

CONFIDENTIAL UP TO AND INCLUDING December 31, 9999
IMPORTANT

This master dissertation contains confidential information and/or confidential research results proprietary to Ghent University or third parties. It is strictly forbidden to publish, cite or make public in any way this master dissertation or any part thereof without the express written permission of Ghent University. Under no circumstance this master dissertation may be communicated to or put at a disposal of third parties. Photocopying or duplicating it in any other way is strictly prohibited. Disregarding the confidential nature of this master dissertation may cause irremediable damage to Ghent University.

The stipulations above are in force until the embargo date.

Oxidation of cyclic compounds: a combined modelling and experimental study

Kevin De Ras

Student number: 01408695

Supervisors: Prof. dr. ir. Kevin Van Geem, Prof. dr. ir. Joris Thybaut
Counsellor: Florence Vermeire

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Chemical Engineering

Academic year 2018-2019

Luck is great,
but most of life is hard work.

— Iain Duncan Smith

Acknowledgements

Nu het einde van dit thesisjaar nadert, neem ik de tijd om een welgemeend woord van dank te uiten aan iedereen die mij het afgelopen jaar geholpen en gesteund heeft.

Allereerst wil ik starten met prof. dr. ir. Kevin Van Geem te bedanken voor het vertrouwen en de mogelijkheid om op dit uitdagende onderwerp onderzoek te mogen verrichten. Ik kijk er alvast naar uit om op hetzelfde elan verder te werken de komende jaren. Daarnaast wil ik mijn beide promotoren bedanken, prof. dr. ir. Kevin Van Geem en prof. dr. ir. Joris Thybaut, voor de feedback en opvolging van deze thesis.

Mijn allergrootste dank gaat ongetwijfeld uit naar Florence Vermeire die mij begeleidde op een manier waar anderen enkel van konden dromen. Hulp was steeds slechts enkele meters verwijderd en met jouw ervaring waren problemen vaak zeer snel opgelost. Je hebt mij in de wondere wereld van automatic kinetic model generation geïntroduceerd en zonder jouw tips & tricks, feedback op mijn teksten, *etc.* was deze thesis nooit tot het voorliggende resultaat gekomen.

Ik wil eveneens de andere doctoraatsstudenten bedanken, Cato Pappijn, Pieter Plehiers en Ruben Van de Vijver, jullie hebben me meermaals uit de nood geholpen met Genesys en andere computerproblemen.

Moreover, I would like to thank Marko Djokic for taking the time to teach me the basics of experimental research and always staying ready to help me.

Een speciaal woord van dank gaat uit naar mijn mede “Team Genesys”-leden, Morgane Janssens, David De Pauw en Pieter-Jan Verberckmoes. De groepssfeer die er was op ons verdiep is niet te overtreffen en ik mis nu al onze (sportieve) pauzes en de toffe babbels die we hadden over de meest uiteenlopende onderwerpen. Ook Florian Wéry en Axel Vernimmen wil ik in het bijzonder bedanken voor de aangename breaks tussen het werken door.

Uiteraard kan ik het MaChT board hier niet vergeten vermelden voor alle fijne activiteiten die we samen georganiseerd en beleefd hebben het afgelopen jaar.

Tenslotte wil ik ook nog Nele, mama en papa bedanken die er altijd voor mij waren en mij steeds gesteund hebben in mijn doen en laten de afgelopen jaren.

Kevin De Ras
Gent, Juni 2019

Declaration concerning the accessibility of the master thesis

Undersigned,

Kevin De Ras

Graduated from Ghent University, academic year 2018-2019 and is author of the master thesis with title:

Oxidation of cyclic compounds: a combined modelling and experimental study

The author(s) gives (give) permission to make this master dissertation available for consultation and to copy parts of this master dissertation for personal use. In the case of any other use, the copyright terms have to be respected, in particular with regard to the obligation to state expressly the source when quoting results from this master dissertation.

June 2019

Kevin De Ras

Oxidation of cyclic compounds: a combined modelling and experimental study

Kevin De Ras

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Chemical Engineering

Academic year 2018 – 2019

Promotors: Prof. dr. ir. Kevin M. Van Geem, Prof. dr. ir. Joris W. Thybaut

Counsellor: dr. ir. Florence Vermeire

GHENT UNIVERSITY

Faculty of Engineering and Architecture

Department of Materials, Textiles and Chemical Engineering

Chairman: Prof. dr. Paul Kiekens

Abstract

Nowadays, there is a strong drive to design new and optimize existing large-scale chemical processes in the pursuit of sustainability and increased profitability. Theoretical and computational advances have made predictive first principles based modelling of gas-phase processes feasible. This is in sharp contrast with the limited fundamental understanding of reactions important in many liquid-phase processes. The liquid-phase oxidation of cyclohexane to cyclohexanone and cyclohexanol is investigated, since it is an important industrial process for the synthesis of nylon-6(6) which is only partially understood. Genesys, an in-house automatic kinetic model generation code, is used to construct a kinetic model for the gas-phase oxidation of cyclohexane to confirm well-functioning of the framework for the gas phase. Potential energy surfaces are constructed to uncover the important reaction paths. Experiments from literature are used for validation. Experimental trends are predicted well with one adapted parameter. Subsequently, an algorithm to calculate the Gibbs free energy of solvation is implemented in Genesys as a first extension to introduce liquid-phase effects. The latter is validated with available databases and allows calculation of the Gibbs free energy of solvation with a mean absolute error of 2.5 kJ mol⁻¹.

Keywords: Automatic kinetic model generation, computational chemistry, cyclohexane oxidation, liquid-phase modelling

Oxidation of cyclic compounds: a combined modelling and experimental study

Kevin De Ras

Promotors: prof. dr. ir. Kevin M. Van Geem*, prof. dr. ir. Joris W. Thybaut
Counsellor: dr. ir. Florence H. Vermeire

Abstract

Nowadays, there is a strong drive to design new and optimize existing large-scale chemical processes in the pursuit of sustainability and increased profitability. Theoretical and computational advances in the last decade have made predictive first principles based modelling of gas-phase processes feasible. This is in contrast with the limited fundamental understanding of reactions important in liquid-phase processes. The liquid-phase oxidation of cyclohexane to cyclohexanone and cyclohexanol is investigated, since it is an important industrial process for the synthesis of nylon-6 and nylon-6,6 which is only partially understood. Genesys, an in-house developed automatic kinetic model generation code, is used to construct a kinetic model for the gas-phase oxidation of cyclohexane to confirm well-functioning of the framework for the gas phase. Potential energy surfaces are constructed to uncover the important reaction pathways. To include accurate reactivity of smaller (oxygenated) hydrocarbons, the generated model from Genesys is merged with AramcoMech 2.0. Experiments from literature are used to assess the final model for its performance. The experimental trends of the conversion and major product formation are predicted well, especially in the low-temperature oxidation zone after fitting of one parameter. For this work, this low-temperature oxidation regime of the gas phase is of interest since the same type of reactions take place during liquid-phase oxidation. Subsequently, an algorithm to calculate the Gibbs free energy of solvation at 298 K and standard conditions is implemented in Genesys as a first extension to introduce liquid-phase effects. The latter is validated with a data set containing both *ab initio* and experimental values. The implemented scheme allows calculation of the Gibbs free energy of solvation with a mean absolute error of 2.5 kJ mol⁻¹ for the complete validation set.

Keywords

Automatic kinetic model generation, computational chemistry, cyclohexane oxidation, liquid-phase modelling

I. INTRODUCTION

An accurate chemical microkinetic model is an extremely powerful and valuable tool. Integrated within an overall process model, design and optimization of the chemical process becomes possible. Currently these topics are on the mindset of many people, considering that process optimization can result

in significant energy savings and associated reduction of greenhouse gas emissions, more specifically CO₂ emissions, as well as reduction of waste streams.

The last decade, great progress has been made in the construction of reliable microkinetic models for a variety of technologically important gas-phase processes including pyrolysis and oxidation. The corresponding models typically contain thousands of species and accompanied several tens of thousands of reactions. Manual construction of a microkinetic model by hand becomes thereby tedious, error prone and often incomplete. To prevent this, automatic microkinetic model generation codes have been developed, *e.g.* the in-house developed code Genesys [1].

First principles based modelling of gas-phase processes is currently feasible due to the increased computational and theoretical advancements the last decade. This is in sharp contrast with the limited fundamental knowledge concerning modelling of important liquid-phase processes. In this work the liquid-phase oxidation of cyclohexane to cyclohexanone and cyclohexanol (*cf.* Figure 1) is investigated. This is an important process in the production chain of nylon-6 and nylon-6,6 for which the chemistry is only partially understood on the molecular level. Other industrial relevant liquid-phase processes include the production of phenol and terephthalic acid by oxidation of respectively cumene and *p*-xylene. The long-term objective is therefore to extend the framework of Genesys, which is currently tailored to the gas-phase, to also allow automatic microkinetic model generation for liquid-phase processes.

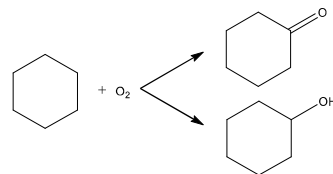


Figure 1: Reaction scheme for oxidation of cyclohexane to cyclohexanone and cyclohexanol.

First, a gas-phase model is constructed with Genesys for the oxidation of cyclohexane to gain insight in the elementary chemistry between cyclohexane and molecular oxygen. From the developed microkinetic gas-phase model, several reaction families can be reused for the development of a liquid-phase model. Especially since literature indicates that in the low temperature gas-phase oxidation zone similar reactions occur as during the liquid-phase oxidation. The microkinetic model is assessed for its performance by experimental data from literature.

* E-mail: Kevin.VanGeem@UGent.be .

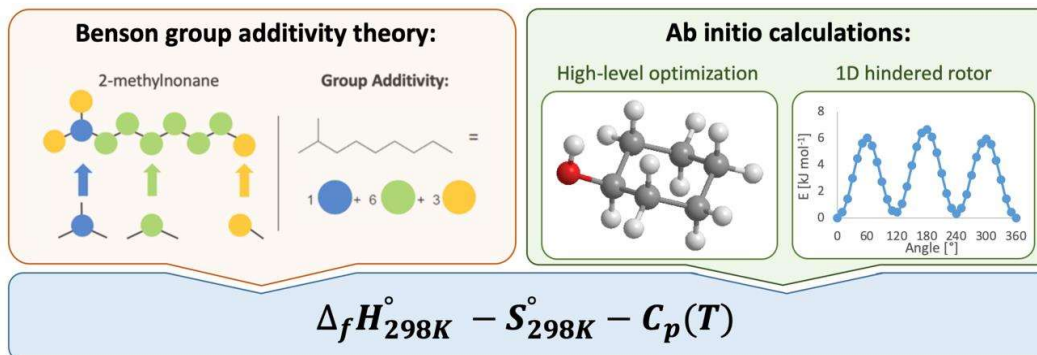


Figure 2: Benson group additivity theory and ab initio calculations for thermodynamic parameter assignment.

A second part concerns implementation of a calculation algorithm for the Gibbs free energy of solvation at 298 K and standard conditions in Genesys via the empirical Abraham equation, as a first inclusion of liquid-phase effects. Assessment is made of the accuracy by comparison of calculated and tabulated solvation energies which are obtained from an experimental database and from ab initio calculations.

II. METHODOLOGY

Automatic microkinetic model generation with Genesys requires specification of the reaction families which are used for reaction network generation. Reaction termination criteria have to be defined to limit the reaction network to the essential chemistry. In addition, accurate thermodynamic parameters for all species and kinetic parameters for all reactions in the network are crucial for model development.

A. Kinetic model construction

Kinetic model generation starts with user-defined reactants. These are subjected to the defined reaction families giving birth to new products, which can subsequently be subjected to the reaction families themselves giving birth to second-generation, third-generation, and so on products. The kinetic model generation iteratively continues until the reaction network is considered complete by user-defined criteria.

Typical acyclic hydrocarbon oxidation reaction families are defined in the model, e.g. hydrogen abstraction by molecular oxygen, molecular oxygen addition, chain-branching reactions, hydrogen shifts, cyclic ether formation, β -scission of carbon-carbon and carbon-hydrogen bonds and α -scission of carbonyl.

To avoid models with tens of thousands of reactions where most of them do not significantly affect the kinetic model, several reaction rules are introduced. These criteria or rules determine when to stop the generation of new reactants which allows one to attain a manageable size of the microkinetic model for reactor simulations and optimization, that still contains the essential chemistry. In the generated model for example, reactions can only generate products with maximal two oxygen atoms. Due to the latter (severe) constraint chain-branching reactions do not take place, which are therefore separately included.

When the model is generated both accurate thermodynamic and kinetic parameters are supplied for oxidation reactions starting from cyclohexane. The often pressure dependent kinetics of small molecules are not yet available in Genesys, in addition the used thermodynamic calculation scheme fails for small compounds. Therefore, the Genesys model is merged with a base mechanism before reactor simulations are

performed in ANSYS Chemkin. AramcoMech 2.0 [2] is used which includes thermodynamic parameters for C_1 to C_4 (oxygenated) compounds and associated reactions.

B. Thermodynamic and kinetic parameter assignment

In order to generate a reliable microkinetic model, it is necessary to assign both accurate thermodynamic parameters to all species, i.e. $\Delta_f H^\circ_{298K}$, S°_{298K} , $C_p(T)$, and (modified) Arrhenius parameters to all reactions present in the kinetic model. The latter are necessary to calculate the reaction rates and equilibrium compositions.

When thermodynamic parameters, are not available in the ab initio database of Genesys, these can be calculated via Benson group additivity theory [3] or ab initio calculations, as indicated in Figure 2. Benson group additivity is a powerful scheme which allows calculation of the thermodynamic parameters in which every group, which a molecule is built from, contributes and includes several corrections for special interactions within the molecule. A group is defined as a polyvalent atom in the molecule together with its immediate neighbors. The final outcome only deviates 4 kJ mol⁻¹ on average compared with ab initio calculations and experimental measurements. However, for cyclic compounds the GAVs are less accurate due to ring strain in six-membered ring structures.

Quantum mechanical calculations have been carried out in Gaussian V16 to obtain thermodynamic parameters for important molecules where GAVs are missing or unreliable. High-level geometry optimization is performed at the CBS-QB3 level of theory, see Figure 2. To take into account the different rotational conformers of a molecule, one-dimensional hindered rotor potentials have been calculated at the B3LYP/6-31G(D) level of theory which are included as an extra contribution to the partition function. The outcome is manually corrected with bond additive corrections and spin-orbit corrections.

For assignment of kinetic parameters, a similar approach is applied. Rate rules and kinetic parameters from literature are used for several reactions. A group additive scheme is used for the Arrhenius and pre-exponential parameters of β -scission and hydrogen abstraction reactions as proposed by Saeys et al. [4]. In addition, ab initio calculations are carried out for the most important reaction paths which allows calculation of the modified Arrhenius parameters via transition state theory.

III. GAS-PHASE MODEL FOR CYCLOHEXANE OXIDATION

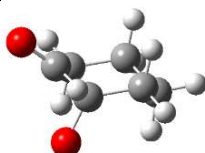
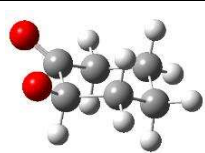
A microkinetic model for the gas-phase oxidation of cyclohexane is developed focused on the low-temperature

oxidation regime since the reactions proceeding are similar as the reactions proceeding during liquid-phase oxidation. Ab initio calculations are carried out to obtain accurate thermodynamic parameters and to uncover the important reaction paths by construction of a potential energy surface. The final model is assessed with experiments for its performance.

A. Ab initio calculations for thermodynamic parameter calculation and potential energy surface construction

To start, accurate thermodynamic parameters are calculated for the important species. For accuracy, it is observed that it is important to treat both axial and equatorial positioned substituents on the six-membered ring since the energy difference between the geometries can reach beyond 10 kJ mol⁻¹. For example consider 2-oxyocyclohexan-1-olate in Table 1, the equatorial positioned geometry is unfavorable due to interaction between the oxy group and the carbonyl group. This is less for the axial positioned geometry. The difference in enthalpy of formation at 298 K amounts to 18.9 kJ mol⁻¹. In case of a single substituted species without radicals and double bonds in the six ring, the equatorial position is always favored both for the enthalpy of formation and the entropy. In addition to repulsive interactions, also attractive interactions can be formed such as a hydrogen bond which do decrease both the enthalpy of formation and the entropy.

Table 1: Spatial arrangement and enthalpy of formation for 2-oxyocyclohexan-1-olate of the axial and equatorial positioned oxy group.

	Axial	Equatorial
		
$\Delta_f H_{298K}$ [kJ mol ⁻¹]	-164.7	-145.9

The potential energy surface (PES) for the addition of molecular oxygen to the cyclohexyl radical has been constructed by high-level ab initio calculations at the CBS-QB3 level of theory. Ab initio calculated (modified) Arrhenius parameters are used to replace rate rule parameters which are developed for acyclic alkanes. The most important reaction pathways are depicted in Figure 4. From the PES, it is clear that the peroxide radical can be converted to a hydroperoxide by a hydrogen shift reaction with a barrier for the 1,5-shift and 1-6 shift of 108 kJ mol⁻¹ and 129 kJ mol⁻¹ respectively at 0 K. This barrier is below the dissociation reaction enthalpy to molecular oxygen and the cyclohexyl radical for which an enthalpy difference between reactants and products of 156 kJ mol⁻¹ is calculated. The barrier for the 1,4-shift amounts to 164 kJ mol⁻¹ which is slightly higher than the enthalpy of reaction of the reverse dissociation reaction. It can be expected that the 1,5-shift will also occur because it is entropically favored compared to the other shifts. These formed hydroperoxide radicals can react further mainly through cyclic ether formation and β -scission reactions, as depicted in Figure 4. Alternatively, oxygen addition on the hydroperoxide radicals leads to chain branching, which is not shown on the PES.

Besides the hydrogen shifts, also hydroperoxide elimination can take place with formation of cyclohexene for which the activation barrier amounts to 134 kJ mol⁻¹, which is below the reverse dissociation to molecular oxygen and the cyclohexyl radical. The unimolecular decomposition reaction considered with the highest enthalpy of formation is a 1,3- hydrogen shift for which 113 kJ mol⁻¹ is released due to the formation of the energetically favorable cyclohexanone. However, an activation energy of 165 kJ mol⁻¹ makes proceeding of this reaction unlikely. Other hydrogen shifts, carbon-carbon and carbon-hydrogen β -scissions are calculated and do not play an important role except for one β -scission possible after the 1,5-shift which forms the hex-5-enal and hydroxyl radical. This reaction has an energy barrier equal to 106 kJ mol⁻¹ (but still an energy barrier of 162 kJ mol⁻¹ compared to the starting products) with an enthalpy change of -70 kJ mol⁻¹ which is entropically favorable due to ring-opening.

B. Model assessment

The developed microkinetic model of Genesys is validated with experiments published in literature, performed in a jet-stirred reactor at l'Université de Lorraine by the group of Prof. Battin-Leclerc [5]. The experimental conditions are a pressure of 1.07 atm, a reactor volume of 85 cm³, an equivalence ratio of 1.0 and a molar inlet flow of cyclohexane of 0.667 dilution with helium. The model predicts well the experimental conversion in the low-temperature oxidation region, as depicted in Figure 3, when the alternative activation energy of the modified Arrhenius equation for the mentioned β -scission, a sensitive reaction, is lowered from 24.2 kJ mol⁻¹ to 12.2 kJ mol⁻¹. The on-set temperature is a bit early, but both the experimental and modelled low-temperature oxidation peak reach a maximum around 610 K which do also qualitatively agree (50% versus 40% conversion). For the high-temperature oxidation zone, the on-set temperature is correct, but the increase in conversion is predicted too slow. Extra ab initio calculations for the important reactions in this region are suggested. Major model formation is predicted rather well and follows the experimental trend in the temperature range between 500 and 750 K, formation of carbon monoxide, as depicted in Figure 5, is underestimated in this region.

