mechanical Design*, 130, 2008.

J. Jung and M. Perrut. Particle design using supercritical fluids: Literature and patent survey. *Journal of Supercritical Fluids*, 20:179–219, 2001.

O. E. Kalinovskaya and J. K. Vij. Molecular dynamics of iso-amyl bromide by dielectric spectroscopy, and the effects of a nonpolar solvent, 2-methylpentane, on the spectral features. *Journal of Chemical Physics*, 111:10979–10985, 1999.

S. C. L. Kamerlin, C. L. Shina, M. Haranczyk, and A. Warshel. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pK(a), redox reactions and rolvation free energies. *Journal of Physical Chemistry B*, 113:1253–1272, 2009.

M. J. Kamlet and R. W. Taft. Solvatochromic comparison method.1.Beta-scale of solvent hydrogen-bond acceptor(HBA) basicities. *Journal of American Chemical Society*, 98:377–383, 1976.

V. Karantzi, X. Qin, M. El Halwagi andF. Eljack, and M. Eden. Simultaneous process and molecular design through property clustering techniques: A visualization tool. *Industrial and Engineering Chemistry Research*, 46:3400–3409, 2007.

A. T. Karunanithi, L. E. K. Achenie, and R. Gani. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Industrial and Engineering Chemistry Research*, 44: 4785–4797, 2005.

J. C. Keck. Variational theory of chemical reaction rates applied to three-body recombinations. *Journal of Chemical Physics*, 32:1035–1050, 1960.

P. Khare, B. P. Baruah, and P. G. Rao. Application of chemometrics to study the kinetics of coal. *Fuel*, 90:3299–3305, 2011.

J.R. Khurma, O. Muthu, S. Munjal, and B. D. Smith. Total-pressure vapor-liquid equilibrium data for binary systems of 1-chlorobutane with ethyl acetate, acetonitrile, nitromethane and acetone. *Journal of Chemical Engineering Data*, 28:86–93, 1983.

J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *Journal of Chemical Physics*, 58:351–361, 1934.

A. Klamt. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *Journal of Physical Chemistry*, 99:2224–2235, 1995.

A. Klamt and G. Schüürmann. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of Chemical Society, Perkin Transactions 2*, pages 799–805, 1993.

A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco, and J. Luque. Commentary on the performance of continuum solvation methods. a comment on "universal approaches to solvation modeling". *Accounts of Chemical Research*, 42:489–492, 2009.

J. A. Klein, D. T. Wu, and R. Gani. Computer aided mixture design with specified property constraints. *Computers and Chemical Engineering*, 16:S229–S236, 1992.

J. L. Klepeis and C. A. Floudas. Comparative study of global minimum energy conformations of hydrated peptides. *Journal of Computational Chemistry*, 20:636–654, 1999.

J. L. Klepeis, M. J. Pieja, and C. A. Floudas. Hybrid global optimization algorithms for protein structure prediction: Alternating hybrids. *Biophysical Journal*, 84:869–882, 2003.

D. Kodama, N. Kubota, Y. Yamaki, H. Tanaka, and M. Kato. High pressure vapor-liquid equilibria and density behaviors for carbon dioxide + methanol system at 313.15 k. *Netsu Bussei*, 10:16–20, 1996.

H J Koh, K L Han, H W Lee, and I Lee. Kinetics and mechanism of the pyridinolysis of phenacyl bromides in acetonitrile. *Journal of Organic Chemistry*, 65:4706–4711, 2000.

I. Komaromi and J. M. J. Tronchet. Quantum chemical reaction path and transition state for a model cope (and reverse cope) elimination. *Journal of Physical Chemistry A*, 101:3554–3560, 1997.

A. Kordikowski, A. P. Schenk, R. M. Van Nielen, and C. J. Peters. Volume expansions and vapor-liquid equilibria of binary mixtures of a variety of polar solvents and certain near-critical solvents. *Journal of Supercritical Fluids*, 8:205–216, 1995.

J. Kozlowski, A. Maciejewski, M. Milewski, and W. Urjasz. Photochemical reactions of T_1 benzopyranthione in 3-methylpentane. *Journal of Physical Organic Chemistry*, 12:47–52, 1999.

D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52:119–139, 1951.

H. Kwart and M. Brechbiel. Role of solvent in the mechanism of amine oxide thermolysis elucidated by the temperature dependence of a kinetic isotope effect. *Journal of American Chemical Society*, 103:4650–4652, 1981.

H. Kwart, T. J. George, R. Louw, and W. Ultee. Transition-state structures in sulfoxide and amine oxide thermolysis. *Journal of American Chemical Society*, 100:3927–3928, 1978.

K. J. Laidler. *Chemical Kinetics*. Harper & Row Publishers, New York, third edition, 1987.

B. Lee and F. M. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55:379–400, 1971.

B. Lee and F. M. Richards. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *Journal of Physical Chemistry*, 76:2754–2759, 1972.