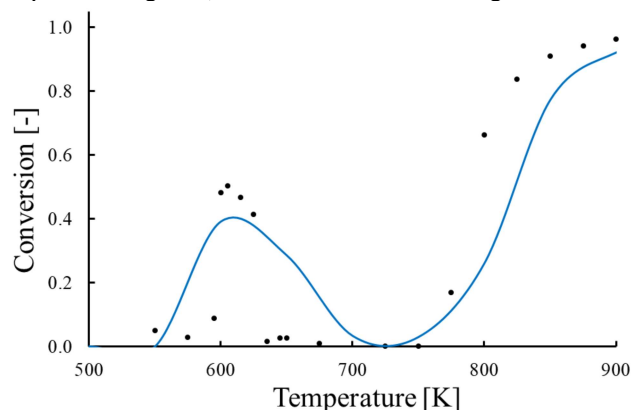


Figure 3: Plot of the cyclohexane conversion in function of the reactor temperature [K] for the oxidation of cyclohexane. The blue line indicates model predictions and black dots represent experimental measurements from Nancy [5].

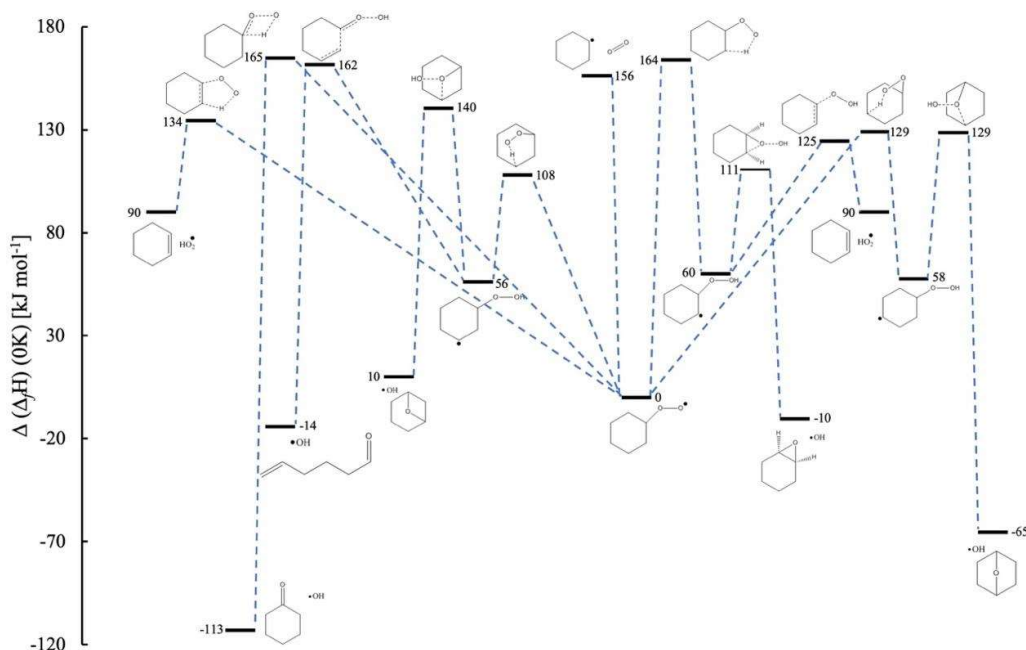


Figure 4: Relevant part of the potential energy surface for the addition of molecular oxygen to the cyclohexyl radical, i.e. the radical formed after hydrogen abstraction from cyclohexane. The values are enthalpies of formation calculated at the CBS-QB3 level of theory at 0 K relative to the cyclohexyl peroxide radical. Reactions are only depicted when the activation energy is below 180 kJ mol⁻¹.

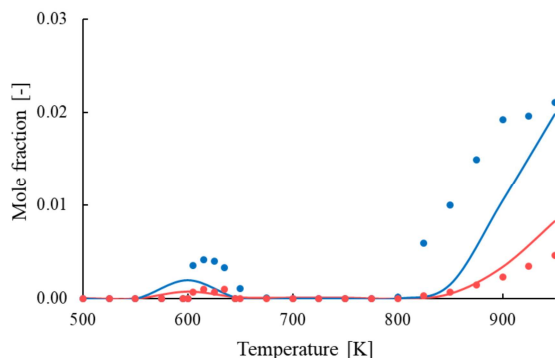


Figure 5: Carbon monoxide (blue) and carbon dioxide (red) production rate in function of the temperature. Lines represent the model predictions and dots represent experimental measurements.

IV. FAST ASSIGNMENT OF LIQUID-PHASE THERMODYNAMIC PARAMETERS WITH GENESYS

Automatic kinetic model generation for liquid-phase processes requires the fast assignment of thermodynamic parameters to solute molecules surrounded with solvent. The Gibbs free energy of a molecule in a solvent, or more generally a liquid phase (G_{liq}) can be calculated by the sum of the Gibbs free energy of a molecule in the gas phase (G_{gas}) adding up the Gibbs free energy of solvation (ΔG_{solv}), cf. Eq. 1, since it are state functions.

$$G_{liq} = G_{gas} + \Delta G_{solv} \quad \text{Eq. 1}$$

To obtain gas-phase parameters, the framework is available in Genesys and now for calculation of ΔG_{solv} an algorithm is included in Genesys which assigns a value based on the availability of data. The algorithm first searches the database with experimental and ab initio data if the combination of solute and solvent is tabulated. Carrying out on-the-fly ab initio

calculations allows to obtain an accurate ΔG_{solv} in the solvent of interest, but this requires a high computational effort. Therefore, a correlation, i.e. the Abraham equation [6], is used to calculate the Gibbs free energy of solvation at standard conditions and 298 K. The latter is an empirical based correlation which calculates the gas-solvent partition coefficient (K_{gas-s}) based on solvent (a, b, c, e, l, s) and solute descriptors (A, B, E, L, S), cf. Eq. 2.

$$\ln K_{gas-sol} = c + aA + bB + eE + lL + sS \quad \text{Eq. 2}$$

Via the relation between the Gibbs free energy and the partition coefficient, the Gibbs free energy of solvation is obtained. Solvent descriptors are tabulated for most common solvents. If solvent descriptors are not available in the database, the correlation cannot be used. Solute descriptors are tabulated for solutes, but in addition these can be calculated from a group additive scheme, i.e. Platts fragments method [7], that calculates the unknown solute descriptors of a molecule based on the Platts fragments the molecule is built from. It is impossible to calculate ΔG_{solv}° if solute descriptors are not tabulated and if not all non-hydrogen atoms can be assigned a contribution via Platts fragments. Currently, Platts fragments are missing for radical species and the peroxide group, both important for oxidation processes.

Validation of both the Abraham equation and the Platts fragments methodology is carried out with two data sets. One set with 3922 data entries in 20 different solvents is obtained from ab initio calculations post-processed by COSMOtherm and the other set is the FreeSolv database with 643 Gibbs free energies of hydration constructed by Mobley et al. [8]. Typical values for the Gibbs free energy of solvation amount to -30 to 0 kJ mol⁻¹.

First, the Abraham equation is tested by calculating the Gibbs free energy of solvation at standard conditions and 298 K for the data points of which both solute and solvent descriptors are tabulated. In Figure 6, the result is depicted of tabulated minus

calculated values with the Abraham equation for ΔG_{solv}° . For 1438 of the 4565 data points, ΔG_{solv}° can be calculated with a mean absolute error (MAE) of 1.1 kJ mol⁻¹. A small error is expected since tabulated solute descriptors are regressed from experimental data. However, 1-octanol stands out as solvent, indicated in Figure 6, with a significant higher MAE of 2.2 kJ mol⁻¹.

Subsequently, the Platts fragments method was implemented with C, H, O, N-containing fragments and validated with the same data set. From the complete data set with 4565 entries, 3549 combinations can be calculated using the Platts fragments method for determination of the solute descriptors which amounts to three times more than only with tabulated solute descriptors. However, the MAE increases to 2.5 kJ mol⁻¹ with 82.6% of the compounds calculated within 4 kJ mol⁻¹. Even for exceptional high energies the calculation of ΔG_{solv}° using Platts fragments is good, e.g. the Gibbs free energy of hydration of mannose amounts to -106.6 kJ mol⁻¹ and the calculated value amounts to -107.7 kJ mol⁻¹.

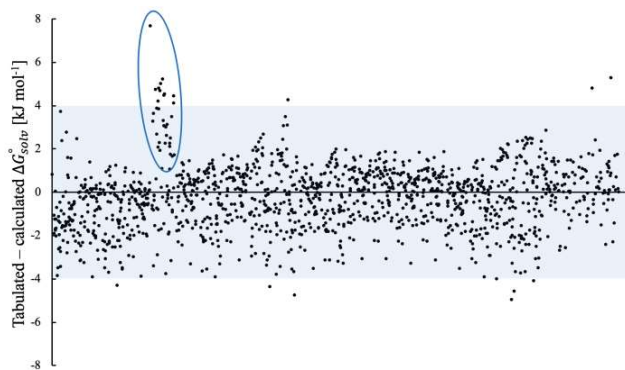


Figure 6: Plot of tabulated minus calculated values for ΔG_{solv}° with the Abraham equation in kJ mol⁻¹ using tabulated solvent and solute descriptors, points in the shaded area are calculated with an error smaller than 4 kJ mol⁻¹. Outliers for the solvent 1-octanol are encircled.

However, if the results with calculated solute descriptors via Platts fragments are considered more in detail, cf. Figure 7, it stands out that several solvents perform better than other ones. Especially the polar solvents perform worse such as water, ethyl acetate, acetonitrile, etc. Moreover, several outliers are detected, covering all functionalities, e.g. acyclic, nitrous, oxygenated compounds which in general do not systematically return in the other solvents. One exception is carbon dioxide which is for every solvent an outlier. Carbon dioxide is via Platts fragments composed out of two double bonded oxygen atoms and a carbon atom bond to two species. With solute descriptors for CO₂ via Platts fragments, inaccurate calculations are performed with a MAE of 10.0 kJ mol⁻¹. As carbon dioxide plays an important role in oxidation processes, it is important to have good values for the solvation of CO₂ in solvents. Therefore, new solute descriptors were regressed for CO₂, listed in Table 2, via least-squares regression of the COSMOtherm data, which results in an improved MAE of only 0.4 kJ mol⁻¹.

Table 2: Regressed solute descriptors for carbon dioxide.

A	B	E	L	S
0.017	0.040	-1.133	0.480	1.101



Figure 7: Plot of tabulated minus calculated values for ΔG_{solv}° with the Abraham equation in kJ mol⁻¹ for tabulated solvent descriptors and solute descriptors calculated by the Platts fragments method. Different colors indicate different solvents: red shades for acyclic hydrocarbon solvents, orange shades for cyclic hydrocarbon solvents, blue shades for oxygenated hydrocarbon solvents and green shades for nitrogen-containing solvents.

V. CONCLUSIONS & FUTURE WORK

A microkinetic model has been constructed for the low-temperature oxidation region with the in-house automatic kinetic model generator Genesys. The generated model is assessed with experiments performed in a jet-stirred reactor at l'Université de Lorraine by the group of Prof. Battin-Leclerc. Experimental trends are well-described by reducing the alternative activation energy of the modified Arrhenius parameter from 24.4 kJ mol⁻¹ to 12.4 kJ mol⁻¹ for the β -scission of a hydroperoxide radical which forms a sensitive reaction. The developed model can be improved by performing ab initio calculations for the high-temperature reaction pathways to obtain a model valid in the low- and high-temperature region. Several elements, e.g. the reaction families, of the developed low-temperature gas-phase model can be reused for the construction of the liquid-phase model since literature states that the same reactions proceed.

Genesys has been extended with an algorithm to calculate the Gibbs free energy of solvation at 298 K and standard conditions via the empirical Abraham equation. Solvent descriptors are available for most common solvents. Solute descriptors are tabulated for several solutes and if unavailable these can be calculated via Platts fragments for most C, H, O and N containing molecules with a mean absolute error of 2.5 kJ mol⁻¹. Though, the peroxide group and radicals, both important for oxidation reactions, are missing which will have to be regressed.

REFERENCES

1. N.M. Vandewiele, K.M. Van Geem, M.-F. Reyniers, and G.B. Marin, *Genesys: Kinetic model construction using chemoinformatics*. Chemical Engineering Journal, 2012. **207-208**: p. 526-538.
2. L.M. Cai, H. Pitsch, S.Y. Mohamed, V. Raman, J. Bugler, H. Curran, and S.M. Sarathy, *Optimized reaction mechanism rate rules for ignition of normal alkanes*. Combustion and Flame, 2016. **173**: p. 468-482.
3. S.W. Benson, F.R. Cruickshank, D.M. Golden, G.R. Haugen, H.E. O'Neal, A.S. Rodgers, R. Shaw, and R. Walsh, *Additivity rules for the estimation of thermochemical properties*. Chemical Reviews, 1969. **69**(3): p. 279-324.
4. M. Saeys, M.-F. Reyniers, V. Van Speybroeck, M. Waroquier, and G.B. Marin, *Ab Initio Group Contribution Method for Activation Energies of Hydrogen Abstraction Reactions*. ChemPhysChem, 2006. **7**(1): p. 188-199.

5. Z. Serinyel, O. Herbinet, O. Frottier, P. Dirrenberger, V. Warth, P.A. Glaude, and F. Battin-Leclerc, *An experimental and modeling study of the low- and high-temperature oxidation of cyclohexane*. Combustion and Flame, 2013. **160**(11): p. 2319-2332.
6. M.H. Abraham, *Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes*. Chemical Society Reviews, 1993. **22**(2): p. 73-83.
7. J.A. Platts, D. Butina, M.H. Abraham, and A. Hersey, *Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach*. Journal of Chemical Information and Computer Sciences, 1999. **39**(5): p. 835-845.
8. D.L. Mobley and J.P. Guthrie, *FreeSolv: a database of experimental and calculated hydration free energies, with input files*. Journal of Computer-Aided Molecular Design, 2014. **28**(7): p. 711-720.

Table of contents

1. INTRODUCTION.....	1
1.1. INDUSTRIAL PRODUCTION OF CYCLOHEXANONE AND CYCLOHEXANOL.....	5
1.2. DISSERTATION OUTLINE	9
1.3. REFERENCES	10
 2. LITERATURE SURVEY - KINETIC MODELLING IN THE LIQUID PHASE	 13
2.1. INTRODUCTION	15
2.2. THERMODYNAMIC AND PHYSICAL ASPECT OF SOLVATION	17
2.3. SOLVATION MODELS FOR ESTIMATION OF GIBBS FREE ENERGY OF SOLVATION	22
2.3.1. Classification of solvation models	22
2.3.2. Empirical solvation models.....	23
2.3.3. Discrete solvation models.....	24
2.3.3.1 Quantum mechanical treatment	24
2.3.3.2 Molecular mechanics treatment	25
2.3.3.3 Combined quantum mechanical and molecular mechanics treatment	25
2.3.4. Continuum solvation models	26
2.3.4.1 General concepts and assumptions	26
2.3.4.2 General continuum models	27
2.3.4.3 COSMO-RS by Klamt	30
2.3.4.4 Solvation model x (SMx) by Cramer and Truhlar	33
2.4. LINEAR FREE ENERGY RELATIONSHIP (LFER) METHODS	37
2.4.1. Historical development	38
2.4.2. Theoretical background of the Abraham model.....	40
2.4.2.1 Solute descriptor 'V' and 'L'	42
2.4.2.2 Solute descriptor 'S' and 'E'	43
2.4.2.3 Solute descriptor 'A' and 'B'	43
2.4.2.4 Lower-case constants in the Abraham LSER.....	44
2.4.3. Determination of Abraham solute parameters	45
2.4.3.1 Solute descriptor 'L' and 'V'	45
2.4.3.2 Solute descriptor 'S' and 'E'	45
2.4.3.3 Solute descriptors 'A' and 'B'	46

2.4.4.	Experimental determination of descriptors	47
2.4.5.	Fast estimation methods for solute parameters using group additivity	48
2.4.5.1	Platts fragments approach	48
2.4.5.2	Application of the Abraham equation in automatic model generators	50
2.4.5.3	Machine learning approach	52
2.4.6.	Temperature dependency of the solvation free energy	53
2.5.	VALIDATION AND ACCURACY OF PREDICTED THERMODYNAMIC PARAMETERS	57
2.5.1.	Accuracy tests	57
2.5.2.	Blind challenges	57
2.5.3.	Experimental data set: Minnesota Solvation database	59
2.5.4.	Experimental data set: FreeSolv database	59
2.6.	LIQUID-PHASE EFFECTS ON KINETICS	61
2.6.1.	Diffusional limitations	61
2.6.2.	Experimental determination of reaction rates	62
2.6.3.	Theoretical determination of reaction rates	63
2.6.4.	Example 1: Hydrogen abstraction reactions	64
2.6.5.	Example 2: β -scission	65
2.6.6.	Example 3: Hydrolysis of formamide	66
2.7.	CONCLUSION	68
2.8.	REFERENCES	71
 3 . METHODOLOGY - GENESYS AND AB INITIO CALCULATIONS.....		79
3.1.	GENESYS: AUTOMATIC MICROKINETIC MODEL GENERATOR	80
3.1.1.	User-defined input	80
3.1.2.	Termination criteria	83
3.1.3.	Reaction generation	84
3.1.4.	Assigning thermodynamic properties	85
3.1.4.1	Group additivity theory	86
3.1.4.2	Group additivity algorithm	87
3.1.5.	Assigning kinetic parameters	89
3.1.6.	Output and post-processing	90
3.1.6.1	CHEMKIN	90
3.1.6.2	The AramcoMech base mechanism	91
3.2.	AB INITIO CALCULATIONS	93
3.3.	CONCLUSIONS	97
3.4.	REFERENCES	98

4 . GAS-PHASE MODEL FOR CYCLOHEXANE OXIDATION	101
4.1. INTRODUCTION	102
4.2. PUBLISHED EXPERIMENTAL AND KINETIC MODELLING STUDIES	104
4.3. KINETIC MODEL CONSTRUCTION	107
4.3.1. Kinetic model generation with Genesys	107
4.3.2. New introduced GAVs for -OOH and -OO•	110
4.4. AB INITIO CALCULATIONS RESULTS	112
4.4.1. Thermodynamic parameters	112
4.4.1.1 Single substituted six-membered ring structures	112
4.4.1.2 Double substituted six-membered ring structures	115
4.4.2. Hydrogen abstractions	118
4.4.3. Energy scans	119
4.4.3.1 Addition of molecular oxygen to the cyclohexyl radical	119
4.4.3.2 Cyclohexyl hydroperoxide oxygen-oxygen scission	120
4.4.3.3 Cyclohexane conformer search	121
4.4.4. Potential energy surface	122
4.5. PERFORMANCE OF THE DEVELOPED KINETIC MODEL	128
4.6. CONCLUSIONS	131
4.7. REFERENCES	133
 5 . FAST ASSIGNMENT OF LIQUID-PHASE THERMODYNAMICS WITH GENESYS	 137
5.1. INTRODUCTION	138
5.2. STATE OF THE ART FOR THE LIQUID-PHASE OXIDATION OF CYCLOHEXANE	139
5.2.1. Literature modelling and experimental work	139
5.2.2. Experimental GC×GC analysis	141
5.3. EXTENDING GENESYS TO A LIQUID-PHASE ENVIRONMENT	143
5.3.1. Molecules	143
5.3.2. Reaction families	144
5.3.3. Reaction rules	145
5.4. FAST ASSIGNMENT OF THERMODYNAMIC DATA: ALGORITHM TO ASSIGN THE GIBBS FREE ENERGY OF SOLVATION	146
5.5. VALIDATION AND RESULTS	149
5.5.1. Available validation data	149
5.5.1.1 COSMOtherm	149
5.5.1.1 FreeSolv database	150
5.5.2. Abraham solute descriptors validation	150
5.5.3. Platts fragments validation	151

5.5.3.1	Complete data set	152
5.5.3.2	Validation for cyclohexane as solvent.....	155
5.5.4.	Improving the performance for carbon dioxide	156
5.6.	CONCLUSIONS.....	158
5.7.	REFERENCES	160
 6 . CONCLUSIONS AND FUTURE OUTLOOK		163
6.1.	CONCLUSIONS	164
6.1.1.	Cyclohexane gas-phase oxidation model	164
6.1.2.	Fast assignment of liquid-phase thermodynamic parameters with Genesys	165
6.2.	FUTURE OUTLOOK AND PERSPECTIVES.....	167
6.2.1.	Cyclohexane gas-phase oxidation model	167
6.2.2.	Extension of Genesys to the liquid phase	167
 Appendices		
A . ACTIVITY COEFFICIENT MODELS		169
A.1.	STANDARD AND REFERENCE STATES.....	171
A.2.	UNIQUAC MODEL	172
A.3.	NRTL MODEL	174
A.4.	UNIFAC MODEL.....	175
A.5.	REFERENCES	177
 B . THERMODYNAMIC PARAMETERS FROM AB INITIO CALCULATIONS.....		179
 C . SOLVENT DESCRIPTORS, SOLUTE DESCRIPTORS & PLATTS FRAGMENTS.....		193
C.1	ABRAHAM SOLVENT DESCRIPTORS	194
C.2	SOLUTE DESCRIPTORS	195
C.3	PLATTS FRAGMENTS	200
C.4	REFERENCES	201

Glossary

Ab initio

A Latin term for “from first principles”. It refers to the fact that the results are obtained by applying the established fundamentals of nature without assumptions or experimental input. Ab initio methods determine the energy of a system by solving the corresponding Schrödinger equation.

Arrhenius equation

The mathematical expression of the rate coefficient $k = A \exp\left(-\frac{E_a}{RT}\right)$ where A is the temperature independent pre-exponential factor and E_a is the activation energy.

Basis set

Set of Gaussian functions used to describe the molecular orbitals in ab initio calculations.

Chemical reaction

Chemical reactions are processes whereby one or more species, i.e. the reactants, are converted into one or more different species, i.e. the products, by rearrangements of the constituent atoms.

Density Functional Theory

Computational method that derives the properties of a molecule based on the determination of the electron density of the molecule. Unlike the wave function, which has no physical reality but is only a mathematical construction, electron density is a physical characteristic of all molecules.

Diffusion

Movement of the center-of-mass of a molecule in a reaction mixture.

Elementary reaction step

The transformation of reactants to products in a single step with a single transition state, *i.e.* without passing over an intermediate.

Enthalpy

H , thermodynamic quantity that is calculated from the internal energy, *i.e.* U , as $H = U + pV$, with p the pressure and V the volume of the system.

Entropy

S , thermodynamic quantity that is related to the disorder, randomness/disorder of the system. A system with a larger number of states which can be occupied will have a higher entropy.

Gibbs free energy

G , thermodynamic quantity that is calculated as $G = H - TS$ with H the enthalpy, T the absolute temperature and S the entropy.

Group additivity theory

Methodology which allows the prediction of molecular properties based on the different groups a molecule is built from.

Level of theory

Approach to solve the Schrödinger equation. In general, there are two degrees of freedom: the treatment of electron correlation and the basis set.

Lumping

Grouping of species, which are considered isomers or homologous species with similar properties, in order to reduce the total number of species in a microkinetic model.

Mechanism

A sequence of elementary steps in which reactants are converted into products, often through the formation of intermediates.

Microkinetic model

A kinetic model that described processes using only elementary reaction steps.

Modified Arrhenius equation

The mathematical expression of the rate coefficient $k = AT^n \exp\left(-\frac{B}{RT}\right)$ where A is the temperature independent pre-exponential factor, T the absolute temperature, n and B constants.

Partition function

A measure for the statistical properties of a system. For a canonical ensemble, the partition function is the Boltzmann sum over the different microstates the system can occupy.

Reaction family

A class of reactions that are characterized by the same pattern of molecular rearrangements.

Solute

The minor component in a solution, dissolved in a solvent.

Solvent

A solvent is the component of a solution that is present in the greatest amount. It is the substance in which the solute is dissolved.

Transition state

Saddle point on the potential energy surface along the minimum energy path. A normal mode analysis on the transition state structure yields exactly one imaginary frequency.

Nomenclature

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ASC	Apparent surface charge
BAC	Bond additive correction
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
CBS	Complete basis set
CDK	Chemical Development Kit
COSMO-RS	Conductor like screening model for real solvents
DFT	Density functional theory
FID	Flame ionization detector
GAV	Group additive value
GC	Gas chromatography
Genesys	Generation of reacting systems
HF	Hartree-Fock
HIR	Hindered internal rotor
InChI	International Chemical Identifier
IRC	Intrinsic reaction coordinate
LCT	Laboratory for Chemical Technology
LFER	Linear free energy relationship
LSER	Linear solvation energy relationship
MAE	Mean average error
MC	Monte Carlo

MM	Molecular mechanics
MS	Mass spectrometry
NNI	Non-nearest neighbor interactions
NIST	National Institute of Standards and Technology
PCM	Polarizable continuum model
PES	Potential energy surface
QM	Quantum mechanics
SCRF	Self-consistent reaction field
SMx	Solvation model x
SOC	Spin orbit correction
RES	Resonance correction
RMG	Reaction Mechanism Generator
RMS	Root mean square
RSC	Ring strain correction
TST	Transition state theory
SMARTS	SMILES Arbitrary Target Specification
SMILES	Simplified Molecular Input Line Entry System
XML	Extended Markup Language

Roman symbols

<i>A</i>	Pre-exponential factor	s^{-1} or $\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$
\tilde{A}	Single-event pre-exponential factor	s^{-1} or $\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$
<i>C_p</i>	Heat capacity	$\text{J mol}^{-1} \text{K}^{-1}$
<i>E_a</i>	Activation energy	kJ mol^{-1}
<i>H</i>	Enthalpy	kJ mol^{-1}
<i>k</i>	Rate coefficient	s^{-1} or $\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$
\tilde{k}	Single-event rate coefficient	s^{-1} or $\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$
<i>n</i>	Constant	-
<i>n_e</i>	Number of single events	-
<i>n_{opt}</i>	Number of optical isomers	-

q	Molecular partition function	-
R	Universal gas constant	$\text{J mol}^{-1} \text{K}^{-1}$
S	Entropy	$\text{J mol}^{-1} \text{K}^{-1}$
T	Temperature	K

Greek symbols

ΔG_{solv}	Gibbs free energy of solvation	kJ mol^{-1}
$\Delta_f H$	Enthalpy of formation	kJ mol^{-1}
$\Delta_r H$	Reaction enthalpy	kJ mol^{-1}
ΔH_{solv}	Enthalpy of solvation	kJ mol^{-1}
$\Delta^\ddagger H$	Standard enthalpy of activation	kJ mol^{-1}
ΔS_{solv}	Entropy of solvation	$\text{J mol}^{-1} \text{K}^{-1}$
$\Delta^\ddagger S$	Standard entropy of activation	$\text{J mol}^{-1} \text{K}^{-1}$
κ	Tunneling coefficient	-

Subscripts

elec	Electronic
rot	Rotational
ref	Reference
trans	Translational
solv	Solvation
vib	Vibrational

Subscripts

‡	Transition state
$^\circ$	Standard

1.

Introduction

1 . INTRODUCTION.....	1
1.1. INDUSTRIAL PRODUCTION OF CYCLOHEXANONE AND CYCLOHEXANOL.....	5
1.2. DISSERTATION OUTLINE	9
1.3. REFERENCES	10

Nowadays, one cannot imagine a world without chemical processes, providing society with a broad spectrum of daily used products. Design and operation of the associated chemical plants requires a thorough understanding of the process, from the molecular scale to the scale of the entire installation. Because experiments are expensive, time-consuming and can only cover a rather narrow range of conditions, simulations are often preferred. The core of these process simulations is the model describing the chemistry itself, *i.e.* the microkinetic model [1].

An accurate chemical microkinetic model is an extremely powerful and valuable tool. The availability of such a model allows integration of the reactor compartment, *i.e.* the place where the chemical transformations take place, into the model covering the overall process. Associated with the overall process model, design and optimization of the process becomes possible [2]. Currently these topics are on the mindset of many people, considering that process optimization can result in significant energy savings and associated reduction of greenhouse gas emissions, more specifically CO₂ emissions, as well as reduction of waste streams.

In the last decade, great progress has been made in the construction of reliable kinetic models for a variety of technologically important processes including combustion, pyrolysis and oxidation of hetero-atomic mixtures in the gas phase [3-6]. However, for a lot of production processes such kinetic models are not available yet, since it is time consuming and very difficult to construct reliable models. The corresponding models typically contain thousands of species, and hence, as indicated in Figure 1-1 several thousands of associated reactions [7]. The general trend in the amount of reactions as function of the amount of species in a kinetic model is linear, although an exponential trend would be expected. Furthermore, increasing the number of heavy atoms, *i.e.* the number of non-hydrogen atoms, in the reactant molecule(s) increases exponentially the size of the reaction network and generation of models by hand becomes thereby tedious, error prone and often incomplete. To prevent this, automatic microkinetic model generation codes have been developed. Some examples of such codes are RMG developed at MIT (based on the earlier codes XMG and NETGEN) [8], RING developed at the University of Minnesota [9], EXGAS developed at l'Université de Nancy [10], the in-house developed tool Genesys [11] and more are listed in a recent review by Van de Vijver et al. [7] considering the general framework of automatic kinetic model generation codes.

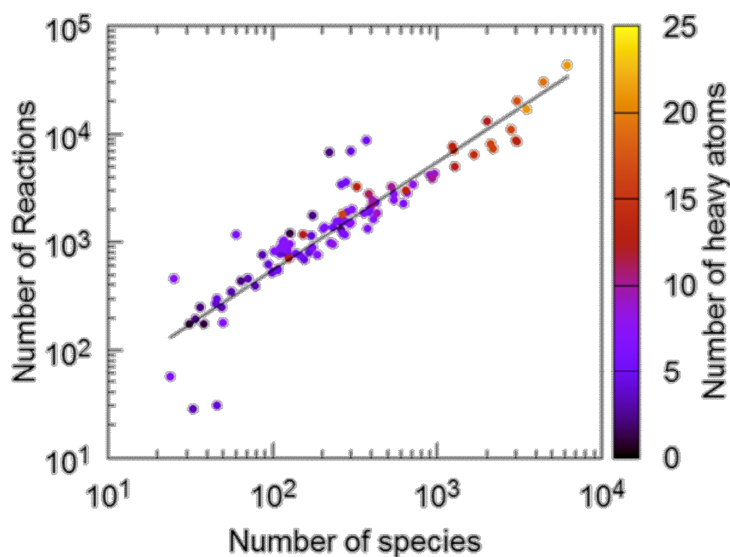


Figure 1-1: The number of reactions as function of the number of species in various pyrolysis and combustion kinetic models published in the last two decades. The secondary axis indicates the number of heavy (non-hydrogen) atoms in the reactant molecule(s). The color bar indicates the size of the reactant molecule. The number of reactions follows a linear trend as function of the number of species and can reach several tens of thousands for large reactants [7].

Accurate (micro)kinetic models for the oxidation of cyclic compounds are crucial to simulate the combustion of typical commercial gasoline, diesel and jet fuels, since these can contain considerable amounts of cycloalkanes, *e.g.* they can represent at least 30 % of diesel fuel [12]. Cyclohexane is therefore frequently used as a simple model surrogate to describe the chemistry of cycloalkanes in real fuels. A lot of research has already been performed on the oxidation of cyclic compounds in the gas phase with several models being available in the open literature for the oxidation of cyclohexane. The combination of automatic kinetic model generation with affordable quantum chemical calculations, to resolve the scarcity of accurate chemical and thermodynamic parameters, forms an outstanding framework for kinetic modelling. Theoretical and computational advances have made predictive first principles based modelling of several important gas-phase processes, *e.g.* steam cracking [13, 14], nowadays feasible. This is in sharp contrast with the limited fundamental understanding and limited performed research on production processes carried out in the liquid phase. Understanding the theoretical aspects for liquid-phase oxidations is essential to allow automatic model generation for relevant processes and is therefore the key objective for further improvement of such processes.

The given master dissertation fits in a bigger project, for which the objective will be to expand the current gas-phase framework of the in-house developed automatic microkinetic model generation tool Genesys, by extending the application domain to liquid-phase processes. The latter extension will require adaptations to the way kinetic and thermochemical parameters are retrieved by including the presence of a solvent. As proof of concept of the newly developed framework, the oxidation of cyclohexane to cyclohexanone and cyclohexanol will be investigated. The oxidation of cyclohexane is

an industrial important process, since it is an essential step in the nylon-6 and nylon-6,6 production chain which is up-to-date still not fully understood.

In the first part of this master dissertation, it will be started with the construction of a gas-phase kinetic model with Genesys which will prove the power of the latter framework. The generated gas-phase model will be validated with experimental data obtained from literature. The second part is to include adaptations in Genesys which will allow the calculation of thermodynamic properties for species in the liquid phase. Therefore, an empirical method for calculation of the Gibbs free energy of solvation at standard conditions will be implemented. The latter will be validated with experimental and ab initio calculated results. The extended framework of Genesys will not be limited to the investigated process, *i.e.* the liquid-phase oxidation of cyclohexane, but will also be applicable to other liquid-phase processes with similar operating conditions. Some other industrial applications include the liquid-phase oxidation of cumene and p-xylene to synthesize respectively phenol and terephthalic acid, the formation of engine and fuel injector deposits and free radical polymerizations in solvents [15].

1.1. Industrial production of cyclohexanone and cyclohexanol

Several important oxidation processes which play an important role in the petrochemical industry are performed in the liquid phase. In this master dissertation, the focus will be on the production of cyclohexanone and cyclohexanol by the liquid-phase oxidation of cyclohexane. Similar to gas-phase oxidations, liquid-phase oxidations have a reaction network with a lot of species and a lot of associated reactions. The latter makes it interesting to extend and apply the developed framework of automatic kinetic model generation for gas-phase processes also to liquid-phase processes.

Cyclohexanone and cyclohexanol are important precursors for the production of nylon-6 and nylon-6,6, as indicated in Figure 1-2. The liquid-phase oxidation of cyclohexane, with a global capacity of 6 million tons per year, is one of the most important production routes for cyclohexanone and cyclohexanol. It is industrially performed in the liquid phase at 140 - 180 °C and 0.8 - 2.0 MPa either uncatalyzed or with a soluble cobalt catalyst [16]. However, the envisaged applicability is broader since the fundamental knowledge of this specific process can be extended to processes with similar operating conditions. For example, the oxidation of p-xylene to terephthalic acid (an important building block for the production of PET) with a capacity of 30 million tons per year and the oxidation of cumene for the production of both phenol and acetone with an annual production capacity of 5 million tons.

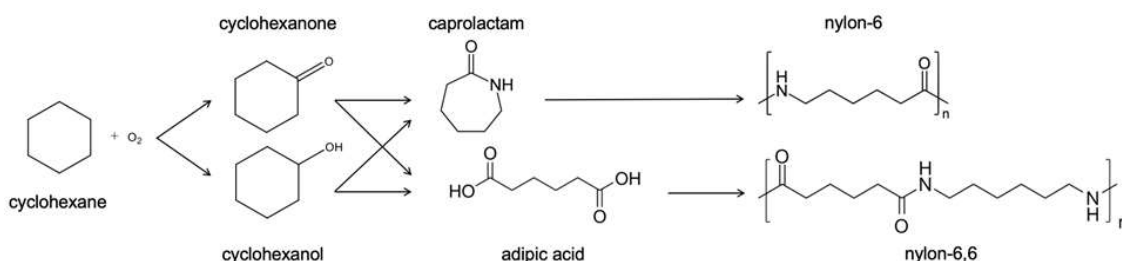


Figure 1-2: Production chain starting with the oxidation of cyclohexane to eventually form nylon-6 and nylon-6,6.

For the oxidation of cyclohexane, two reactants are required, *i.e.* molecular oxygen and cyclohexane. Oxygen is supplied by the addition of compressed air to the first reactor. The main production route for cyclohexane production is the hydrogenation of benzene with a noble metal catalyst [17]. Subsequently cyclohexane is oxidized in a series of agitated reactors with the formation of cyclohexanone, cyclohexanol and cyclohexyl hydroperoxide, this is the so-called air-oxidizing part. The total residence time in these reactors amounts to 15 - 60 minutes. In another reactor, the cyclohexyl hydroperoxide intermediate is further reacted into cyclohexanol and cyclohexanone by using a catalyst. The catalyst has a big influence on the obtained cyclohexanol to cyclohexanone ratio. For example, the normal cobalt-catalyzed deperoxidation results in a ratio of

approximately 3.5 mol cyclohexanol per mol cyclohexanone. However, if chromium(III) is added during deperoxidation, the dehydration of cyclohexyl hydroperoxide is promoted leading to a lower cyclohexanol to cyclohexanone molar ratio of 0.4. An intermediate molar ratio of 1.4 is possible with nickel or palladium as catalyst in the deperoxidation step. Because the products cyclohexanone, cyclohexanol and the intermediate cyclohexyl hydroperoxide are more readily oxidized than cyclohexane, the conversion is kept low, *i.e.* a cyclohexane conversion of around 5 %, to maximize the yield and avoid byproducts formation [15]. The byproducts of the oxidation include a wide range of mono- and dicarboxylic acids, esters, aldehydes and other ring-opened products. In this master dissertation, the focus will only be on the first step, *i.e.* the air-oxidizing part, and more specific the autoxidation case for which no catalyst is used.

Although, the liquid-phase oxidation of cyclohexane is the primary production route of interest, there are still other conventional production routes used in industry, as indicated in Figure 1-3, for the production of cyclohexanol and cyclohexanone.

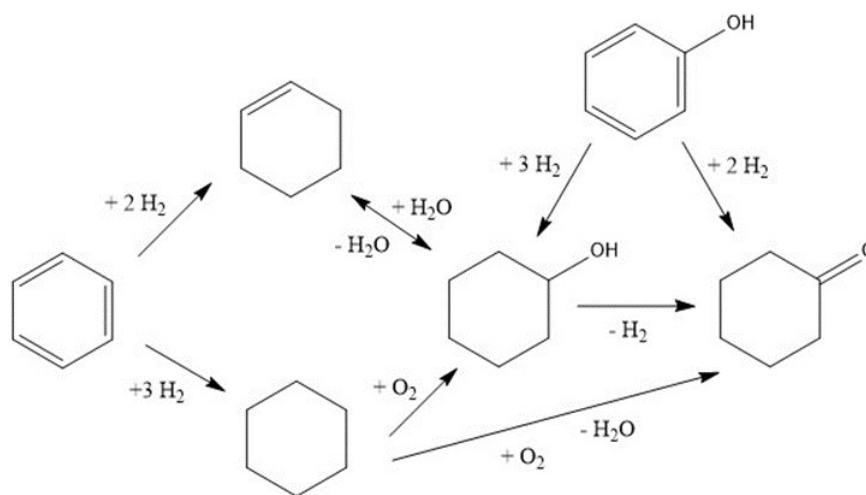


Figure 1-3: Overview of production routes to cyclohexanol and cyclohexanone.

Cyclohexanol, cyclohexanone or a mixture of both depending on the used metallic catalyst, can be produced by hydrogenation of phenol, which was also the first industrial production route. Cyclohexanol can be prepared in high yield, by either gas- or liquid-phase hydrogenation of phenol. Frequently applied for the gas-phase hydrogenation is a supported nickel catalyst which can contain copper, cobalt or manganese. For the latter, yields have been described to achieve 98 % [18]. To obtain cyclohexanone the hydrogenation can be carried out in the gas phase at 140 - 170 °C and atmospheric pressure with a supported noble metal catalyst, *e.g.* palladium, platinum or ruthenium. Yields as high as 95 % have been obtained for complete conversion of phenol. Commercial liquid-phase hydrogenation of phenol with a supported palladium catalyst on carbon results in yields higher than 99 % cyclohexanone with a phenol conversion of 90 %. With a nickel catalyst on the other hand, a selectivity to cyclohexanol of

99.9 % can be obtained. The hydrogenation conditions can be adjusted to produce mixtures of cyclohexanol and cyclohexanone which is desired for adipic acid producers.