P. Lek-Utaiwan, B. Suphanit, N. Mongkolsir, and R. Gani. Integrated design of solvent-based extractive separation processes. In B. Braunschweig and X. Joulia, editors, *Computer-Aided Chemical Engineering*, volume 25, pages 121–126, 2008.

P. Lek-Utaiwan, B. Suphanit, N. Mongkolsiri, and R. Gani. Integrated design of solvent-based extractive separation processes including experimental validation. In *European Symposium on Computer Aided Process Engineering -19*, page 201, 2009.

C. T. Liao, W. J. Tzeng, , and F. S. Wang. Mixed-integer hybrid differential evolution for synthesis of chemical processes. *Journal of the Chinese Institute of Chemical Engineers*, 32:491–502, 2001.

G. Liessmann, W. Schmidt, and S. Reiffarth. Recommended thermophysical data. In *Data compilation of the Saechsische Olefinwerke Boehlen Germany*, 1995.

H. Lin and D. G. Truhlar. QM/MM: What have we learned, where are we, and where do we go from here? *Theoritical Chemistry Accounts*, 117:185–199, 2007.

A. Lymperiadis, C. S. Adjiman, A. Galindo, and G. Jackson. A group contribution method for associating chain molecules based on the statistical associating fluid theory (saft-γ). *Journal of Chemical Physics*, 127:234903, 2007.

C. D. Maranas. Optimal computer-aided molecular design: A polymer design case study. *Industrial and Engineering Chemistry Research*, 35:3403–3414, 1996.

E. C. Marcoulaki and A. C. Kokossis. On the development of novel chemical using a systematic optimisation approach. Part II. Solvent design. *Chemical Engineering Science*, 55:2547–2561, 2000.

Y. Marcus, editor. *Solvent mixtures - Properties and selective solvation*. Marcel Dekker, New York, Basel, 2002.

A. Marenich, C. J. Cramer, and D. G. Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by bulk dielectric constant and atomic surface tensions. *Journal of Physical Chemistry B*, 113:6378–6396, 2009.

J. Marrero and R. Gani. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183:183–208, 2001.

G. Matheron. Principles of geostatistics. *Economic Geology*, 58:1246–1266, 1963.

D.V. Matyushov and R. Schmid. Properties of liquids at the boiling point: Equation of state, internal pressure and vaporization entropy. *Ber. Bunsen-Ges.*, 98:1590–1595, 1994.

ARKI Consulting and Development. URL `www.conopt.com`.

Chemical Engineering Magazine, December 2010.

GAMS Development Corporation, 2011. URL `www.gams.com`.

Process Systems Enterprise, 1997-2009. URL `www.psenterprise.com/gproms`.

D. A. McQuarrie. *Statistical mechanics*. University Science Books, Californis, 2000.

N. Menschutkin. Beiträge zur kenntnis der affinitätskoeffizienten der alkylhaloide und der organischen amine. *Zeitschrift für Physikalische Chemie*, 5:589–600, 1890a.

N. Menschutkin. Über die affinitätskoeffizienten der alkylhaloide und der amine. *Zeitschrift für Physikalische Chemie*, 6:41–57, 1890b.

S. Miertuš, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum - A direct utilization of ab-initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, 55:117–129, 1981.

A. A. Mirzaliev, Sh. G. Shakhmuradov, and S. O. Guseinov. Investigation of the isobaric heat capacity of nitriles at various temperatures. *Izv. Vyssh. Uchebn. Zaved. Neft Gaz*, pages 55–58, 1987.

B. R. Nielsen, R. Gani, and A. Fredenslund. Computer aided molecular design by group contribution method. In H. T. Bussemaker and P. D. Iedema, editors, *Computer Applications in Chemical Engineering, Process Technology Proceedings*, volume 9, pages 227–232. Elsevier, Amsterdam, The Netherlands, 1990.

J. B. Nikolić, G. S. Ušćumlić, and I. O. Juranić. Solvent and structural effects on the kinetics of the reactions of 2-substituted cyclohex-1-enylcarboxylic and 2-substituted benzoic acids with diazodiphenylmethane. *International Journal of Chemical Kinetics*, 39:664–671, 2007.

O. Odele and S. Macchietto. Computer-aided molecular design; A novel method for optimal solvent selection. *Fluid Phase Equilibria*, 82:47–54, 1993.

M. Orozco and F. J. Luque. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chemical Reviews*, 100:4187–4225, 2000.

J. Ortega, C. Ràfols, E. Bosch, and M. Rosés. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 3. The E(t)-(30) polarity of binary mixtures of hydroxylic solvents. *Journal of Chemical Society Perkin Transactions 2*, pages 1497–1503, 1996.

A. I. Papadopoulos and P. Linke. Multiobjective molecular design for integrated process-solvent systems synthesis. *AIChE Journal*, pages 1057–1070, 2006.

A. I. Papadopoulos, M. Stijepovic, P. Linke, P. Seferlis, and S. Voutetakis. Toward optimim working fluid mixtures for organic rankine cycles using molecular design and sensitivity analysis. *Industrial and Engineering Chemistry Research*, 52:12116–12133, 2013.