Another production route for cyclohexanol is via the synthesis and hydration of cyclohexene. First selective hydrogenation of benzene to cyclohexene is carried out after which non-reacted benzene and cyclohexene are separated by extractive distillation. The cyclohexene is subsequently hydrated with a heterogeneous nickel catalyst and for a conversion of 50 % results in a mixture of about 35 % cyclohexene, 15 % cyclohexane and 50 % unreacted benzene. The selectivity is however highly influenced by the presence of iron and sulfur impurities. Therefore, only very pure benzene and vessels with an inline inert coating material are used. The actual hydration of cyclohexene is performed in a slurry reactor but is limited by the equilibrium constant, which restricts the conversion to approximately 14 %. Although, yields of 95 % and higher are achieved with not a lot of byproduct formation, the disadvantage is that no cyclohexanone is produced with this synthesis method and that the investment cost is higher than that of a conventional oxidation plant [16].

Besides the conventional oxidation process with air, it is also possible to produce cyclohexanol by oxidation of cyclohexane with anhydrous metaboric acid, also known as the Halcon process. A slurry of anhydrous metaboric acid is added to the first stage of several staged air oxidation vessels. During the reaction, cyclohexyl hydroperoxide is trapped as cyclohexyl perborate ester which reacts further with cyclohexane to form cyclohexanol and the borate ester. The advantage is that the borate ester is stable and does not have the intention to form ring-opened products. Hydration of the borate ester ends up with cyclohexanol and boric acid, the latter can be dehydrated and recycled as metaboric acid to the air oxidizer. Notwithstanding, a higher investment cost and production cost (for recovery of the boric acid), a higher yield of cyclohexanone and cyclohexanol is obtained (88 %) and a cyclohexanol to cyclohexanone ratio of 10 is achieved.

Cyclohexanol can also be dehydrogenated to form cyclohexanone which is performed by passing the vapors through a tube furnace at 400 – 450 °C without a catalyst. The purity of the resulting cyclohexanone stream amounts to 98 – 99 %. Catalytic dehydrogenation can be carried out with a variety of catalysts, *e.g.* chromium oxide – copper, copper chromate or nickel, at milder conditions with an even higher selectivity.

Both cyclohexanol and cyclohexanone are precursors for adipic acid and caprolactam [19]. Around 90 % of the caprolactam is synthesized from cyclohexanone. For this synthesis, cyclohexanone is first converted to the oxime (=NOH functional group) [16]. Treatment of the oxime with a strong acid, mostly sulfuric acid induces a Beckmann rearrangement which results in the bisulfate salt of caprolactam. Neutralization of the salt with ammonia releases the caprolactam and cogenerates ammonium sulfate. For the production of caprolactam starting from cyclohexanol, the cyclohexanol is first mildly oxidized

or catalytically dehydrogenated to obtain cyclohexanone which can be converted into caprolactam by the latter route. A mixture of cyclohexanol and cyclohexanone with nitric acid (HNO_3) can be converted catalytically by ammonium vanadate into adipic acid [20]. Nevertheless, the latter production route is claimed to produce 5 to 8 % of the anthropogenic emission of the greenhouse gas nitrous oxide (N_2O), since nitrous oxide is formed as byproduct [21].

Finally, nylon-6, also called polyamide 6 is synthesized by ring-opening polymerization from heating caprolactam above $260\text{ }^\circ\text{C}$ in an inert atmosphere [22]. In Europe, the demand of nylon-6 amounts to a million tons per year with an annual production of 240 ktons per year by the biggest producer of Europe, *i.e.* BASF. Nowadays, nylon-6 is the most significant construction material used in many industries, *e.g.* automotive industry, aircraft industry, clothing industry and electrotechnical industry. Nylon-6,6 is synthesized by polycondensation of equal amounts of hexamethylenediamine and adipic acid [23]. As byproduct water is formed which is removed from the reactor to drive the equilibrium to polymer. The worldwide production amounted to 2 million tons in 2011 which was mainly consumed by the fiber market for the production of carpets and textiles. The rest is consumed in other applications as a thermoplastic mostly by injection molding for 3D structural objects, *e.g.* auto applications, pipes and electro-insulating elements.

1.2. Dissertation Outline

The objective of this master dissertation is twofold. A first goal is to develop a validated gas-phase model for the oxidation of cyclohexane, as a proof of concept, which will confirm the proper functioning of the automatic kinetic model generation framework for gas-phase processes, more specifically the gas-phase oxidation of cyclohexane. The second goal is to get familiar with the effect of liquid-phase effects on thermodynamic and kinetic parameters for kinetic modelling. As a first step towards a fully functioning liquid-phase framework for automatic kinetic model generation, the implementation of an automatic fast assignment algorithm for thermodynamic parameters in Genesys is intended.

In Chapter 2, a literature survey concerning the effect of a liquid phase on kinetic modelling, and more specifically determination of kinetic and thermodynamic parameters for liquid phase processes, is elaborated.

In Chapter 3, the working principle of automatic kinetic model generation and more specifically Genesys is covered. Moreover, the methodology for performing *ab initio* calculations, carried out to obtain accurate gas-phase thermodynamic and kinetic parameters, is explained.

Chapter 4 covers the development and results of the gas-phase model for the oxidation of cyclohexane. *Ab initio* calculations are performed to construct a potential energy surface (PES) for important reactions starting from cyclohexane after the addition of molecular oxygen, as well as calculations to obtain accurate thermodynamic parameters for associated species. To validate the developed model, a comparison is made with experimental data obtained from literature and other models proposed in literature.

In Chapter 5, the first extension of Genesys to include liquid-phase effects is described which will be limited to the assignment of thermodynamic parameters. An empirical methodology is implemented for the calculation of the Gibbs free energy of solvation at 298 K and standard conditions which requires solvent and solute descriptors. The working principle, *i.e.* the algorithm, implemented in Genesys is described. To investigate the accuracy of the newly implemented code, it is validated with experimentally determined and *ab initio* calculated Gibbs free energies of solvation.

In the final Chapter, the conclusions are summarized, and the future work is listed.

1.3. References

1. R. Van de Vijver, B.R. Devocht, K.M. Van Geem, J.W. Thybaut, and G.B. Marin, *Challenges and opportunities for molecule-based management of chemical processes*. Current Opinion in Chemical Engineering, 2016. **13**: p. 142-149.
2. K. Toch, J.W. Thybaut, and G.B. Marin, *A systematic methodology for kinetic modeling of chemical reactions applied to n-hexane hydroisomerization*. AIChE Journal, 2015. **61**(3): p. 880-892.
3. F. Battin-Leclerc, *Detailed chemical kinetic models for the low-temperature combustion of hydrocarbons with application to gasoline and diesel fuel surrogates*. Progress in Energy and Combustion Science, 2008. **34**(4): p. 440-498.
4. F.H. Vermeire, H.H. Carstensen, O. Herbinet, F. Battin-Leclerc, G.B. Marin, and K.M. Van Geem, *Experimental and modeling study of the pyrolysis and combustion of dimethoxymethane*. Combustion and Flame, 2018. **190**: p. 270-283.
5. R. De Bruycker, I. Amghizar, F.H. Vermeire, T. Nyman, M. Hakola, and K.M. Van Geem, *Steam cracking of bio-derived normal and branched alkanes: Influence of branching on product distribution and formation of aromatics*. Journal of Analytical and Applied Pyrolysis, 2016. **122**: p. 468-478.
6. M.V. Khandavilli, M. Djokic, F.H. Vermeire, H.-H. Carstensen, K.M. Van Geem, and G.B. Marin, *Experimental and Kinetic Modeling Study of Cyclohexane Pyrolysis*. Energy & Fuels, 2018. **32**(6): p. 7153-7168.
7. R. Van de Vijver, N.M. Vandewiele, P.L. Bhoorasingh, B.L. Slakman, F. Seyedzadeh Khanshan, H.-H. Carstensen, M.-F. Reyniers, G.B. Marin, R.H. West, and K.M. Van Geem, *Automatic Mechanism and Kinetic Model Generation for Gas- and Solution-Phase Processes: A Perspective on Best Practices, Recent Advances, and Future Challenges*. International Journal of Chemical Kinetics, 2015. **47**(4): p. 199-231.
8. C.W. Gao, J.W. Allen, W.H. Green, and R.H. West, *Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms*. Computer Physics Communications, 2016. **203**: p. 212-225.
9. S. Rangarajan, A. Bhan, and P. Daoutidis, *Language-oriented rule-based reaction network generation and analysis: Description of RING*. Computers & Chemical Engineering, 2012. **45**: p. 114-123.
10. V. Warth, F. Battin-Leclerc, R. Fournet, P.A. Glaude, G.M. Come, and G. Scacchi, *Computer based generation of reaction mechanisms for gas-phase oxidation*. Computers & Chemistry, 2000. **24**(5): p. 541-560.
11. N.M. Vandewiele, K.M. Van Geem, M.-F. Reyniers, and G.B. Marin, *Genesys: Kinetic model construction using chemo-informatics*. Chemical Engineering Journal, 2012. **207-208**: p. 526-538.
12. Z. Serinyel, O. Herbinet, O. Frottier, P. Dirrenberger, V. Warth, P.A. Glaude, and F. Battin-Leclerc, *An experimental and modeling study of the low- and high-temperature oxidation of cyclohexane*. Combustion and Flame, 2013. **160**(11): p. 2319-2332.
13. M. Dente, E. Ranzi, and A.G. Goossens, *Detailed prediction of olefin yields from hydrocarbon pyrolysis through a fundamental simulation model (SPYRO)*. Computers & Chemical Engineering, 1979. **3**(1): p. 61-75.
14. K. Van Geem, M. F. Reyniers, and G. B. Marin, *Challenges of Modeling Steam Cracking of Heavy Feedstocks*. Vol. 63. 2008. 79-94.

15. I. Hermans, P.A. Jacobs, and J. Peeters, *To the Core of Autocatalysis in Cyclohexane Autoxidation*. Chemistry – A European Journal, 2006. **12**(16): p. 4229-4240.
16. M.T. Musser, *Cyclohexanol and Cyclohexanone*, in *Ullmann's Encyclopedia of Industrial Chemistry*. 2011, Wiley-VCH Verlag GmbH & Co. KGaA.
17. G. Luft, A. Laurent, F. Recasens, F. Trabelsi, and G. Weickert, *Chapter 5 - Industrial Reaction Units*, in *Industrial Chemistry Library*, A. Bertucco and G. Vetter, Editors. 2001, Elsevier. p. 243-350
18. G. Rinck, Ulmer, B., Kluy, G., *Comparative consumer testing*. Betriebswirtschaftliche Forschung Und Praxis, 1966. **18**(10): p. 608-608.
19. W.B. Fisher and J.F. VanPeppen, *Cyclohexanol and Cyclohexanone*, in *Kirk-Othmer Encyclopedia of Chemical Technology*. 2000, John Wiley & Sons, Inc.
20. V.N. Parmon, G.I. Panov, A. Uriarte, and A.S. Noskov, *Nitrous oxide in oxidation chemistry and catalysis: application and production*. Catalysis Today, 2005. **100**(1): p. 115-131.
21. R.A. Reimer, C.S. Slaten, M. Seapan, T.A. Koch, and V.G. Triner, *Adipic Acid Industry — N₂O Abatement*, in *Non-CO₂ Greenhouse Gases: Scientific Understanding, Control and Implementation: Proceedings of the Second International Symposium, Noordwijkerhout, The Netherlands, 8–10 September 1999*, J. van Ham, et al., Editors. 2000, Springer Netherlands: Dordrecht. p. 347-358
22. F.F. Liu, J.M. Hurley, N.P. Khare, and K.B. McAuley, *Mathematical Modeling of Nylon 6/6,6 Copolymerization: Beneficial Influence of Comonomers on Degree of Polymerization in Batch Reactor*. Macromolecular Reaction Engineering, 2017. **11**(5): p. 19.
23. R.J. Palmer, *Polyamides, Plastics*, in *Encyclopedia of Polymer Science and Technology*. 2002, John Wiley & Sons, Inc.

2.

Literature survey - Kinetic modelling in the liquid phase

2 . LITERATURE SURVEY - KINETIC MODELLING IN THE LIQUID PHASE	13
2.1. INTRODUCTION	15
2.2. THERMODYNAMIC AND PHYSICAL ASPECT OF SOLVATION	17
2.3. SOLVATION MODELS FOR ESTIMATION OF GIBBS FREE ENERGY OF SOLVATION	22
2.3.1. Classification of solvation models	22
2.3.2. Empirical solvation models	23
2.3.3. Discrete solvation models	24
2.3.3.1 Quantum mechanical treatment	24
2.3.3.2 Molecular mechanics treatment	25
2.3.3.3 Combined quantum mechanical and molecular mechanics treatment	25
2.3.4. Continuum solvation models	26
2.3.4.1 General concepts and assumptions	26
2.3.4.2 General continuum models	27
2.3.4.3 COSMO-RS by Klamt	30
2.3.4.4 Solvation model x (SMx) by Cramer and Truhlar	33
2.4. LINEAR FREE ENERGY RELATIONSHIP (LFER) METHODS	37
2.4.1. Historical development	38
2.4.2. Theoretical background of the Abraham model	40
2.4.2.1 Solute descriptor 'V' and 'L'	42
2.4.2.2 Solute descriptor 'S' and 'E'	43
2.4.2.3 Solute descriptor 'A' and 'B'	43
2.4.2.4 Lower-case constants in the Abraham LSER	44
2.4.3. Determination of Abraham solute parameters	45

2.4.3.1	Solute descriptor ‘L’ and ‘V’	45
2.4.3.2	Solute descriptor ‘S’ and ‘E’	45
2.4.3.3	Solute descriptors ‘A’ and ‘B’	46
2.4.4.	Experimental determination of descriptors	47
2.4.5.	Fast estimation methods for solute parameters using group additivity	48
2.4.5.1	Platts fragments approach	48
2.4.5.2	Application of the Abraham equation in automatic model generators	50
2.4.5.3	Machine learning approach	52
2.4.6.	Temperature dependency of the solvation free energy	53
2.5.	VALIDATION AND ACCURACY OF PREDICTED THERMODYNAMIC PARAMETERS	57
2.5.1.	Accuracy tests	57
2.5.2.	Blind challenges	57
2.5.3.	Experimental data set: Minnesota Solvation database	59
2.5.4.	Experimental data set: FreeSolv database	59
2.6.	LIQUID-PHASE EFFECTS ON KINETICS	61
2.6.1.	Diffusional limitations	61
2.6.2.	Experimental determination of reaction rates	62
2.6.3.	Theoretical determination of reaction rates	63
2.6.4.	Example 1: Hydrogen abstraction reactions	64
2.6.5.	Example 2: β -scission	65
2.6.6.	Example 3: Hydrolysis of formamide	66
2.7.	CONCLUSION	68
2.8.	REFERENCES	71

2.1. Introduction

Accurate chemical kinetic models are extremely powerful and valuable. Many significant public policy and business decisions are made on the basis of kinetic model predictions. For example, the Montreal Protocol, which imposed a worldwide ban on production of certain halocarbons, was based on a kinetic model for the ozone layer [1]. To construct a reliable kinetic model for a chemical process, it is necessary to have both accurate kinetic parameters and thermodynamic parameters [2].

In the past decades, a lot of fundamental knowledge in the chemistry on the molecular scale of gas-phase processes has been gained, which resulted in reliable models for novel process design and full-scale process optimization. Validation of these models has involved a lot of experimental work and provided empirical data both for kinetic and thermodynamic parameters to be stored in databases which now can be used as input for model generation. Nevertheless, for a lot of compounds data is missing, wherefore a wide variety of estimation methods for thermochemical and kinetic parameters for gas-phase reactions have been developed, *e.g.* group additivity values (GAVs), empirical correlations like Evans-Polanyi [3], electronic structure calculations and more recently also machine learning techniques [4] are applied. Since model generation for oxidations involves in most cases several thousands, even up to several ten thousands of reactions, fast estimation methods, such as GAVs, are highly necessary to keep the computational time reasonable. Despite the great progress and good agreement of model predictions with experiments for gas-phase kinetic model generation, liquid-phase processes have been investigated to a much lesser extent and the framework for automatic kinetic model generation is also much less developed. In this literature survey, it is attempted to gain insight in both fast estimation techniques and developed quantum chemistry models to predict the effect of the liquid phase on the reaction rate coefficients and the thermodynamic properties with a view on applications for (computer-aided) (micro)kinetic modelling.

Many cases are known where the gas-phase and liquid-phase reaction rate coefficients differ by several orders of magnitude [5-7]. This can be a direct consequence of the change in barrier height for reaction, *cf.* Figure 2-1, due to the differences in solvation energy between reactants and transition states or other effects such as diffusion limitations, *e.g.* the solvent cage effect [8]. As a result, reaction pathways unimportant in the gas phase may become important in the liquid phase, and vice versa. Thermodynamic equilibria are affected as well by differences in solvation free energy between reactants and products. Moreover, a change in type of solvent can also significantly change the reaction rate coefficients. As an example, the rate coefficients differ three orders of magnitude for the hydrogen abstraction from phenol to cumyloxyl radical with methanol as solvent compared to isooctane as solvent at 298 K [2]. Another example is the equilibrium coefficients (in mol l⁻¹) at 298 K for the dissociation of N₂O₄ that differs with a factor of thousand for reaction performed in the gas phase and in cyclohexane [9]. Hence, the presence

of a liquid phase cannot be neglected for modelling purposes. Estimation of the liquid-phase thermochemistry and kinetics requires an accurate prediction of solvation energies, which is a key subject during this literature survey.

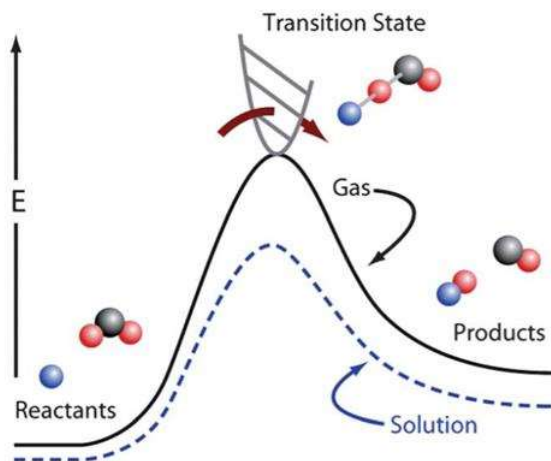


Figure 2-1: Difference in potential energy surface for a gas-phase (black) and liquid-phase (blue) reaction [8].

2.2. Thermodynamic and physical aspect of solvation

A number of different effects contribute to the solvation energy, and it is often convenient to consider them separately. In Figure 2-3, an example is given of a hypothetical solvation path starting from a molecule in a real non-ideal gas mixture to the same molecule in a real non-ideal solution [10]. Eight discrete events are considered during this transfer which all have an associated change in the Gibbs free energy. Taking into account solvation effects ideally requires a model which allows one to approximate all these energy contributions. However, most solvation models described in the literature only estimate the energy change accompanied by transfer from an ideal gas to a dilute state in a pure solvent, as indicated in Figure 2-3 by the blue arrow. The free energy change accompanied by this transition is referred to as ΔG_{solv}^{app} since it is an approximation of the true solvation energy, *i.e.* ΔG_{solv}^{true} . Though, not only gas-solvent transitions are of interest, also the difference in Gibbs free energy related to a transfer of solutes from one solvent to another is sometimes of importance. As the Gibbs free energy is a state function, the energy change by transfer between different solvents (A and B) can be approximated as the total energy change associated with a solvent A-to-gas followed by a gas-to-solvent B process. A good solvation model that accurately describes energy changes for gas-solvent transitions is thus sufficient to also estimate properties related to solvent-to-solvent transitions. Calculations to account for Gibbs free energy contributions from mixed non-ideal solvents and concentrated solutions as well as for non-ideal gas phase contributions are more difficult to perform and are typically not considered in the models further described.

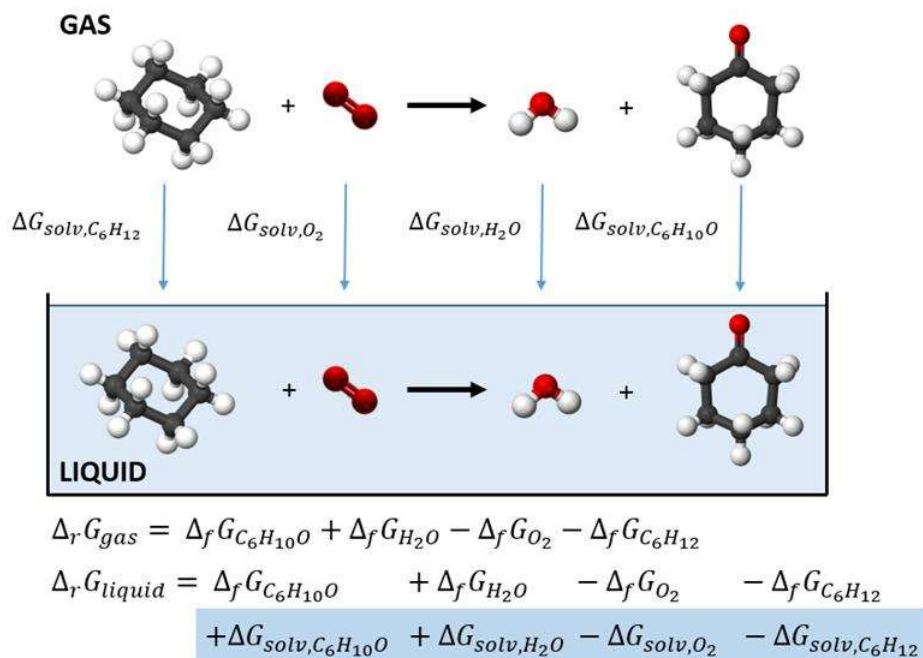


Figure 2-2: Thermodynamic cycle describing the global change in Gibbs free energy of the global oxidation reaction to cyclohexanone in the gas phase and in the liquid phase.

The Gibbs free energy of a molecule in the liquid phase can be written as the sum of the Gibbs free energy of formation of the molecule with the true energy change of solvation as indicated in equation (2-1) which is also depicted in Figure 2-2. The latter is the most correct definition of the Gibbs free energy in a liquid, however it is mostly simplified to equation (2-2), by considering ΔG_{solv}^{app} instead of ΔG_{solv}^{true} .

$$\Delta G_{f,solution} = \Delta G_{f,gas} + \Delta G_{solv}^{true} \quad (2-1)$$

$$\Delta G_{f,solution} \cong \Delta G_{f,gas} + \Delta G_{solv}^{app} \quad (2-2)$$

The formation free energy for most molecules in the gas phase can be obtained by experimental data or by conventional ab initio calculations. Gas-phase electronic structure calculations are nowadays well established and result in reliable outcomes. Without going further into detail in the gas phase, it is mentioned that different methods have been developed which treat the molecules with different levels of theory, *e.g.* Hartree Fock [11], density functional theory (DFT) [12], classical physics (Molecular Mechanics) with corresponding basis sets. From these electronic structure calculations, more specifically from the partition functions, the absolute energy of molecules at 0 K is obtained, which still has to be converted to the enthalpies of formation, *i.e.* $\Delta_f H_{298 K}^\circ$. The latter is done by the atomization energy method which converts the enthalpy of the molecule to the enthalpy of formation at 298 K, by subtracting the experimentally determined atomization enthalpies of the constituting gaseous atoms from the ab initio calculated atomization enthalpy of the species [13].

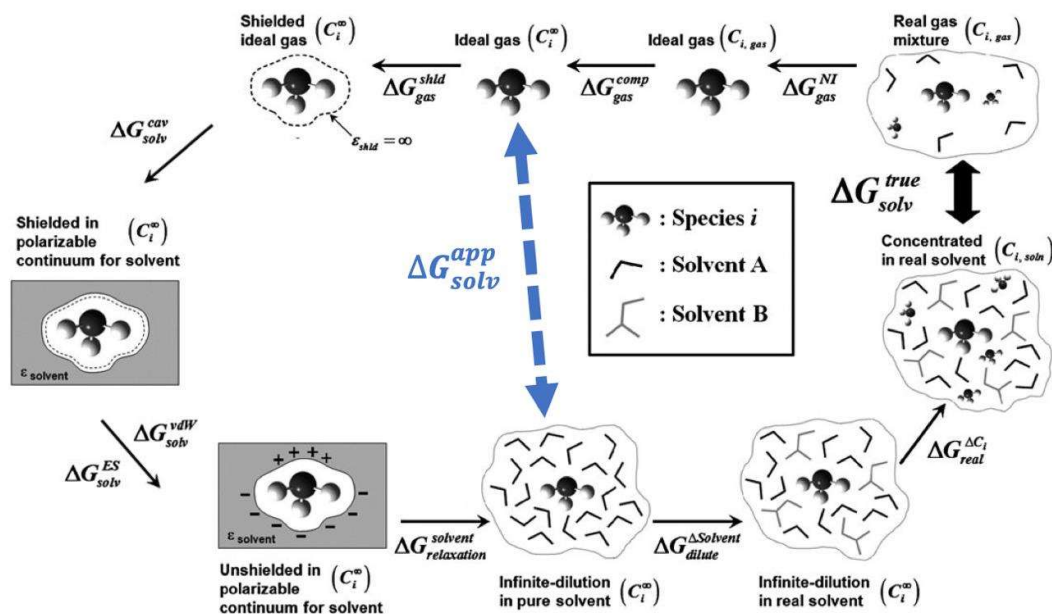


Figure 2-3: Thermochemical cycle for the solvation of molecules from a real gas to a real solution. The blue arrow denotes the transition which is approximated in most solvation models [10].

The Gibbs free energy related to the solvation can be split up into several contributions indicated in equation (2-3) which correspond with the thermochemical cycle indicated in Figure 2-3.

$$\begin{aligned} \Delta G_{solv}^{true} = & \Delta G_{gas}^{NI} + \Delta G_{gas}^{comp} + \Delta G_{gas}^{shld} + \Delta G_{solv}^{cav} + \Delta G_{solv}^{ES} + \Delta G_{solv}^{rep} + \Delta G_{solv}^{disp} \\ & + \Delta G_{relaxation}^{solvent} + \Delta G_{dilute}^{\Delta solvent} + \Delta G_{real}^{\Delta C_i} \end{aligned} \quad (2-3)$$

From this same thermochemical cycle, a definition is given for ΔG_{solv}^{app} by equation (2-4).

$$\Delta G_{solv}^{app} = \Delta G_{gas}^{shld} + \Delta G_{solv}^{cav} + \Delta G_{solv}^{ES} + \Delta G_{solv}^{rep} + \Delta G_{solv}^{disp} + \Delta G_{relaxation}^{solvent} \quad (2-4)$$

A brief description of the eight steps and the according changes in Gibbs free energy, *i.e.* ΔG , are provided next. In a first step, the gas-phase mixture containing the solute must first be taken from the real, *i.e.* non-ideal, state to a hypothetical ideal gas mixture with the same concentration of solute molecules. The presence of solvent molecules is thereby temporarily neglected but will be added back in a later step. Energy changes, *i.e.* ΔG_{gas}^{NI} , related to this type of transitions can be calculated using fugacity relationships and equations of state. Next the ideal gas mixture is expanded, which reduces the concentration of the species to so-called infinite dilution or dilute limit, accompanied with a change in Gibbs free energy ΔG_{gas}^{comp} . The corresponding infinite dilution concentration, *i.e.* C_i^∞ , is by definition small enough that the molecule can accurately be treated as isolated with respect to the surrounding molecules in the gas phase. The molecule is subsequently wrapped in a hypothetical shell that blocks all interactions between neighbor molecules, this serves to screen the molecule and its charge from the surrounding environment. This contribution, ΔG_{gas}^{shld} , is actually equal to zero, since the molecule is already in the dilute, non-interacting ideal gas phase. In a next step, the shielded molecule is inserted into a polarizable continuum representation of the solvent with a corresponding change in free energy of ΔG_{solv}^{cav} . In general, a liquid phase has a much lower free volume than a gas phase, which results in a significant loss of translational entropy. Also, the introduction of cavity formation for the solute in the solvent will possibly induce entropy and enthalpy changes. The actual entropy loss depends on both the size of the solute molecule and the properties of the solvent. Next, the shield is removed from the molecule, which allows interactions between the continuum dielectric and the solute. Two free energy terms are associated with this process, the electrostatic solvation energy, *i.e.* ΔG_{solv}^{ES} , and the solvation energy originating from van der Waals interactions, which is typically written as a repulsion, *i.e.* ΔG_{solv}^{rep} and dispersion term, *i.e.* ΔG_{solv}^{disp} . The polarized continuum models, a certain type of solvation models *cf.* next section, assume a continuum representation of the solvent and do not take into account the contributions further discussed. The indicated blue arrow in Figure 2-3 is thus not applicable for this type of solvation models. In a next step, the continuum representation of the pure solvent is “converted” into a discrete version of the pure solvent which involves a “solvent relaxation” energy, *i.e.* $\Delta G_{relaxation}^{solvent}$. With this conversion to a real solvent, molecules are introduced which will change the solvent and solute

orientations since the continuum model cannot accurately capture the discrete interactions. Good techniques to capture this effect are Monte Carlo or Molecular Dynamics approaches, the reader is referred to the references for more information concerning these approaches [14-16]. Thereafter, the environment is changed from the dilute limit in a pure solvent to the dilute limit in a real multi-component solvent mixture accounted for by $\Delta G_{dilute}^{\Delta solvent}$. The accompanied Gibbs free energy change will depend on the real solvent composition. In a final step, the concentration of the solute is increased from infinite dilution to the actual solute concentration in the solvent, this can be accompanied with the appearance of solute-solute interactions. The infinite-dilution of a liquid solution is defined as a solution that contains so much solvent molecules that when you add more solvent, there is no change in concentration. As a result, no matter how much solvent is added to the solution, the properties of the solute and the system will not change. Moreover, the infinite-dilution limit makes sure that the solute molecule is only surrounded with solvent molecules and that no solute-solute interactions can take place. This excess energy correction can be presented in the form of an activity coefficient. The accompanied change in free energy is indicated with $\Delta G_{real}^{\Delta C_i}$.

Other thermodynamic properties which are linked to transition of molecules from one phase to another phase are the Henry coefficient and partition coefficients. Henry's law states that the dissolved amount of molecules from a gas phase is directly proportional to its partial pressure above the liquid. The proportionality factor is called Henry's coefficient or Henry's constant. Multiple versions of Henry's law are formulated in which the concentration in both phases can have different units and thus care should be taken to make sure one uses the Henry coefficient with correct units. Equation (2-5) is an example of Henry's law in which the Henry coefficient is defined in common units $\text{mol l}^{-1} \text{atm}^{-1}$.

$$C_A = H \cdot P_A \quad (2-5)$$

With C_A the concentration in the liquid phase of component A [mol l^{-1}], H the Henry coefficient and P_A the partial pressure of component A in the gas phase [atm]. Note that the Henry coefficient is a function of temperature and that the temperature dependency can be described by a van 't Hoff equation. The partition coefficient is in essence a generalization of Henry's law which equals the ratio of concentrations of a compound in a mixture of two immiscible phases at equilibrium. This measure is therefore an indication of the solubility of a compound in the two phases. An important example is the octanol-water partition coefficient (K_{ow}) which indicates the potential of a compound to bioaccumulate in organisms, cf. equation (2-6).

$$K_{ow} = \frac{C_o}{C_w} \quad (2-6)$$

With C_O the concentration of a certain molecule in the 1-octanol phase and C_W the concentration of the molecule in the water phase. The link with bioaccumulation is related to the polarity of 1-octanol which is very similar to the polarity of human tissue [17].

In the next section different solvation models are introduced and the different categories to which these models belong, but it should be stressed again that not all models take into account all eight thermodynamic steps as previously discussed.

2.3. Solvation models for estimation of Gibbs free energy of solvation

Computational modelling of liquids and solutions is a rather difficult research topic and multiple research groups around the globe have developed models to capture the solvation effect [10, 18, 19] and more specifically the Gibbs free energy of solvation. The properties of molecules in a liquid phase are influenced by the interactions with fluctuating neighboring molecules, therefore calculation requires an efficient sampling and averaging methodology over all the possible arrangements of solute and solvent molecules. Obviously, this is a very complicated and highly computational task, and thus the need for simplifications is imminent.

Many efficient solvation models have been developed in the last decades to capture the solvation effects. A thorough understanding of these solvation models is important to understand the assumptions made and the accompanied limitations for each group of models. Most solvation models allow a direct calculation of the desired property, *i.e.* the Gibbs free energy of solvation, but also related properties can be obtained from these solvation models, *e.g.* Henry coefficients. However, in detail discussion of all solvation models is out of the scope of this literature study, therefore in the following sections the most important elements of the different categories are discussed as well as a more detailed discussion of two examples of popular, well described solvation models, *i.e.* COSMO-RS [20] and SMx (with x depicting the version number) [18, 21].

2.3.1. Classification of solvation models

The many developed solvation models are classified in different categories depending on the fundamental origin of the model and the approximations which are made for solute and solvent molecules. A first distinction can be made between *ab initio* and empirical methods. Empirical methods or data driven models allow property predictions if sufficient (accurate) experimental data is available for the same or structurally similar compounds. Examples of empirical methods to estimate thermodynamic equilibrium properties are the G^E -models like UNIQUAC [22] and NRTL [23]. UNIFAC [24] is a similar model in the way that it estimates the activity of species in non-ideal mixtures but makes use of group-contribution theory [20, 25]. Of main interest in this work are the linear free energy relationships which do relate Gibbs free energy related quantities, *e.g.* equilibrium coefficients and partition coefficients, via a linear relationship with empirical constants. The Abraham model [26, 27] is discussed since it is the most frequently used relationship in numerous applications including the implementation in automatic kinetic mechanism generation codes, *e.g.* it is used in RMG [28] to estimate the effect of a liquid phase on thermodynamic parameters [29]. The advantage of empirical methods is the low computational effort and the associated accuracy. A trade-off has to be made between the

computational calculation time of a method and the corresponding accuracy, generally a higher computational method results in higher accuracy and the other way around. Note that accurate experimental data is mostly not available for reactive intermediates and transition states despite that it is important for the estimation of reaction rate coefficients.

The increasing computational power has led to the development of atomistic simulation methods with a fundamental description of molecular interactions. These models are referred to as *ab initio* models as they do not require direct experimental input. The strength of these models is the ability to model the system of interest on the molecular level with theoretical fundamentals, *i.e.* first principles based. Several models are available which can again be split up into different categories. A first distinction is made between models that treat all the molecules in a system equally, and methods that treat the solute and solvent molecules differently. Models in the first category surround the solute molecules with a large number of solvent molecules on the same level of theory which are represented explicitly, represented in Figure 2-4 (a) and (b). The other extreme is represented in Figure 2-4 (f), a continuum model treats only the solute molecules at a molecular level and the solvent is represented by a continuum defined by a small number of properties, *e.g.* the dielectric constant or the cohesive energy. Between these extremes, discrete models are available treating the solute and solvent molecules at different levels of theory (Figure 2-4 (c)) and continuum models which explicitly account for some solvent molecules with molecular mechanics or with quantum mechanics. The combination of the solute and some solvent molecules treated on the same or a different level of theory in a continuum is also referred to as a super-molecule approach, for example indicated in Figure 2-4 (d) and (e).

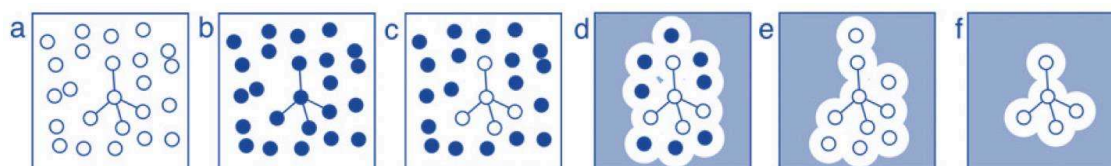


Figure 2-4: From discrete to continuum models. Filled dots represent molecules treated by force field calculations, white dots represent molecules treated by quantum mechanics, colored space represents a continuum dielectricum [10].

2.3.2. Empirical solvation models

A lot of empirical based models have been published in literature related to estimation of properties for species in a solvent. Activity coefficient models and linear free energy relationship methods are the ones discussed in this research to estimate the Gibbs free energy of solvation. Activity coefficient models are used to take into account that solvation models mostly assume an ideal solvent, but in reality a non-ideal solvent has to be considered. This discrepancy can be overcome by calculation of an activity coefficient which can be used to estimate the chemical potential of species more accurately. The actual description and general aspects of some activity coefficient models, the ones which are frequently used in modelling

software, can be found in Appendix A. For more specific information concerning the derivation and the complete set of equations, the reader is referred to more detailed literature.

The linear free energy relationships are not discussed in this section, although that they belong to the category of empirical methods. A separate section is devoted to these methods since they are an interesting starting point for (computer-aided) development of microkinetic models.

2.3.3. Discrete solvation models

Discrete solvation models surround the solute molecule with a large number of solvent molecules, each modeled separately, as indicated in Figure 2-4 (a), (b) and (c). The objective is a detailed description of the sample that is representative for the complete system. Discrete methods are popular for biomolecular systems for which quantum mechanical polarized continuum methods are less used [30]. It is known that the presence of a solvent can introduce notable changes in the molecular structure both for the nuclear and electronic distribution [31]. For example, nuclear structure changes may arise from the tendency of polar solvents to stabilize compounds with large charge separation, *cf.* Figure 2-5. The example is given of glycolaldehyde, for which a transition from the circular conformer to the extended conformer in the gas phase is less favorable, since it has a Gibbs free energy change of +5.4 kJ mol⁻¹. In contrast to a Gibbs free energy change of -12.5 kJ mol⁻¹ in an aqueous solution, which makes the transition favorable in these environments [31]. Moreover, the solvent can also influence the equilibrium concentrations between different isomers. The amount of possible arrangements of the solvent molecule around a solute is countless. Therefore, the solvation free energy and related properties, *e.g.* the effect of the solvent on the geometry, are averaged properties sampled over all the accessible conformations of the molecular system. Monte Carlo (MC) [32] and Molecular Dynamics (MD) [33] algorithms are used thereby for sampling.

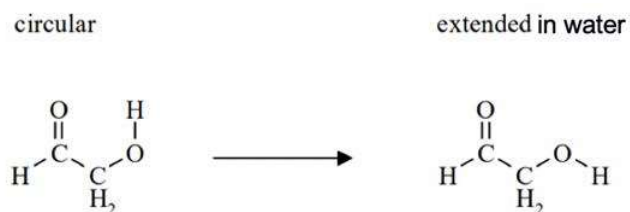


Figure 2-5: Conformational equilibrium of glycolaldehyde (HCOCH_2OH) between circular and extended form in the gas phase and in an aqueous solution [31].

2.3.3.1 Quantum mechanical treatment

The most obvious method is the representation of solute and solvent at the same level of theory. The highest detailed simulations are at a quantum mechanical level of treatment usually with density functional theory as used method in combination with common used basis sets, but also the less detailed

Hartree Fock method is used in literature. An example of a fully quantum mechanical approach is given by the Car-Parrinello method [34], which uses density functional theory and molecular dynamics to average the properties. However, the large number of solvent molecules that need to be treated explicitly in combination with the many MC or MD moves results in very expensive computational tasks. Despite the high computational expense, these methods are not always more accurate than less detailed approaches such as Molecular Mechanics (MM), in part because commonly-used DFT functionals do not accurately describe van der Waals forces. A quantum mechanical method using DFT supplemented with dispersion corrected potentials has been proposed by Aeberhard et al. to better describe these weak interactions which might lead to improved solvation energy calculations [35, 36]. For calculations at this high level of theory, *i.e.* QM treatment of both solute and solvent molecules, the computational effort can amount to a hundred of CPU-hours, already for systems with species containing four heavy atoms and having hydrogen donor and acceptor abilities.