S. J. Patel, D. Ng, and M. S. Mannan. Inherently safer design of solvent processes at the conceptual stage: Practical application for substitution. *Journal of Loss Prevention in the Process Industries*, 23:483–491, 2010.

P. R. Patkar and V. Venkatasubramanian. Polymer design case study. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 303–317. Elsevier, Amsterdam, The Netherlands, 2003a.

P. R. Patkar and V. Venkatasubramanian. Genetic algorithms based CAMD. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 95–128. Elsevier, Amsterdam, The Netherlands, 2003b.

R G Pearson, S H Langer, F V Williams, and W J McGuire. Mechanism of the reaction of α-Haloketones with weakly basic nucleophilic reagents. *Journal of the American Chemical Society*, 74:5130–5132, 1952.

A. Peneloux, E. Rauzy, and R. Freze. A consistent correction for Redlich-Kwong-Soave volumes. *Fluid Phase Equilibria*, 8:7–23, 1982.

F. E. Pereira, E. Keskes, A. Galindo, G. Jackson, and C. S. Adjiman. Integrated solvent and process design using a saft-vr thermodynamic description: High-pressure separation of carbon dioxide and methane. *Computers and Chemical Engineering*, 35:474–491, 2011.

L. N. Petrova. Vapor-pressures in systems acetone-water-carbazole and acetone-water-phenanthrene. *Coke and Chestry USSR*, pages 47–48, 1974.

E. N. Pistikopoulos and S. K. Stefanis. Optimal solvent design for environmental impact minimization. *Computers and Chemical Engineering*, 22:717–733, 1998.

E. N. Pistikopoulos, S. K. Stefanis, and A. G. Livingston. A methodology for minimum environmental impact analysis. In M. M. El-Halwagi and D. M. Petrides, editors, *Pollution prevention via process and product modifications*, volume 90, page 139. AIChE Symposium Series, 1994.

J. A. Platts, S. P. Oldfield, M. M. Reif, A. Palmucci, E. Gabano, and D. Osella. The RP-HPLC measurement and QSPR analysis of $\log P_{o/w}$ values of several Pt(II) complexes. *Journal of Inorganic Biochemistry*, 100:1199–1207, 2006.

B. E. Poling, J. M. Prausnitz, and J. P. O'Connell. *Properties of gases and liquids*. McGraw-Hill, 2007.

G. Power, M. Nagaraj, J. K. Vij, and G. P. Johari. Debye process and dielectric state of an alcohol in a nonpolar solvent. *Journal of Chemical Physics*, 134:044525, 2011.

W. H. Press, S. Teukolsky, W. T. Vetterling, and B. P. Flannery, editors. *Numerical recipes: the art of scientific computing*. Cambridge University Press, 1986.

E. J. Pretel, P. A. López, S. B. Bottini, and E. A. Brignole. Computer-aided molecular design of solvents for separation processes. *AIChE Journal*, 40:1349–1360, 1994.

C. Ràfols, M. Rosés, and E. Bosch. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 1. A comparison of several preferential solvation models for describing E(t)-(30) polarity of dipolar hydrogen-bond acceptor-cosolvent

mixtures. *Journal of Chemical Society Perkin Transactions 2*, pages 1607–1615, 1995.

C. Ràfols, M. Rosés, and E. Bosch. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. 5. Preferential solvation of solvatochromic indicators in mixtures of propan-2-ol with hexane, benzene, ethanol and methanol. *Journal of Chemical Society Perkin Transactions 2*, 38:243–248, 1997.

Yu. L. Rastorguev and Yu. A. Ganiev. Investigation of the heat capacity of selective solvents. *Izv. Vyssh. Uchebn. Zaved. Neft Gaz*, pages 79–82, 1967.

N. Ray and S. Bagchi. Use of a solvatochromic probe for study of solvation in ternary solvent mixture. *Journal of Physical Chemistry A*, 109:142–147, 2005.

C. Reichardt. Solvents and solvents effects: An introduction. *Organic Process Research and Development*, 11:105–113, 2007.

C. Reichardt and T. Welton. *Solvents and solvents effects in organic chemistry*. Wiley-VCH, 2011.

M. Reta, R. Cattana, and J. J. Silber. Kamlet-Taft solvatochromic parameters of nonaqueous binary mixtures between *n*-hexane and 2-propanol, tetrahydrofurane, and ethyl acetate. *Journal of Solution Chemistry*, 30:237–252, 2001.

J. A. Riddick, W. B. Bunger, and T. K. Sakano, editors. *Organic solvents*. WILEY, New York, 1986.

G. Ritzoulis, D. Missopolinou, S. Doulami, and C. Panayiotou. Relative permittivities, densities, refractive indices and ultrasound velocities of the binary systems of gamma-butyrolacetone with methanol, ethanol, 1-butanol, and 1-octanol. *Journal of Chemical Engineering Data*, 45:636–641, 2000.