2.3.3.2 *Molecular mechanics treatment*

Due to the high computational cost, it is sometimes more common to use the pure classical molecular mechanics treatment which relies upon parameterized force fields describing the molecular interactions for the complete system of solute and solvent molecules. One of the major difficulties with this approach is the lack of classical force-field potentials that are sophisticated enough to reproduce the bulk behavior of the solvent, such as long-range dielectric effects. Also, the accurate determination of the many parameters in force-field potentials and the challenge of adequately sampling the configuration space of the solvent molecule positions can form a difficulty. Performing classical molecular mechanics simulations are exceptionally useful for the determination of structural and dynamic information, *e.g.* the solvation free energy can be obtained by these simulations. The difference in free energy between two related states from MD and MC simulations can be obtained by two popular methods: Free Energy Perturbation (FEP) [37] and Thermodynamic Integration (TI) [38]. The thermodynamic cycle integration technique, for example, calculates the work, related to the solvation Gibbs free energy, required to gradually change one solvent molecule (from a bulk of molecules) into a solute molecule by gradually changing the force field parameters.

2.3.3.3 *Combined quantum mechanical and molecular mechanics treatment*

With methods combining quantum mechanics and molecular mechanics (QM/MM), a part of the system is treated quantum mechanically and the remainder is modeled using a molecular mechanics force field. Generally, the solute particles are treated with the highest level of theory, *i.e.* QM, and the solvents with a lower level of theory, *i.e.* MM. A combined treatment of the solvent with molecular mechanics, and the solute with quantum mechanics, requires adaptations to account for the solvent phase in the quantum mechanical calculations and the other way around. Usually point charges are assigned to atomic sites of

the solvent molecules. In addition polarizable sites can be added on the solvent molecules to account for polarization of the solvent. Popular molecular mechanics supporting quantum mechanical codes are GROMACS [39] and CHARMM [40]. Computational cost is currently large for calculations with a larger basis set, although advances may speed them up in the future. An example of this hybrid approach is the Effective Fragment Potential (EFP) method [41] which represents the molecules surrounding the quantum mechanical zone as fragments. Each fragment contributes a one-electron potential by Coulombic, induction and repulsion interactions to the *ab initio* electronic Hamiltonian of the solute, both Hartree Fock and DFT calculations can be performed. For simple solvent molecules, *e.g.* water, the computational time is in the order of minutes even for a large basis set such as pVTZ. However, when the solvent molecules become more complicated, *e.g.* acetone or DMSO, the computational time rises to above 100 CPU-hours for the pVTZ basis set [41].

2.3.4. Continuum solvation models

In the polarized continuum solvation models, the solvent molecules are not explicitly considered but are represented by a continuum dielectricum, as indicated in Figure 2-4 (d), (e) and (f). Also in this paragraph, two well-known solvation models from literature are described in some more detail. Both models are classified in the category of continuum solvation models. The first one is the COSMO-RS model developed by Klamt et al. and which has already been used at the LCT amongst others by Gilles Desmet and Maarten Sabbe [42-45]. The second solvation model is a collection of solvation models so-called SMx with x the version number developed by Cramer and Truhlar [18].

2.3.4.1 General concepts and assumptions

The concept of a continuum solvation model can be translated into a division of the system into two parts: a focused part (F) and the remainder (R). The focused part is the treatment of the solute molecule possibly extended with incorporation of some solvent molecules, *cf.* Figure 2-4 (d) and (e), and the treatment of the surrounding solvent molecules are the remainder. The Hamiltonian of the complete system may then be written as in equation (2-7).

$$\hat{H}^{FR}(f, r) = \hat{H}^F(f) + \hat{H}^R(r) + \hat{H}^{int}(f, r) \quad (2-7)$$

Where f and r indicate the degrees of freedom of the F and R parts. The F part is treated with more molecular detail than the R part. In the continuum solvation models the complete $\hat{H}^R(r)$ is eliminated and the total Hamiltonian is reduced to an effective Hamiltonian for the solute in the form indicated in equation (2-8).

$$\hat{H}_{eff}^{FR}(f, r) = \hat{H}^F(f) + \hat{H}^{int}(f, r) \quad (2-8)$$

In this approach, there is no need to have a detailed description of the solvent, only a good description of the interactions is sufficient. The contribution of the R part to the Hamiltonian has to be removed, since a continuum representation of this part will be introduced in the F part of the Hamiltonian, in the other case it would count twice the presence of solvent. In the discrete solvation models the term $\hat{H}^R(r)$ is retained by incorporation of solvent molecules in the system of interest, but the degrees of freedom of R can be reduced by limiting the amount of solvent molecules considered. However, elimination of the solvent Hamiltonian is not sufficient to eliminate the r degrees of freedom since they also appear in the interaction Hamiltonian. An almost complete elimination is obtained by introduction of an appropriate solvent response function $Q(\vec{r}, \vec{r}')$ in which \vec{r} and \vec{r}' indicate position vectors and not the complete set of solvent coordinates. In the effective Hamiltonian, an electronic potential function, *i.e.* \hat{V}^{int} , which is a function of the solvent response function and the amount of solute parameters, is introduced which accounts for the interactions between solvent and solute molecules, as indicated in equation (2-9).

$$\hat{H}_{eff}^{FR}(f, r) = \hat{H}^F(f) + \hat{V}^{int}[f, Q(\vec{r}, \vec{r}')] \quad (2-9)$$

Dielectric continuum solvation models are a good framework for the solvent description when modelling the solute with quantum mechanics. The list of solvation models using this continuum methodology is large, with the general framework in mind some different types are discussed next.

2.3.4.2 General continuum models

A lot of the developed solvation models can be categorized in the family of continuum solvation models or also referred to as polarizable continuum models (PCM). The solvent is not explicitly solved, but described as an infinite, isotropic dielectric into which the solute is placed. Models differ in the way the interaction potential, *i.e.* \hat{V}^{int} , is described between solute and solvent. The solute polarizes the dielectric, which also polarizes the solute. The shape and size of the cavity in which the solute is inserted depends on the method used and will affect the interaction potential, but usually this cavity is divided into a finite number of elements called tesserae. In most continuum solvation methods, the apparent surface charge (ASC) is solved iteratively for each tessera. Thereby, the reaction field of the solvent is self-consistent with the wave function, these methods are also referred to as self-consistent reaction field (SCRF) models. As a result, the keyword to perform calculations in the presence of a solvent, *i.e.* to use a solvation models, in Gaussian [46] is “SCRF” [47]. The electrostatic-dielectric interaction is usually the biggest contribution to the solvation free energy, still additional terms must be added to account for entropy losses, dispersion interactions and other effects. Although these methods are now all classified under the category of ab initio methods, *i.e.* based on first principles, some of these methods still contain parameters in the model which are derived from regression of model predictions to experimental data. Furthermore, most of these models require the input of a dielectric constant of the solvent which has to

be determined by experiments. For most solvents, the dielectric constant is readily available in literature, e.g. the dielectric constant of cyclohexane amounts to 2.0 at 20 °C [48]. Note that the latter parameter is temperature dependent. Sometimes these methods are also referred to as semi-empirical quantum mechanical solvation models.

The original PCM model, which is now referred to as DPCM [49] in which the D stands for dielectric, uses the normal component of the electric field for determination of the apparent surface charges on the tesserae. The integral equation formalism PCM, or short IEFPCM [50, 51], instead uses the electrostatic potential in its derivation. Another variant of PCM is CPCM [19], in which the C stands for COSMO, the conductor-like screening model (COSMO) introduced by Klamt and Schüürmann. This latter method has already been applied for studies at the LCT [44, 45] and is further elaborated in the latest section. To obtain the ASC, COSMO replaces the surrounding medium by a conductor, with an infinite dielectric constant. An extension to COSMO is the conductor-like screening model for real solvents, *i.e.* COSMO-RS [25]. COSMO-RS uses the screening charge densities available from COSMO calculations to model interactions between the molecular surfaces of all the molecules in the liquid. Therewith the thermodynamic properties of mixed fluids, including mixtures of different solvents can be predicted. Most of these methods are executed at the density functional theory level for electronic structure calculations.

An alternative for determining the ASC is expressing the solute charge distribution as a multipole expansion. The easiest model is the Onsager reaction field model, which assumes the cavity (in which the solute is placed) to have a fixed spherical shape in the solvent field. Inside the center of the sphere a dipole is inserted. When a dipole is present in the solute it will induce a dipole to the surrounding medium, and the electric field applied by the solvent dipole will in turn interact with the molecular dipole. This in turn will result in a net stabilization. However, when a system is considered which does not have a dipole moment, *i.e.* the dipole moment of the solute amounts to zero, there will be no interaction with the surrounding field and therefore the Onsager model calculations will give the same results as for the gas phase. This is an inherent limitation of the Onsager approach. Although its simplicity the Onsager model can be used with advanced quantum mechanical approaches.

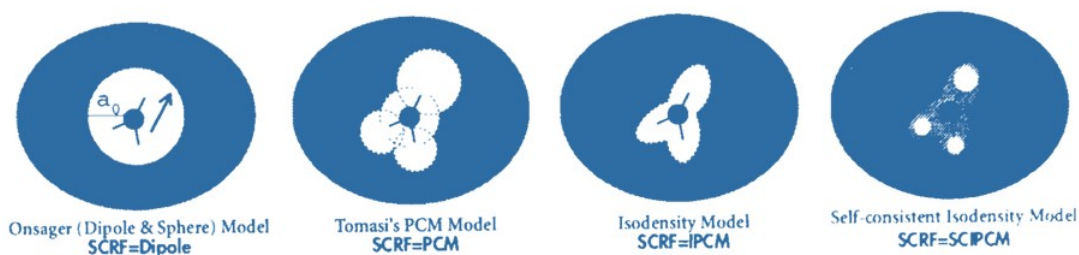


Figure 2-6: Representation of different multipole expansion approaches [52].

Extensions of this simple spherical model include expanding the series to a cavity with multiple moments, expanding over multiple centers and changing shape of the cavity, so-called multipole extension, as indicated by several models in Figure 2-6. For example, the IEFPCM, also called Tomasi's PCM, defines the cavity as a union of different atomic spheres each carrying a dipole, these spheres may overlap each other. Another model is the isodensity PCM (IPCM) [53], which defines the cavity as an isodensity surface of the molecule. This isodensity is determined by an iterative process in which a SCFR cycle is performed and converged using the current isodensity cavity. The resulting wavefunction is then used to compute an updated isodensity surface, and the cycle is repeated until the cavity shape no longer changes upon completion of the SCF. An isodensity surface is a very natural, intuitive shape for the cavity since it corresponds to the reactive shape of the molecule (rather than being a simpler, pre-defined shape such as a sphere or a set of overlapping spheres). However, a cavity defined as an isosurface and the electron density are necessarily coupled. The self-consistent isodensity polarized continuum Model (SCI-PCM) was designed to take this effect fully into account. It includes the effect of solvation in the solution of the SCF problem. This procedure solves for the electron density which minimizes the energy, including the solvation energy, which itself depends on the cavity, which in turn depends on the electron density. In other words, the effects of solvation are folded into the iterative SCFR computation rather than comprising an extra step afterwards. Hence, SCI-PCM accounts for the full coupled relation between the cavity and the electron density which includes coupling terms neglected by IPCM. The Onsager, Tomasi's, isodensity and self-consistent isodensity model are all included in the Gaussian software. In Table 2-1, the arguments for the keyword "SCRF" and the necessary input to launch the different solvation models in Gaussian is indicated.

Table 2-1: Keywords and required input for continuum models included in Gaussian [47].

Model	Keyword	Required input
Onsager model	SCRF=Dipole	Solute radius in Angstroms, Dielectric constant of the solvent
Default model, integral equation formalism PCM (IEFPCM)	SCRF=PCM	Dielectric constant of the solvent
Isodensity Polarized Continuum Model (IPCM)	SCRF=IPCM	Dielectric constant of the solvent
Self-Consistent Isodensity Polarized Continuum Model (SCIPCM)	SCRF=SCIPCM	Dielectric constant of the solvent

Model	Keyword	Required input
Solvation Model (SM)	SCRF=SM	Dielectric constant of the solvent, refractive index, hydrogen bond acidity, hydrogen bond basicity, surface tension, aromaticity, electronegative halogenicity
COSMO	SCRF=COSMO	Klamt's radii, Optional: Dielectric constant of the solvent
COSMO-RS - Generates the data file necessary for COSMO-RS	SCRF=COSMORS	None, structures may be optimized prior with SCRF=COSMO

Moreover, the Generalized Born (GB) approximation methods are considered as an extreme example of multi-center multipole expansion, with expansions centered on each nucleus but truncated at the first term. A difficulty is the determination of the Born radii for each nucleus. The most widely used GB methods are collected under the name SM x , where x represents the version number, introduced by Cramer and Truhlar. Additional terms are introduced in the solvation models to represent non-electrostatic contributions from the cavitation, dispersion and solvent structural rearrangement on the solvation free energy. Due to the popularity of this model, it is described in more detail in the next section.

A general weakness of the continuum solvation models is the inability to model specific interactions between the solute and nearby solvent molecules. An option is to include some of the solvent molecules in the calculations as a kind of "supermolecule". Other options are quantum mechanical/molecular mechanics/continuum methods in which some solvent molecules are treated with force fields surrounded by a polarizable continuum. An example of such an explicit/implicit solvation model is the general liquid optimized boundary (GLOB) model.

2.3.4.3 COSMO-RS by Klamt

In 1995, the first work on COSMO-RS was published by A. Klamt [20]. The quantum chemical basis of COSMO-RS is COSMO, *i.e.* the CONductor-like Screening MOdel, which is classified in the category of quantum mechanical continuum solvation models. The molecule is treated as if it is embedded in a dielectric medium via a molecular surface or "cavity" that is constructed around the molecule. Normally, the macroscopic dielectric constant of the solvent is used. COSMO is a quite popular model that has been incorporated in many quantum chemistry program packages, *e.g.* Gaussian, Molpro and Turbomole. If combined with accurate quantum mechanical calculations, COSMO-RS has been proven

to produce good results for Henry coefficients and partition coefficients. The necessary input file for COSMO-RS calculations can be generated with Gaussian by using the 'SCRF=COSMORS' keyword. There is an additional option to add as extra input the solvent dielectric constant which will make sure that the lowest energy conformer is searched before the actual COSMO calculation, but this is in essence not necessary for the COSMO calculations. Since DFT methods are available with about the same computational costs, and with more consistent inherent electrostatics, the use of HF-level calculations is at present not recommended [54]. COSMO*therm* is the tool developed by the group of Klamt to perform the property estimation of liquids by calculation of the chemical potential of almost any molecule in almost any pure or mixed liquid at variable temperatures [55]. This is the key for the prediction of for example solubility, partitioning, vapor pressures, complete phase diagrams and the Gibbs free energy of solvation.

From the perspective of dielectric theory, vacuum with a dielectric constant which amounts to 1, and a conductor, corresponding with a dielectric constant of infinity, are the extremes to embed solutes. Where most computational chemistry methods start with a molecule in vacuum as reference, the COSMO-RS theory starts from molecules embedded in a conductor for fluid phase simulations. Before performing COSMO-RS calculations, one has thus to perform COSMO calculations to obtain the self-consistent state, *i.e.* the energy, the electron density, the polarization charge densities, and the geometry of the molecule in a virtual conductor. This state is also referred as "the COSMO state".

The conceptual value of the COSMO state as reference state for molecules in the liquid phase was only detected in the context of the COSMO-RS (Real Solvent) theory. In a perfect conductor all interactions are completely screened on the conductor interface, *i.e.* on the surface of the solute, by the conductor polarization charge density σ . Therefore, for an ensemble of molecules virtually moving in a conductor, there are no intermolecular interactions. Hence each molecule in the COSMO state can just be considered individually. This decoupling makes the COSMO state a very attractive, clean reference state. In Figure 2-7, an example is given of a σ -profile for water, *i.e.* a histogram of the molecular COSMO surface with respect to the polarization charge density. In the profile of water, two pronounced peaks are observed at -1.5 e nm^{-2} and $+1.5 \text{ e nm}^{-2}$, which result from the polar hydrogen atoms and the lone pair regions of the oxygen atom, respectively. From this picture it is also easy to introduce the color coding of the COSMO cavities with respect to σ ; blue stands for a surface area with negative polarization charge density, *i.e.* positively polar parts of the molecule, green for neutral parts of the surface and red for positive parts of the COSMO surface. The sign inversion between the polarization charge density and the molecular polarity just originates from the fact that the conductor has to compensate for the molecular polarity by opposite polarization charge density.

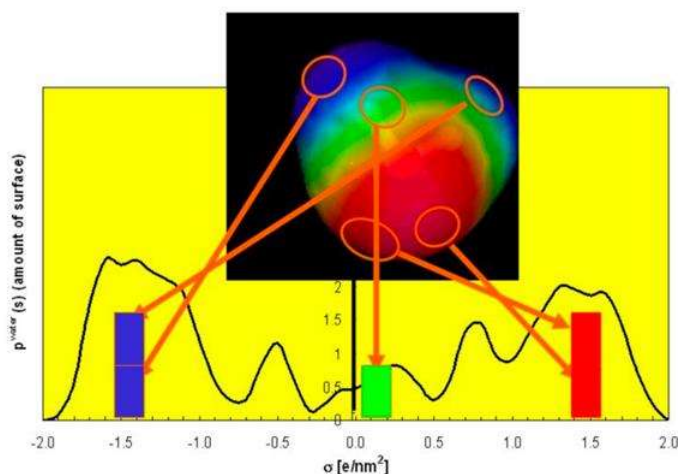


Figure 2-7: σ -profile of water, x-axis: σ , *i.e.* the polarization charge density, [$e \text{ nm}^{-2}$] and y-axis: amount of surface (non-normalized) [25].

In reality molecules are not moving separately in an ideal conductor but do have interaction with each other. Starting from a reference ensemble of molecules swimming in the conductor, in which each molecule has its COSMO “state” with its corresponding energy and its polarization charges, it is tried to approximate a closely packed liquid system by the iterative introduction of molecular contacts. Thus, from a thermodynamic point of view, the concept is a free enthalpy or Gibbs free energy concept of non-compressible fluids.

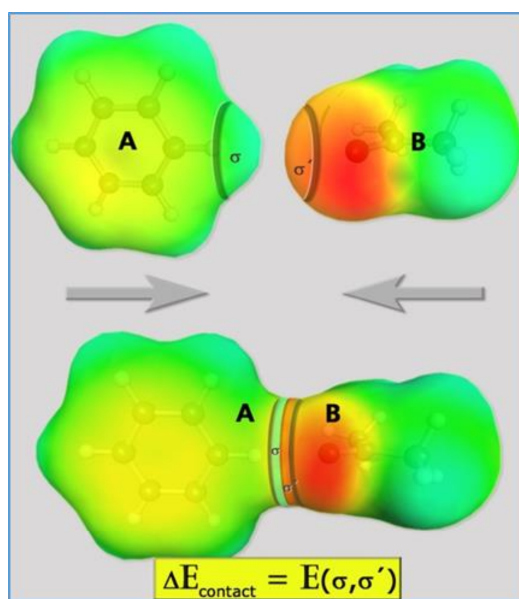


Figure 2-8: Schematic representation of the COSMO-RS interaction concept (ΔE is the difference in energy when both molecules are in contact and when they are separated, σ and σ' are the polarization charge density at the contact surfaces) [25].