M. Rosés, U. Buhvestov, C. Ráfols, F. Rived, and E. Bosch. Solute-solvent and solvent-solvent interactions in binary solvent mixtures. Part 6. A quantitative measurement of the enhancement of the water structure in 2-methylpropan-2-ol-water

and propan-2-ol-water mixtures by solvatochromic indicators. *Journal of Chemical Society Perkin Transactions 2*, pages 1341–1348, 1997.

K. Ruzicka. Thermophysical datas. *Recueil des Travaux Chimiques des Pays-Bas*, 54:574, 1935.

N. V. Sahinidis and M. Tawarmalani. Applications of global optimization to process and molecular design. *Computers and Chemical Engineering*, 24:2157–2169, 2000.

M. R. V. Sahyun and D. J. Cram. Studies in stereochemistry. XXXV. Mechanism of E_i reaction of amine oxides. *Journal of American Chemical Society*, 85:1262–1268, 1963.

M. Salomon. Electrolyte solvation in aprotic solvents. *Journal of Power Sources*, 26: 9–21, 1989.

M. Salomon. Conductance of solutions of lithium bis(trifluoromethanesulfone)imide in water, propylene carbonate, acetonitrile and methyl formate at 25 C. *Journal of Solution Chemistry*, 22:715–725, 1993.

A. P. Samudra and N. V. Sahinidis. Optimization-based framework for computer-aided molecular design. *AIChE Journal*, 59:3686–3701, 2013.

S. I. Sandler. *Chemical and engineering thermodynamics*. Wiley, 1999.

E. E. Santiso and K. E. Gubbins. Multi-scale molecular modelling of chemical reactivity. *Molecular Simulation*, 30:699–748, 2004.

N. V. Sastry and S. R. Patel. Excess volumes and dielectric properties for (methyl methacrylate + a branched alcohol) at T = 298.15 K and T = 308.15 K. *Journal of Chemical Thermodynamics*, 32:1669–1682, 2000.

S. Schiavo and B. Scrosati. Conductance and ion-pair formation of tetrabutylammonium bromide in aqueous solvent mixtures at 25 C. *Zeitschrift für Physikalische Chemie*, 102:9–24, 1976.

H. M. Senn and W. Thiel. QM/MM methods for biomolecular systems. *Angewandte Chemie-International Edition*, 48:1198–1229, 2009.

M. Shariati-Rada, M. Irandoust, T. Amini, M. Hasani, and M. Shamsipur. Chemometrics exploration of the kinetics of the reaction between amlodipine and 1,2-naphthoquinone-4-sulfonate. *Chemometrics and Inteligent Laboratory Systems*, 122:84–92, 2013.

T. J. Sheldon, C. S. Adjiman, and J. L. Cordiner. Pure component properties from group contribution: Hyrogen-bond basicity, hydrogen-bond acidity, Hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectic constant. *Fluid Phase Equilibria*, 231:27–37, 2005.

T. J. Sheldon, M. Folić, and C. S. Adjiman. Solvent design using a quantum mechanical continuum solvation model. *Industrial and Engineering Chemical Research*, 45: 1128–1140, 2006.

R. M. Shirke, A. Chaudhari, N. M. More, and P. B. Patil. Temperature dependent dielectric relaxation study of ethyl acetate - alcohol mixtures using time domain technique. *Journal of Molecular Liquids*, 94:27–36, 2001.

A Shunmugasundaram and S Balakumar. Kinetics of reaction of phenacyl bromide with 3- & 4-substituted pyridines & 4-substituted 4-styrylpyridines. *Indian Journal of Chemistry*, 24A:775–777, 1985.

K. H. Simmrock, R. Janowsky, and A. Ohnsorge. Critical data of pure substances. In *Dechema Chemistry Data Series, Frankfurt*, volume 2, 1986.

K. P. Singh, S. Gupta, P. Ojha, and P. Rai. Predicting adsorptive removal of chlorophenol from aqueous solution using artificial intelligence based modeling approaches. *Environmental Science and Pollution Research International*, 40:2271–2287, 2013.

M. Sinha, L. E. K. Achenie, and G. M. Ostrovsky. Environmentally benign solvent

design by global optimization. *Computers and Chemical Engineering*, 23:1381–1394, 1999.

M. Sinha, L. E. K. Achenie, and R. Gani. Blanket wash solvent blend design using interval analysis. *Industrial and Engineering Chemistry Research*, 42:516–527, 2003a.

M. Sinha, L. E. K. Achenie, and G. M. Ostrovsky. Case study in optimal solvent design. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 247–259. Elsevier, Amsterdam, The Netherlands, 2003b.

R. D. Skwierczynski and K. A. Connors. Solvent effects on chemical processes.7.Quantitative description of the composition dependence of the solvent polarity measure E-T(30) in binary aqueous-organic solvent mixtures. *Journal of Chemical Society Perkin Transactions 2*, pages 467–472, 1994.

R. A. Stairs and E. Buncel. Principal component analysis of solvent effects on equilibria and kinetics - A hemisphere model. *Canadian Journal of Chemistry*, 84: 1580–1591, 2006.

I. Stanescu and L. E. K. Achenie. A theoritical study of solvent effects on Kolbe-Schmitt reaction kinetics. *Chemical Engineering Science*, 61:6199–6212, 2006.

H. Struebing, editor. *Identifying optimal solvents for reactions using quantum mechanics and computer-aided molecular design.* PhD Thesis, Imperial College London, 2011.

H. Struebing, P. G. Karamertzanis, E. N. Pistikopoulos, A. Galindo, and C. S. Adjiman. Solvent design for a Menschutkin reaction by using CAMD and DFT calculations. In S. Pierucci and B. G. Ferraris, editors, *20th European Symposium on Computer Aided Process Engineering*, volume 28, pages 1291–1296, 2010.

H. Struebing, Z. Ganase, P. G. Karamertzanis, E. Siougkrou, P. Haycock, P. M. Piccione, A. Armstrong, A. Galindo, and C. S. Adjiman. Computer-aided molecular

design of solvents for accelerated reaction kinetics. *Nature Chemistry*, 2013. doi: 10.1038/nchem.1755.

B. Subramaniam. Gas-expanded liquids for sustainable catalysis and novel materials: Recent advances. *Journal of Supercritical Fluids*, 254:1843–1853, 2010.

R. W. Taft and M. J. Kamlet. Solvatochromic comparison method.2.Alpha-scale of solvent hydrogen-bond donor(HBD) acidities. *Journal of American Chemical Society*, 98:2886–2894, 1976.

M. Tawarmalani and N. V. Sahinidis. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*, 103:225–249, 2005.

D. G. Truhlar. Inverse solvent design. *Nature Chemistry*, 5:902–903, 2013.

C. H. Twu, J. E. Coon, and J. R. Cunningham. A new generalized alpha function for a cubic equation of state. Part 1. Peng-Robinson equation. *Fluid Phase Equilibria*, 105:49–59, 1995.

R. Vaidyanathan and M. El-Halwagi. Global optimization of nonconvex nonlinear programs via interval analysis. *Computers and Chemical Engineering*, 18:889–897, 1994.

J. M. Vinson. CAMD for solvent selection in industry - I. In L. E. K. Achenie, R. Gani, and V. Venkatasubramanian, editors, *Computer Aided Molecular Design: Theory and Practice*, volume 12, pages 213–228. Elsevier, Amsterdam, The Netherlands, 2003.

D. P. Visco, R. S. Pophale, M. D. Rintoul, and J. L. Faulon. Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *Journal of Molecular Graphics and Modelling*, 20:429–438, 2002.

Y. Wang and L. E. K. Achenie. Computer aided solvent design for extractive fermentation. *Fluid Phase Equilibria*, 201:1–18, 2002.

M. Wei, G. T. Musie, D. H. Busch, and B. Subramaniam. CO_2-expanded solvents: Unique and versatile media for performing homogeneous catalytic oxidations. *Journal of American Chemical Society*, 124:2513–2517, 2002.

D. C. Weis and D. P. Visco. Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. *Computers and Chemical Engineering*, 34:1018–1029, 2010.

D. Wen and S. V. Olesik. Chromatography of substituted benzoic acids with methanol-water carbon dioxide mixtures. *Journal of Chromatography A*, 931:41–52, 2001.

K. N. West, J. P. Hallett, R. S. Jones, D. Bush, C. L. Liotta, and C. A. Eckert. CO_2-induced miscibility of fluorous and organic solvents for recycling homogeneous catalysts. *Industrial and Engineering Chemistry Research*, 43:4827–4832, 2004.

S J Winston, P J Rao, B Sethuram, and T N Rao. The LFER correlations for S_N2 reactions of phenacyl bromide. *Indian Journal of Chemistry*, 35A:979–982, 1996.

V. T. Wyatt, D. Bush, J. Lu, J. P. Hallet, C. L. Liotta, and C. A. Eckert. Determination of solvatochromic solvent parameters for the characterization of gas-expanded liquids. *Journal of Supercritical Fluids*, 36:16–22, 2005.

W. F. K. Wynne-Jones and H. Eyring. The absolute rate of reactions in condensed phases. *Journal of Chemical Physics*, 3:492–502, 1935.

K. Ye, H. Freund, Z. Xie, B. Subramaniam, and K. Sundmacher. Prediction of multicomponent phase behavior of CO_2-expanded liquids using CEoS/GE models and comparison with experimental data. *Journal of Supercritical Fluids*, 67:41–52, 2012.

S D Yoh, K T Shim, and K A Lee. Studies on the quaternization of tertiary amines (II). Kinetics and mechanism for the reaction of substituted phenacyl bromides with substituted pyridines. *Journal of the Korean Chemical Society*, 25:110–118, 1981.