For example, two molecules are considered A and B, which may be different or from the same chemical nature, and their virtual distance is now reduced until they make close contact with each other as depicted in Figure 2-8. As there are no interactions between conductor-screened molecules, there is no energy change coupled to this approximation. Only in the final step, when the surfaces of both molecules will chemically interact with each other and the conductor is removed between them at a certain surface area, a change of energy is accompanied since the AB complex is now screened by the conductor as a whole. When the polarization charge densities of A and B, *i.e.* σ and σ' , on their contact surface area are equal but just opposite in sign, then the electrostatic energy change amounts to zero because in total there is zero screening charge between the two molecules shortly before and after making the close contact. Thus, the electrostatic energy and the polarization charge densities on the remainder of the conductor surface do not change during optimal contact.

While the considered approach above just represents a single liquid-like configuration of our molecular ensemble, the only way to calculate the macroscopic thermodynamic properties of a liquid system is application of statistical thermodynamics, *i.e.* to calculate thermodynamic averages over all possible configurations of the liquid via the Boltzmann distribution. This is usually done employing MC or MD techniques in the context of force-field based energy expressions. The resulting chemical potential can subsequently be used for calculation of other thermodynamic properties, *e.g.* activity coefficients, solubility, partition coefficients, vapor pressure and the Gibbs free energy of solvation. It is obvious that such averaging always goes along with a loss of information. Statistical averaging of the COSMO-RS energy expression by MC or MD methodologies, in contrast to force-field energy expressions, does not explicitly depend on the full 3D geometry of the ensemble, but boils down to averaging the sum of local surface contact energies. These complicated statistical thermodynamics of the nestled, 3D molecules is reduced to the much simpler statistical thermodynamics of just independently pair-wise interacting surface pieces. The given approach accounts for group-group interactions, *e.g.* hydrogen bonding, mesomeric effects and inductive effects. Generally, it is not necessary to use the same computational level for the COSMO-RS calculations for ΔG_{solv} and the ab initio quantum mechanics level used for calculation of the formation Gibbs free energy and thermal contribution [56]. Below the assumptions are listed up which are included in the COSMO-RS theory:

- The liquid state is incompressible;
- All parts of the molecular surface can be in contact with each other;
- Only pairwise interactions for the molecular surface patches are allowed.

2.3.4.4 Solvation model x (SM x) by Cramer and Truhlar

In 1991, the first solvation model, *i.e.* Solvation Model 1 (SM1), in a series of models has been published by Cramer and Truhlar which is able to describe the solvation of ions and molecules in an aqueous

solution, well suited for description of chemical reaction dynamics and reaction intermediates [57]. In particular the SM8 model is well described in literature and is therefore discussed. It has had meanwhile several predecessors (SM1-7) [58, 59]. The SM8 model is stated to be both a universal model, *i.e.* it can be used for any liquid condensed phase for which certain bulk properties can be obtained, and it is the most accurate continuum solvation model present at the moment of publishing in 2008 for prediction of Gibbs free energies of solvation. More recent newer solvation models have been published by the same authors, *e.g.* SM12 [60] and SMD [61] where the D stands for density to denote that the full solute electron density is used. While for the other developed models, *i.e.* SMx with x a number, it was necessary to define partial charges on all atoms of the solute molecule. In addition another variant on the solvation models was developed, *i.e.* SMVLE, which explicitly takes into account the volume polarization and local electrostatics as new elements beyond these already available in previous SMx versions [62]. Though, the theoretical principles behind the solvation models remain the same. Similar to COSMO-RS, it allows the calculation of the Gibbs free energy of solvation and partition coefficients, but the model is also particularly effective for determination of pK_a values and oxidation and reduction potentials in contrast to COSMO-RS. Both Hartree-Fock and density functional theory methods are used for the electronic structure calculations [63, 64].

For the solvation models, the Generalized Born approximation is used. In this approach, the quantum mechanical charge distribution of the solute is represented by atom-centered monopoles, *i.e.* partial atomic charges q_k on atom k . The resulting reaction field potential, *i.e.* V_k , for atom k by these charges is then given by equation (2-10).

$$V_k = \left(1 - \frac{1}{\varepsilon(T)}\right) \sum_{k'} q_{k'} \gamma_{kk'} \quad (2-10)$$

with $\varepsilon(T)$ the permittivity of the medium which is temperature, *i.e.* T , dependent and $\gamma_{kk'}$ is an actual Coulomb integral between atom k and k' . A good approximation for this latter term is the following formula, *i.e.* equation (2-11).

$$\gamma_{kk'} = \left(r_{kk'}^2 + b_k b_{k'} e^{-\frac{r_{kk'}^2}{d_{kk'} b_k b_{k'}}} \right)^{-\frac{1}{2}} \quad (2-11)$$

with $r_{kk'}$, an interatomic distance, b_k the effective atomic Born radius and $d_{kk'}$, a parameter for the model without direct physical meaning. The different models of Cramer and Truhlar differ in the way the atomic Born radius is calculated and the extra parameter, *i.e.* $d_{kk'}$, which is introduced into the model. Also, for the calculation of the partial atomic charges, different theoretical formulas, *i.e.* charge models (CM), have been used. Determination of the unknown parameter, $d_{kk'}$, is obtained by fitting modelled data to the experimental available data. Also in the formulas for determination of the partial atomic

charges, empirical constants pop up. The earlier solvation models were not universal but were fitted to specific groups of chemical compounds and application for compounds out of this regression group was thus not advised. Although it is unclear for which data the model has been fitted, it performs readily better than other models as indicated in , data taken over from [18]. For temperatures different than the temperature of fitting, *i.e.* 298 K, an extension of the solvation model is available the so-called SM8T model. Due to the input of experimental data in the model, it is often referred to as a semi-empirical quantum mechanical solvation model. In Table 2-2, the better performance of SM8 compared to other common solvation models is indicated for neutral, cationic and anionic species. Note that a lot of parameters are required to launch the Solvation Model (SM) of Truhlar and Cramer (*cf.* Table 2-1) in Gaussian, but all these parameters are available for common used solvents in the Minnesota database freely available for non-commercial usage [65].

Table 2-2: Mean unsigned error [kJ mol⁻¹] in Solvation Free Energies calculated using SM8 and other continuous solvation models^a [47].

Solute class	# samples	SM8	IEF-PCM G03/UA0	C-PCM* GAMESS	COSMO-RS* NWCHEM
Neutral	940	2.5	23.7	10.2	17.9
Cations	124	16.3	53.1	42.6	35.6
Anions	208	19.0	33.3	30.3	62.3

^a The calculations were performed at the mPW1PW/6-31G(d) level of theory, except for the calculations indicated with an asterisks, which used B3LYP in place of mPW1PW.

In general, continuum solvation models differ in the manner in which the potential, *i.e.* V , is constructed. Also the different solvation models of Truhlar and Cramer differ in this way. Once the definition of the potential for the solvation model is known, the electronic structure calculations can be started from which the electrostatics part of the Gibbs free energy of solvation is obtained by as stated in equation (2-12). Note that this formula is generally applicable to all continuum solvation models and thus not solely for the solvation models of Truhlar and Cramer.

$$\Delta G_{ENP} = \langle \Psi^1 | H + \frac{1}{2} V | \Psi^1 \rangle - \langle \Psi^0 | H | \Psi^0 \rangle \quad (2-12)$$

where the subscript ENP denotes electronic, nuclear and polarization components of the free energy and the superscripts 0 and 1 refer respectively to the gas and liquid phase. The factor $\frac{1}{2}$ in the equation originates from the linear response of the surrounding medium to the solute's charge distribution because half of the induced favorable solute-solvent is canceled by the reorganization of the solvent. Since V is a function of Ψ^1 , the corresponding Schrödinger equation is non-linear and requires an iterative solution.

The temperature dependency of the SM8T model is fully situated in the ΔG_{ENP} term. The total free energy of solvation can eventually be calculated by using equation (2-13).

$$\Delta G_{solv}^{\circ} = \Delta G_{ENP} + G_{CDS} + \Delta G_{conc}^{\circ} \quad (2-13)$$

with the term G_{CDS} including the free energy changes associated with solvent cavitation, changes in dispersion energy and possible changes in local solvent structure. The final term ΔG_{conc}° , accounts for the difference in molar concentrations of the two phases in their standard states and can be linked to the dilution steps from real concentration to infinite-dilution or vice versa in Figure 2-3. If this term is omitted, the equation is only valid for mixtures in which the concentration in the gas and liquid phase equals 1 mol l⁻¹ at 298 K [66]. The terms corresponding with ENP and CDS are not separable thermodynamic observables, only their sum is physically meaningful and corresponds with the other steps of the thermodynamic cycle represented in Figure 2-3. Though, one or the other of these terms may be expected to dominate under certain conditions, for example ΔG_{ENP} is expected to be much larger than G_{CDS} for a charged solute, while the reverse can be expected for a large, uncharged solute that lacks any polar functionality.

2.4. Linear Free Energy Relationship (LFER) methods

Parallel to the atomistic-simulation methods, a completely different approach for predicting solvation energies has been developed. In contrast to the previously discussed theoretical models which are mainly based on fundamental knowledge, this approach is based on empirical input. The term Linear Free Energy Relationships (LFER) is used to denote relationships between logarithms of reaction rates and/or equilibrium coefficients, *i.e.* Gibbs free energy related quantities [67]. These relationships are formulated by using empirical constants and are linear with respect to at least one parameter. As is generally the case for thermodynamics, these LFER methods can be used without detailed knowledge of the reaction mechanism. Due to the empirical nature of this approach, the only proof of its correctness is the agreement of the estimation with experiments.

The oldest known and simplest LFER is the Hammett equation [68]. This relationship, developed in organic chemistry, is used to relate equilibrium coefficients and reaction rates for many reactions involving benzoic acid derivatives with meta- and para-substituents. Usually the corresponding relationship are expressed in the form of equation (2-14) and equation (2-15) for respectively the reaction rate and the equilibrium coefficient.

$$\log k = \log k^0 + \rho\sigma \quad (2-14)$$

$$\log K = \log K^0 + \rho\sigma \quad (2-15)$$

In these equations, k and K represents respectively the reaction rate and equilibrium coefficient for a compound with substituent R and the corresponding k^0 and K^0 refers to a reference constant when the substituent is a hydrogen atom. The substituent constant, *i.e.* σ , depends only on the specific substituent and the reaction constant, *i.e.* ρ , depends only on the type of reaction considered. Other examples of LFERs are the Taft equation [69], the Evans-Polanyi principle [3] and the Brønsted equation [70].

Similarly, LFER methods have been developed for liquid-phase properties and more specifically estimation of the Gibbs free energy of solvation, also commonly referred to as linear solvation energy relationships (LSERs). These relationships start from a set of parameters to characterize the solvent and the solute, so called solute descriptors (P_n) and solvent descriptors (p_n). The gas-solvent or solvent-solvent partition coefficient is then estimated by a formula in the form of equation (2-16).

$$K = \sum_n p_n P_n \quad (2-16)$$

This partition coefficient can then be used to estimate to the Gibbs free energy of solvation, as indicated by equation (2-21).

$$K = \exp\left(-\frac{\Delta G_{solv}}{RT}\right) \quad (2-17)$$

If the solute and solvent descriptors are available, this very simple formula is able to predict solvation free energies with an accuracy similar to those of computationally intensive methods. However, the complete set of parameters for solutes is only available for a limited list of stable molecules and also the availability of solvent descriptors is limited. This makes it less useful for estimation of kinetic parameters, especially for solutes, since there is no data for transient species, *i.e.* unstable, highly reactive species for which experiments are difficult, and transition states. Although there is a fundamental aspect to the meaning of the descriptors, it is still an empirical method, as the solute and solvent parameters are obtained from experiments.

2.4.1. Historical development

The earliest LFER method for liquid-phase property estimation was introduced by Kamlet and Taft who developed a “solvatochromic comparison method” [71-73]. Solvatochromism is the term used to describe the phenomenon of a change in color when a solute is dissolved in different solvents. The idea was that the electronic transitions that occur when a solute is transferred to the solvent phase, could provide a quantitative measure of the types and relative strengths of the solute-solvent interactions responsible for these transitions. More specifically, the transitions were attributed to three types of solute-solvent interactions, *i.e.* dipolarity/polarizability related to electrostatic interactions, hydrogen bond donation and hydrogen bond acceptance. Which led to the establishment of three solvent parameters, respectively given by π^* , α and β , one for each type of interaction. Kamlet and Taft proposed a simple and general form for spectroscopic properties XYZ, *e.g.* measures for UV/Vis spectra or X-ray absorption spectra, of a solute in a solvent given by:

$$XYZ = XYZ_0 + s\pi^* + a\alpha + b\beta \quad (2-18)$$

In which XYZ_0 refers to the property XYZ in a reference solvent, *i.e.* XYZ is a property related to the same solute but in another solvent than the reference solvent. The so-called solvatochromic coefficients s , a and b can be seen as sensitivity parameters of the solute property SP to changes in the corresponding solvent parameters. The developed theory was motivated by the fact that the electronic transitions accompanied by a transfer of a solute molecule to the solvent phase (which causes the measurable spectroscopic shifts), could provide a quantitative measure of the types and relative strengths of the solute-solvent interactions.

Following the work on the solvent parameters, Kamlet and Taft extended their work by introduction of solvatochromic solute parameters, with the assumption that the π^* , α and β for a solvent molecule are adequate enough to describe its behavior as a solute in another medium, *i.e.* the solvent parameters are

used as solute parameters. One immediately recognizes a drawback as the chemical environment of a molecule as a solute drastically differs from that as a solvent. In a solvent each molecule is surrounded by molecules like itself, while as a solute it is surrounded by solvent molecules that can have quite different properties, as indicated in Figure 2-9. The ordered hydrogen bond interactions between the water molecules (black dots) are replaced with new less favorable interactions between water and cyclohexane (yellow dots). Note that cyclohexane is immiscible in water since cyclohexane is an apolar molecule and water is a polar molecule, which is an indication of the less favorable interactions between these compounds. Compounds, such as alcohols, which are highly associated as solvents, are expected to behave differently as solvent compared to when single molecules are introduced in another kind of solvent, *i.e.* having other molecules as near neighbors. Although the rough approximations, the idea of using solvent descriptors for an initial estimate for the descriptors of the same species as a solute is viewed as one of the cornerstones in the development of solute descriptors.

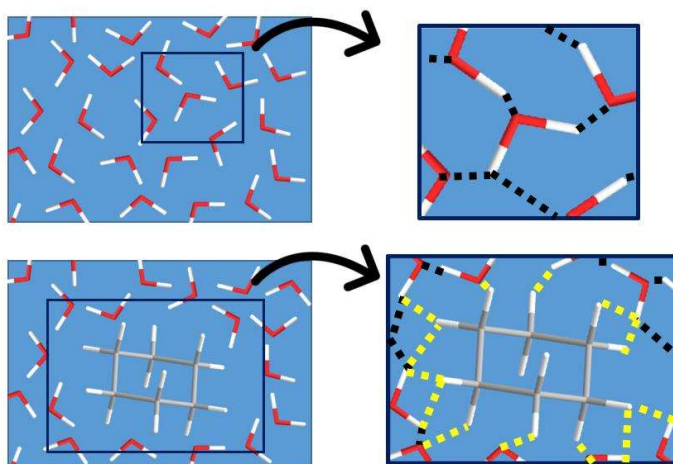


Figure 2-9: Indication of solvent-solvent (blue dots) interactions and solvent-solute (yellow dots) interactions.

Still the solvatochromic comparison method had some disadvantages when looking at the physical connection of the solvation process. Cavity formation is a crucial step in the solvation process, which is not taken into account as a solvatochromic parameter. Moreover, the solvatochromic parameters are derived from spectroscopic energies, which are not expected to reveal all aspects of the physical solvation and do only partially describe the solvation process. These shortcomings of the preliminary model were handled by the introduction of descriptors based on measurements of Gibbs free energy related quantities, *e.g.* partition coefficients and column retention times, leading to new types of LSERs.

A transition originated in which solvatochromic data was replaced with chromatographic data in the determination of new molecular descriptors. Several descriptor formulations have been proposed, the most notable are from Carr [74, 75], Laffort [76], Rohrschneider [77, 78] and Abraham [26, 27]. Currently, the work of Abraham is accepted as state of the art in the development of molecular

descriptors from chromatographic data. None of the other models is as adaptable as the Abraham method for describing systems different to those they were trained on, and moreover, the other models have not been used as widely as the Abraham solvation parameter model to characterize a broad range of chemical, environmental and biological systems. The Abraham method has also been included in an automatic kinetic mechanism generator with successful results [79], therefore this method will be further elaborated.

2.4.2. Theoretical background of the Abraham model

The work of Abraham et al. [27] continued on the fundamentals introduced by Kamlet and Taft [71]. In a review by Abraham et al. [26], the solvatochromic equation was stated to be a good starting point but suggestions were made to account for special classes of solvents and the cavity formation process. Also, the meaning of XYZ was extended to stand for any solute property in contrast to the original equation in which XYZ was limited to spectroscopic properties. Therefore, the notation for the property, *i.e.* XYZ, was also changed to SP, which stands for solute property as written in equation (2-15). Note that this is not the form in which the equation is generally known today, but in literature the symbols used in this older equation still appear.

$$SP = SP_0 + s\pi^* + d\delta + a\alpha + b\beta + h(\delta_h)^2 \quad (2-19)$$

The physical interpretation of π^* , α and β remains as described earlier. The Hildebrand solubility parameter, *i.e.* δ_h , is used to account for the cavity formation. Moreover, an extra term $d\delta$ was added as an empirical polarizability correction factor for solvents which are highly polarizable, *e.g.* polychlorinated aliphatics.

Abraham proposed two equations, depending on whether the solute is transferred from a gas phase to a solvent or transfer between two different solvents. The equation for gas to solvent transfer and solvent-solvent transfer are given by equation (2-20) and equation (2-21) respectively.

$$SP = c + eE + sS + aA + bB + lL \quad (2-20)$$

$$SP = c + eE + sS + aA + bB + vV \quad (2-21)$$

SP can be any Gibbs free energy-related property of the solute, *e.g.* a partition coefficient. The upper-case parameters refer to the solute descriptors and the lower-case parameters represent the difference in the properties of the two phases between which the solute is being transferred (be it gas-solvent or solvent-solvent). Later the different solvent and solute descriptors will be discussed in more detail. When SP represents the partition coefficient for a molecule between a gas phase and a liquid phase, *i.e.* K_{vs} , it can be used to directly calculate the Gibbs free energy of solvation at standard conditions for the transition of the molecule from the vapor phase to the gas phase, as indicated in equation (2-22).

$$K_{vs} = \exp\left(-\frac{\Delta G_{solv}^{\circ}}{RT}\right) \quad (2-22)$$

The application range of this formula is thus very broad since the solvation energy can be calculated for any solvent and solute for which the descriptors are known. Moreover, the Abraham equation can also be modified to account for the presence of ions as set out below:

$$SP = c + eE + sS + aA + bB + vV + j^{+}J^{+} + j^{-}J^{-} \quad (2-23)$$

where J^{+} is used when the solute is a cation and J^{-} when the solute is an anion, both descriptors cannot be included in equation (2-23) at the same time [80].

The Abraham equation is also generally known as the solvation parameter model and is based on the parameterization of the cavity model of solvation, depicted in Figure 2-10 [80]. In this model, transfer of solute molecules from one phase to another is simplified to a three-step process. Firstly, a cavity is formed in the receiving phase of a suitable size to accommodate the solute molecules with the solvent molecules in the same configuration as in the bulk solvent. This cavity formation requires the disruption of solvent-solvent interactions in the receiving phase and this may or may not be favorable for transfer depending on whether the solvent-solvent interactions in the receiving phase are weaker than those in the donating phase. For transfer from a gas phase, where intermolecular interactions are generally assumed to be negligible (*i.e.* an ideal gas), cavity formation in the liquid cannot be compensated for by collapse of a similar cavity in the gas phase. In this situation, the proportion of the free energy of transfer associated with the cavity formation is always unfavorable for solute transfer. In a second step, reorganization of the solvent molecules around the cavity takes place to minimize the disruption created by the cavity and to form a more favorable orientation for possible solute-solvent interactions. The Gibbs free energy change associated with the latter is not expected to be significant since the enthalpy-entropy compensation and is usually negligible. In a third step, the solute is inserted into the cavity and various solute-solvent interactions appear (the same interactions, if any, are collapsed in the donor phase). For non-ionic solutes, these interactions are identified as dispersion, orientation, induction and hydrogen bonding.