L. E. Zerpa, N. V. Queipo, S. Pintos, and J.-L. Salager. An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. *Journal of Petroleum Science and Engineering*, 47:197–208, 2005.

Y. Zhao, N. González-García, and D. G. Truhlar. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *Journal of Physical Chemistry A*, 109:2012–2018, 2005.

T. Zhu, J. Li, G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. Density functional solvation model based on CM2 atomic charges. *Journal of Chemical Physics*, 109: 9117–9133, 1998.

A. M. Zissimos, M. H. Abraham, C. M. Du, K. Valko, C. Bevan, D. Reynolds, J. Wood, and K. Y. Tam. Calculation of Abraham descriptors from experimental data from seven HPLC systems; Evaluation of five different methods of calculation. *Journal of Chemical Society - Perkin Transactions II*, pages 2001–2010, 2002.

Appendix A

Data for MINLP optimisation problem

Information used in the CAMD formulation is given here. The set of groups that are included in the design and subsets used in the formulation are presented, as well as the group contributions to the solvent properties.

A.1 Groups of set G

Groups i	**Upper limit** n_i^U
CH_3	$n_{G,max}y_5 + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}}$
CH_2	$3y_5$
CH	$3y_5$
C	y_5
CH_2=CH*	$y_5+y_M+y_{\mathrm{aCCH}}+y_{\mathrm{aCCH_2}}$
CH=CH*	$y_5+y_M+y_{\mathrm{aCCH_2}}$
CH_2=C*	y_5
CH=C*	y_5
C=C	y_5

Continued on next page

Continued from previous page

Groups i	**Upper limit** n_i^U
aCH	$6y_7+8y_6$
aC	$y_M + 2y_6$
$aCCH_3$	$6y_7+8y_6$
$aCCH_2$	y_{aCCH_2}
aCCH	y_{aCCH}
OH	y_5
aCOH	$6y_7+8y_6$
CH_3CO	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_2CO	$y_5 + y_M + y_{aCCH_2}$
CHO	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_3COO	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_2COO	$y_5 + y_M + y_{aCCH_2}$
CH_3O	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_2O	$y_5 + y_M + y_{aCCH_2}$
CH-O	y_5
$CH_2NH_2{}^*$	$2y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_3NH^*	y_5
CH_2NH^*	$y_5 + y_M + y_{aCCH_2}$
CH_3N^*	$y_5 + y_M + y_{aCCH_2}$
CH_2N^*	y_5
$aCNH_2{}^*$	$6y_7+8y_6$
CH_2CN	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
COOH	$y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CH_2Cl	$2y_5 + y_M + y_{aCCH} + y_{aCCH_2}$
CHCl	y_5
$CHCl_2$	y_5
aCCl	$6y_7 + 8y_6$

Continued on next page

Continued from previous page	
Groups	**Upper limit**
i	n_i^U
CH_2NO_2	$y_5 + y_M + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}}$
$CHNO_2$	$y_5 + y_M + y_{\mathrm{aCCH_2}}$
CH_2SH^*	$y_5 + y_M + y_{\mathrm{aCCH_2}}$
I	$2y_5 + y_M + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}}$
Br	$2y_5 + y_M + y_{\mathrm{aCCH}} + y_{\mathrm{aCCH_2}}$
aCF	$6y_7 + 8y_6$
CH_2S^*	y_5

Table A.1: Groups included in set G and the upper limit on the number of groups allowed to appear at solvent candidates. Groups with an * have been deactivated due to possible reactivity and/or lack of sufficient data for optimal solvent prediction.

A.2 Subsets of set G

G_F	G_{NCE}	G_{CE}	G_A	G_H	G_M
CH_3CO	$CHNO_2$	CH_2NO_2	aCH	I	CH_3
CH_2CO	CH_2CO	CH_3CO	aC	Br	CH_2
CHO	CH_2COO	CH_3COO	$aCCH_3$	CH_2Cl	CH
CH_3COO	CH_2O	CH_3O	$aCCH_2$	$CHCl_2$	C
CH_2COO	$CH{=}CH^*$	CHO	aCCH	CHCl	C=C
CH_3O	CH_2NH^*	COOH	aCOH	$CHCl_3$	$CH_2{=}CH^*$
CH_2O	CH_3N^*	CH_2CN	aCCl		$CH{=}CH^*$
CH-O	CH_2SH^*	CH_2Cl	aCF		$CH_2{=}C^*$
CH_2CN		I	$aCNH_2^*$		$CH{=}C^*$
COOH		Br			
CH_2Cl		$CH_2{=}CH^*$			
CHCl		$CH_2NH_2^*$			
$CHCl_2$					
CH_2NO_2					
$CHNO_2$					
OH					
I					
Br					
CH_3NH^*					
CH_2S^*					
$CH_2NH_2{}^*$					
CH_2NH^*					
CH_3N^*					
CH_2N^*					
CH_2SH^*					

Table A.2: Subsets of set G. G_F subset of functional groups, G_{NCE} subset of non-chain-ending groups, G_{CE} subset of chain-ending groups, G_A subset of aromatic groups, G_H subset of halogenated groups and G_M subset of main groups. Groups with an * have been deactivated.

A.3 Sets used for the calculation of the dielectric constant, refractive index and dipole moment

G_{1-9}	G_1	G_3	G_{n_D}	G_{HC}
CH_3	OH	CH_2Cl	CH_3O	CH_3
CH_2	CH_3CO	CHCl	CH_2O	CH_2
CH	CH_2CO	$CHCl_2$	CH-O	CH
C	CHO		CH_2Cl	C
CH_2=CH*	CH_2CN		CHCl	CH_2=CH*
CH=CH*	CH_2NO_2		$CHCl_2$	CH=CH*
CH_2=C*	$CHNO_2$		I	CH_2=C*
CH=C*			Br	CH=C*
C=C			aCCl	C=C
			aCF	aCH
				aC
				$aCCH_3$
				$aCCH_2$
				aCCH

Table A.3: Sets G_{1-9}, G_1 and G_3, are used to determine the dielectric constant, sets G_{n_D} and G_{HC} are used to define contributions for the refractive index and dipole moment, respectively [Sheldon et al., 2005]. Groups with an * have been deactivated.

A.4 Group contributions for valency, acidity, basicity, dipolarity/polarisability, heat of vaporisation and liquid molar volume.

n_i	v_i	A_i	B_i	S_i	$H_{V,i}$	$V_{m,i}$
CH_3	1	0.000000	0.000000	−0.121946	0.217	0.02614
CH_2	2	−0.001215	0.000000	−0.004475	4.910	0.01641
CH	3	0.000000	0.000000	0.060456	7.962	0.00711
C	4	−0.029098	0.061588	0.220533	10.730	−0.00380
CH_2=CH*	1	0.000000	0.000000	−0.068898	4.031	0.03727
CH=CH*	2	0.000000	0.000000	0.000000	9.456	0.02692
CH_2=C*	2	0.000000	0.027210	0.069770	8.602	0.02697
CH=C*	3	0.000000	0.020261	0.115250	14.095	0.01610
C=C	4	0.000000	0.000000	0.250224	19.910	0.00296
aCH	2	0.000000	0.005373	0.037395	3.683	0.01317
aC	3	0.000000	0.000000	0.125015	6.631	0.00440
$aCCH_3$	2	−0.005007	0.009939	0.048763	8.279	0.02888
$aCCH_2$	3	0.000000	0.052453	0.128524	11.981	0.01916
aCCH	4	0.000000	0.059835	0.220456	13.519	0.00993
OH	1	0.317097	0.362081	0.200678	24.214	0.00551
aCOH	2	0.565501	0.120599	0.353406	34.099	0.01133
CH_3CO	1	0.000000	0.367063	0.485266	15.195	0.03655
CH_2CO	2	0.000000	0.387556	0.567022	19.392	0.02816
CHO	1	0.000000	0.304903	0.489932	12.370	0.02002
CH_3COO	1	−0.01681	0.351460	0.373931	19.342	0.04500
CH_2COO	2	0.000000	0.340981	0.562160	21.100	0.03567
CH_3O	1	−0.059281	0.200768	0.084134	5.783	0.03274
CH_2O	2	0.000000	0.292346	0.193892	9.997	0.02311
CH-O	3	0.000000	0.195516	0.245215	14.620	0.01799
CH_2NH_2*	1	0.143565	0.498168	0.171166	15.432	0.02646

Continued on next page

Continued from previous page						
n_i	v_i	A_i	B_i	S_i	$H_{V,i}$	$V_{m,i}$
CH_3NH^*	1	0.279359	0.273015	0.372262	11.831	0.02674
CH_2NH^*	2	0.141189	0.566291	0.245236	13.067	0.02318
CH_3N^*	2	0.000000	0.396525	0.357309	9.493	0.01913
CH_2N^*	3	−0.039727	0.468704	0.301618	12.636	0.01683
$aCNH_2^*$	2	0.255983	0.248167	0.437387	23.335	0.01365
CH_2CN	1	0.000000	0.234188	0.758708	21.923	0.03313
COOH	1	0.599263	0.309620	0.370092	17.002	0.02232
CH_2Cl	1	0.000000	0.012326	0.182209	11.754	0.03371
CHCl	2	0.000000	0.032895	0.175840	12.048	0.02663
$CHCl_2$	1	0.166271	0.000000	0.261610	17.251	0.04682
aCCl	2	0.000000	−0.063720	0.118214	11.224	0.02414
CH_2NO_2	1	0.000000	0.178291	0.755221	29.640	0.03375
$CHNO_2$	2	0.000000	0.196291	0.666699	29.173	0.02620
CH_2SH^*	1	0.000000	0.112291	0.163076	16.815	0.03446
I	1	0.000000	0.018577	0.196690	14.171	0.02791
Br	1	0.017962	0.004644	0.176422	9.888	0.02143
aCF	2	0.000000	0.000000	0.040433	3.965	0.01727
CH_2S^*	2	0.000000	0.196291	0.311642	14.296	0.02732

Table A.4: Group contributions for valency v_i, acidity A_i, basicity B_i, dipolarity/polarisability S_i, heat of vaporisation $H_{V,i}$ and liquid molar volume $V_{m,i}$. Groups with an * have been deactivated.

A.5 Group contributions for Melting and boiling temperature.

n_i	$T_{m,i}$	$T_{b,i}$
CH_3	0.6953	0.8491
CH_2	0.2515	0.7141
CH	−0.3730	0.2925
C	0.0256	−0.0671
CH_2=CH*	1.1728	1.5596
CH=CH*	0.9460	1.5597
CH_2=C*	0.7662	1.3621
CH=C*	0.1732	1.2971
C=C	0.3928	1.2739
aCH	0.5860	0.8365
aC	1.8955	1.7324
$aCCH_3$	1.0068	1.5653
$aCCH_2$	0.1065	1.4925
aCCH	−0.5197	0.8665
OH	2.7888	2.5670
aCOH	5.1473	3.3205
CH_3CO	2.9588	3.1178
CH_2CO	2.5232	2.6761
CHO	3.0186	2.5388
CH_3COO	2.1657	3.1228
CH_2COO	1.6329	2.9850
CH_3O	1.3643	1.7703
CH_2O	0.8733	1.3368
CH-O	0.2461	0.8924
CH_2NH_2*	3.2742	2.7987
CH_3NH*	2.4034	2.2514

Continued on next page

Continued from previous page		
n_i	$T_{m,i}$	$T_{b,i}$
CH_2NH^*	1.7746	1.8750
CH_3N^*	0.9607	1.3841
CH_2N^*	0.0442	1.1222
$aCNH_2{}^*$	3.9889	3.8298
CH_2CN	2.5760	4.5871
COOH	7.4042	5.1108
CH_2Cl	1.9253	2.6364
CHCl	1.0224	2.0246
$CHCl_2$	2.5196	3.3420
aCCl	1.7134	2.0669
CH_2NO_2	3.2131	4.5311
$CHNO_2$	0.7812	3.8069
CH_2SH^*	2.2992	3.1974
I	1.9444	3.1778
Br	1.7641	2.4231
aCF	0.9782	0.7945
CH_2S^*	1.0063	2.6524

Table A.5: Group contributions for melting $T_{m,i}$ and boiling $T_{b,i}$ temperature. Groups with an * have been deactivated.

A.6 Group contributions for the dipole moment d_i, the aromaticity φ_i, the electronegative halogenicity ψ_i and the number of non-hydrogen atoms ζ_i.

n_i	d_i	φ_i	ψ_i	ζ_i
CH_3	0.0	0	0	1
CH_2	11255.6	0	0	1
CH	29480.8	0	0	1
C	78086.8	0	0	1
CH_2=CH*	3841.2	0	0	2
CH=CH*	48323.4	0	0	2
CH_2=C*	73577.7	0	0	2
CH=C*	24138.9	0	0	2
C=C	0.0	0	0	2
aCH	28212.2	1	0	1
aC	−50745.9	1	0	1
$aCCH_3$	10272.1	1	0	2
$aCCH_2$	−108358.3	1	0	2
aCCH	−44220.2	1	0	2
OH	80002.0	0	0	1
aCOH	52958.9	1	0	2
CH_3CO	400172.6	0	0	3
CH_2CO	353695.2	0	0	3
CHO	412628.5	0	0	2
CH_3COO	222410.7	0	0	4
CH_2COO	256788.1	0	0	4
CH_3O	44572.9	0	0	2
CH_2O	61084.6	0	0	2

Continued on next page

Continued from previous page				
n_i	d_i	φ_i	ψ_i	ζ_i
CH-O	37588.6	0	0	2
CH_2NH_2*	40231.2	0	0	2
CH_3NH*	56361.3	0	0	2
CH_2NH*	13285.8	0	0	2
CH_3N*	66257.9	0	0	2
CH_2N*	−19660.2	0	0	2
$aCNH_2$*	17076.4	1	0	2
CH_2CN	217391.6	0	0	3
COOH	110267.7	0	0	3
CH_2Cl	163844.0	0	1	2
CHCl	145381.6	0	1	2
$CHCl_2$	110685.2	0	2	3
aCCl	778.7	1	1	2
CH_2NO_2	239019.2	0	0	4
$CHNO_2$	177360.5	0	0	4
CH_2SH*	N/A	0	0	2
I	110776.4	0	0	1
Br	66504.4	0	1	1
aCF	−11053.3	1	1	2
CH_2S*	N/A	0	0	2

Table A.6: GCs for dipole moment d_i [Sheldon et al., 2005], aromaticity φ_i, electronegative halogenicity ψ_i and number of non-hydrogen atoms ζ_i. Groups with an * have been deactivated.