In what follows, the theoretical background of the six different solute and solvent descriptors used in equation (2-24) and equation (2-25) is briefly discussed. The descriptors used to calculate the solvation energy of ionic species are not discussed.

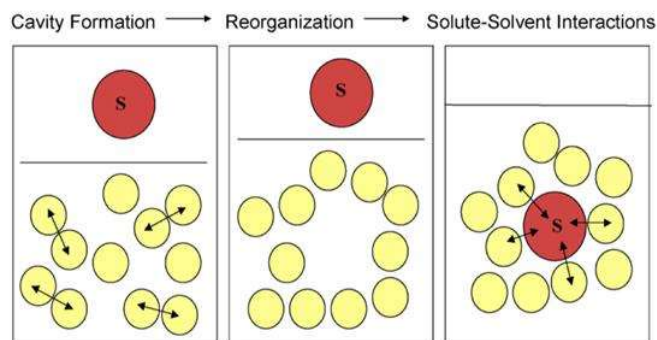


Figure 2-10: Simplified representation of the cavity model of solvation for the transfer of solute S from the gas phase to a liquid with formation of solute-solvent interactions [80].

2.4.2.1 Solute descriptor 'V' and 'L'

It was recognized quite early that the size of the solute molecule had to be included in the LSER formulation. Still, there is some discussion on whether the volume or surface area constitutes the most relevant descriptor, especially for aqueous solutions. A variety of methods are available to characterize the solute size and is also a general issue for modelling of the cavity formation term. Studies suggest that the molecular area rather than the volume may be better correlated for solute partitioning involving water as solvent. This can be explained since molecules with the same volume can have a range of solvent-accessible surface area values by the presence of ring structures, branching or other molecular interactions. Currently most LSER models calculate the cavity term based on the molecular volume, although suggestions are made to allow better modelling of the cavitation energy by addition of an area or shape term [81].

Two different solute descriptors are used to represent the solute size, *i.e.* 'V' and 'L', depending on whether the correlation is for gas-solvent or for solvent-solvent transfers. This **characteristic volume descriptor** is the same for all isomers of a molecule. There is a fundamental difference in interactions when a solute is transferred from the gas phase to the solvent phase compared to a transfer between two solvent phases. Both transfers lead to the disruption of solvent-solvent interactions for the formation of a cavity, but the dispersion interactions contributing to the solvation free energy are different. An isolated solute in the gas phase is not affected by dispersion interactions, by consequence transfer to the solvent phase results in "fresh" dispersion interactions between solute and solvent. In contrast to solvent to solvent transfers, in which the solute already encounters dispersion interactions. Dispersion interactions of a solute are proportional to the size of the molecule, from which follows that the gas-solvent transfer of a solute has an additional dependency on the solute volume. However, it was observed that the solute volume alone is not able to explain the increase in tendency of dispersion interactions for larger molecules. Since it has been impossible to deconvolute the size dependence of the dispersion and the cavity formation contributions they have both taken up in the solute volume descriptor 'L'. This

descriptor is defined as the logarithm of the gas-hexadecane partition coefficient of a solute. The reference for hexadecane as solvent was due to its non-polar and non-hydrogen bonding nature and thus all other contributions cancel out in the LSER equation. The descriptor can be obtained through empirical measurements of the retention volume from gas-liquid chromatography.

2.4.2.2 Solute descriptor 'S' and 'E'

The solute descriptor 'S' is a measure of the solute molecules for electrostatic interactions related to **dipolarity and polarizability effects**. This parameter is similar to the π^* parameter earlier introduced in the solvatochromic equation which was an empirical measure of the solvent dipolarity/polarizability. The two main types of interactions which are related to this descriptor are dipole-dipole interactions and dipole-induced dipole interactions, since descriptors to describe either interaction independently could not be identified. A relation of this parameter to the solute dipole moment resulted in poorer fits, this eventually led to establishment of an empirical solute polarity/dipolarity scale based on gas-chromatography data. The 'S' parameter can be determined along with the hydrogen bonding descriptors using experimental data from water-solvent partition coefficients. The origin of the scale, *i.e.* the zero, is taken for n-alkanes for which the 'S' parameter amounts to zero by definition. By consequence, molecules having a lower tendency to interact via dipole-dipole or dipole-induced dipole interactions than n-alkanes are characterized with a negative value, for example, fluorocarbons and organosilicon compounds.

The descriptor 'E' finds its origin in the δ solvent descriptor from the original solvatochromic approach and takes into account the correction necessary for solvents whose blend of polarizability/dipolarity differed significantly from the one contained in the π^* descriptor. This term is intended to capture the **additional dispersion interactions** between the solute and the solvent beyond what is already accounted for by the cavity and dispersion interaction terms which are respectively included in the 'L' and 'S' solute descriptors. By definition the 'E' descriptor amounts to zero for n-alkanes, and can take both negative and positive values.

2.4.2.3 Solute descriptor 'A' and 'B'

The 'A' and 'B' descriptors model the **hydrogen bonding interactions** between the solute and the neighboring solvent molecules. The hydrogen bonding between solute and solvent molecules can involve the solute as the hydrogen bond donor and the solvent as the acceptor or vice versa. Thereby two categories of hydrogen bonding arise and the difference between them should be noted. The descriptor 'A' is a quantitative measure of the hydrogen bond donor ability, *i.e.* the acidity, of the solute and it is traced back to the α parameter of the solvent as defined by Kamlet and Taft. Similarly, the

solute descriptor ‘B’ originates from the β parameter and is a measure of the hydrogen bond accepting ability, *i.e.* the basicity, of the molecule.

2.4.2.4 Lower-case constants in the Abraham LSER

So far, the interpretation of the solute descriptors has been developed from a physical point of view. However, the lower-case coefficients (c, e, s, a, b, l), *i.e.* the solvent descriptors, that characterize the solvent have not been developed with the same physical background meaning. Although there seem to be a few theoretical developments correlating the solvent descriptors to bulk solvent properties like dielectric constants and refractive indices, they are obtained from an empirical procedure, *i.e.* from regression analysis. Still, it is possible to attach a reasonable amount of qualitative understanding to their values and the trends observed.

An important characteristic of the solvent descriptors is that they are relative, and thus represent the difference in the ability of the two phases, *i.e.* gas-solvent or solvent-solvent, to participate in the various types of interactions. The LSER models reported in literature primarily focus on the gas-solvent transfer and the water-solvent transfer. As such, Abraham parameters have been tabulated for both gas-solvent transfers and water-solvent transfers. In addition, also water-gas transfers have been reported. To check if the parameter sets are perfectly consistent, one would expect that the difference in free energy predicted by Abraham’s LSER would satisfy equation (2-24).

$$\Delta G_{solv}(gas - solvent) = \Delta G_{solv}(gas - water) + \Delta \Delta G_{solv}(water - solvent) \quad (2-24)$$

Which is equivalent with equation (2-25).

$$\begin{aligned} c + eE + sS + aA + bB + lL = & c_{aq} + e_{aq}E + s_{aq}S + a_{aq}A \\ & + b_{aq}B + l_{aq}L + c' + e'E \\ & + s'S + a'A + b'B + v'V \end{aligned} \quad (2-25)$$

If it holds that the solute parameters are uncorrelated, *i.e.* independent of each other, and the water and solvent are perfectly immiscible, the solvent parameters should satisfy several simple identities, *e.g.* equations (2-26).

$$\begin{aligned} c &= c_{aq} + c' \\ e &= e_{aq} + e' \\ s &= s_{aq} + s' \\ a &= a_{aq} + a' \end{aligned} \quad (2-26)$$

For miscible mixtures these equations will not satisfy, since a mixture of water and solvent will form for which the properties cannot be estimated with the Abraham equation. However, usually these equations are not satisfied for solvent parameters reported in literature neither. A part of it can be

explained by the difference in the LL and vV terms which are not equivalent and due to error propagation in the derivation of the solute parameter sets. But it can also indicate that the assumption of uncorrelated solute descriptors is violated, *e.g.* it is observed that solutes with large values for S also tend to have large values for L . Therefore, the review by Green et al. proposes to comprehensively refit the solvent and solute parameters which include the identities as constraints to obtain better results [79].

2.4.3. Determination of Abraham solute parameters

In the previous section the physical interpretation and the structure of the Abraham model has been elaborated. All the different solute-solvent interactions contribute to the Gibbs free energy of solvation, but to use the model in a predictive way, one should be able to determine a priori the solute parameters from the molecular structure or other calculations for the species of interest. Most attempts described in literature are limited to estimation of only one or two parameters. More general methods to estimate all parameters simultaneously by group additivity-based approaches [82, 83] and neural networks [84] have also been reported. For the solvent descriptors the need for these methods is of less importance since for most common solvents these solvent descriptors have been determined experimentally.

2.4.3.1 Solute descriptor 'L' and 'V'

Less research has been reported concerning the estimation methods for the 'L' and 'V' descriptors, which is a measure for the volume of the solute. The descriptor 'V', introduced to include the molecular volume, can be estimated by a solute structure-based additivity scheme to calculate the species characteristic volume, *i.e.* McGowan's volume (V_x), to serve as a measure of the cavity effect [85, 86]. The characteristic volume is then used as a measure for the cavity effect. More recently group additivity methods have been proposed based on multilinear regression to estimate the descriptor 'L' and 'V' of which the Platts method is further elaborated as a method for estimation of all solute parameters at once.

2.4.3.2 Solute descriptor 'S' and 'E'

The solute descriptor 'E', taking into account the correction for the polarizability and dipolarity effects not included in 'S', is the easiest one to estimate as it is not obtained by fitting experimental data. An analytical expression for the descriptor is given by equation (2-27).

$$E = MR_x - 2.83195 V_x + 0.52553 \quad (2-27)$$

With MR_x the solute's molar refraction obtained from refractive index measurements and the quantity V_x is the McGowan's characteristic volume which only depends on the molecular structure of the species. As a result, knowledge from the refractive index and the solute structure allows one to estimate 'E' with little effort. Estimation of the refractive index of a specie is possible with empirical relationships between the Hildebrand solubility parameter δ and the refractive index. The Hildebrand solubility

parameter is on its turn related to the cohesive energy density. Literature values of the refractive index for liquids often vary between different sources, which contributes to the uncertainty in calculated E descriptors.

Several correlations have been researched with more fundamental molecular quantities, *e.g.* usage of feed forward neural networks to correlate ‘S’, describing the electrostatic interactions from dipolarity and polarizability, with a set of 29 structural molecular descriptors by S. Svozil et al. [87]. A similar approach using atomic charges and polarizability from quantum mechanical calculations has been tried to determine the solute descriptor ‘S’ [88]. More recent research has extended this approach by splitting the contribution of polarizability and the electrostatic component. The polarizability is related to the ‘E’ descriptor, which can be derived as discussed from the molar refractive index and the molecular structure, while the electrostatic component is related to the interaction energy between solute and solvent. Finally, the combination of both parts was proposed to obtain the descriptor for polarizability and electrostatic interactions as indicated in equation (2-28).

$$S = \lambda_e E + \lambda_q \Delta U_q \quad (2-28)$$

With ΔU_q the interaction energy between solute and solvent, which can be determined using continuum solvation models and the parameters λ_e and λ_q are obtained by multiple linear regression. An analytical expression was given for the ‘E’ descriptor which allows one to use this formula and calculate ‘S’.

2.4.3.3 Solute descriptors ‘A’ and ‘B’

A lot of research has been dedicated to understanding the molecular origins of the solute parameters introduced in the Abraham equation, especially the ‘A’ and ‘B’ descriptor which are related to hydrogen bond donor and acceptor properties respectively. Attempts have been made to correlate experimental ‘A’ and ‘B’ values with molecular properties, followed by attempts to estimate them using various semi-empirical methods, *e.g.* AM1 [89] and PM3 [90], to model the hydrogen bonding ability of the molecules. However, it was the work of Platts and co-workers who did significant amounts of research to relate ‘A’ and ‘B’ descriptors to fundamental molecular properties. By performing quantum chemical calculations on the isolated molecules and complexes with hydrogen cyanide, it became possible to correlate ‘A’ to the electrostatic potential at the donor hydrogen’s nuclear position. Also, the enthalpy of stabilization from the hydrogen bonding with the reference base is found to be correlated. Analogously, quantum mechanical calculations were performed to correlate the ‘B’ descriptor with atomic and molecular properties, *e.g.* geometry, atomic charges of H-bonded complexes, electric fields and gradients at nuclear positions. This time complexes were made with HF instead of HCN. The response properties having the best correlations with ‘B’ were found to be the H-bond energy, the H-F bond critical point and the electron density at the H-bond critical point. The bond critical point is defined

as the point along the bond path at the interatomic surface, where the shared electron density reaches a minimum.

Still, a less computational method was of interest that could characterize large molecules, *e.g.* for biological research. For research of Cacelli et al. [91], Hartree-Fock calculations were performed to obtain effective nuclear charges which were weighted by ‘atomic factors’ and subsequently correlated with the experimental ‘A’ and ‘B’ descriptors. Also, ‘the atomic factors’ were obtained from regression analysis on a training set with 55 compounds. More recently an estimation scheme has been proposed only using molecular properties of the isolated molecules for ‘A’ and ‘B’. The hydrogen bonding interaction energy is decomposed by the Morokuma method [92] into contributions from electrostatic interactions, polarizability, charge transfer, dispersion, short range exchange repulsion and higher order contributions. The estimation scheme was fitted to a dataset with ‘A’ descriptors of 506 compounds with a single H-bond donor site, which is freely accessible [93]. The model showed good predictive ability which was found to have no significant bias. It is stated qualitative that a similar accuracy as models based on ab initio calculations is achieved although that it allowed faster screening of larger molecules [10].

2.4.4. Experimental determination of descriptors

To apply the Abraham equation, both solute and solvent descriptors are necessary. For the solvent descriptors, the user should rely upon the availability of accurate experimental data, which is the case for most commonly used solvents, in particular for cyclohexane. For the solute descriptors, experimental data is necessary to regress the available methods and accurate experimental data should always be preferred above model estimated data. The most useful methods for determining descriptors experimentally are chromatographic, liquid-liquid partitioning and solubility. Not a single method is suitable to be applied for all compound types, however, to obtain robust descriptor values a combination of methods is mostly the preferred approach. The ‘V’ descriptor is easily calculated for all compounds with a known molecular structure from McGowan’s characteristic volume and ‘E’ can be similarly calculated for all liquids with a known value for the refractive index. The refractive index on its turn can be obtained by performing optical measurements. The ‘E’ descriptor can be estimated for solids in a similar way. Usually a combination of estimation and fitting of experimentally obtained data is used to obtain values which are in accordance. This is done simultaneously with the estimation of the ‘L’, ‘S’, ‘A’ and ‘B’ descriptors, for which experimental approaches are generally required. The estimation methods based on group contribution approaches, *e.g.* Platts fragments, are possible alternative options for the unknown descriptors, but cannot be considered as reliable as experimental methods. In Table 2-3, the solute descriptors are listed for cyclohexanol, cyclohexanone and cyclohexane, and in Table 2-4 the solvent descriptors for cyclohexane are listed. Note that different solvent descriptors are required as input in the Abraham equation for calculating the gas-solvent partition coefficient or the

solvent-solvent partition coefficient corresponding respectively with equation (2-20) and equation (2-21). The descriptors for these compounds are listed up since they are of importance for the further study of the liquid-phase oxidation of cyclohexane.

Table 2-3: Experimental values of Abraham solute descriptors.

Compound	‘E’	‘S’	‘A’	‘B’	‘L’	‘V’
Cyclohexane [94]	0.305	0.100	0	0	2.960	0.845
Cyclohexanol [80]	0.474	0.650	0.237	0.592	3.715	0.904
Cyclohexanone [80]	0.403	0.895	0	0.530	3.759	0.861

Table 2-4: Experimental values of Abraham solvent descriptors for cyclohexane, solvent-solvent descriptors are used for the water-solvent partition coefficient [95].

Applicability	‘c’	‘e’	‘s’	‘a’	‘b’	‘l’	‘v’
Gas-Solvent	0.163	-0.110	0	0	0	1.013	0
Solvent-Solvent	0.159	0.784	-1.678	-3.740	-4.929	0	4.577

2.4.5. Fast estimation methods for solute parameters using group additivity

The methods described until now were related to prediction of the individual descriptors based on fundamental molecular properties. Other methods have been described able to estimate all descriptors together of which Platts method [82, 83] and the neural network approach of Jover et al. [84] are notable and discussed.

2.4.5.1 *Platts fragments approach*

One of the earliest and most used solute descriptor estimation schemes is the fragmental approach developed by research of Platts et al. [82]. A standard least squares regression analysis on a database of 4200 molecules in different solvents was used to develop the contributions belonging to the fragments and a filtration was performed for significance based on a t-test, *i.e.* filtration was performed to only keep fragments which have an essential contribution to the estimation of the descriptors. By introduction of some modifications to account for non-adequately modelled classes of compounds, a set with a total of 81 fragments was obtained, working for all descriptors except the hydrogen bond acidity (the ‘A’ solute descriptor). A new set of 51 fragments was developed to fit the experimental data explicitly for

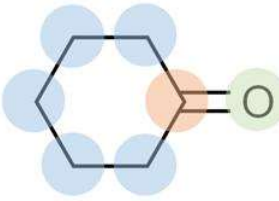
the ‘A’ descriptor. The Platts fragments do contain more complex structural fragments which take into account hydrogen bonding and other interactions between groups in the same molecule. A subset of the “group values” are actually correction factors to consider the presence of specific functional groups, *e.g.* lactones and ortho, meta and para interactions. Note that Platts fragments approach originally uses different symbols for the solute descriptors, but the link and physical meaning of the symbols remains. One drawback of the approach, especially of importance in this thesis, is the lack of data for peroxide linkages, which is currently to be estimated by two ester-functionalities. For an optimal outcome of the liquid-phase model, it can be of interest to regress a value for these peroxide fragments. In Table 2-5, the corresponding fragment contributions of the Platts fragments approach are shown such as implemented in RMG [28], a well-known automatic microkinetic model generator.

Table 2-5: Platts fragments definitions and the corresponding group contributions to Abraham solute descriptors [79].

Fragment	Comments	<i>E</i>	<i>S</i>	<i>B</i>	<i>L</i>	<i>A</i>
-CH ₃	sp ³ , primary	-0.104	-0.075	0.007	0.321	0.000
>CH ₂	sp ³ , secondary	0.000	0.000	0.000	0.499	0.000
>CH-	sp ³ , tertiary	0.089	0.036	0.011	0.449	0.000
>CH _i	sp ³ , quaternary	0.187	0.071	0.037	0.443	0.000
=CH ₂	sp ²	0.045	-0.085	0.019	0.244	0.000
=CH-	sp ² or aromatic	0.068	0.050	0.011	0.469	0.000
=C _i	sp ² or nonfused aromatic	0.180	0.101	0.000	0.624	0.000
C	fused aromatic	-0.104	-0.075	0.007	0.321	0.000
≡C-C	sp	0.040	0.034	0.028	0.332	0.000
≡C-H	sp	0.040	0.034	0.028	0.332	0.082
-OH	aliphatic	0.061	0.247	0.307	0.672	0.345
-OH	phenolic	0.061	0.247	0.307	0.672	0.543
-O-	sp ³ , noncyclic	0.014	0.185	0.211	0.360	0.000
=O	sp ²	-0.041	0.370	0.334	0.495	0.000
-OC(O)-	noncyclic ester	0.067	-0.124	-0.206	0.234	0.000
-OC(O)O-	carbonate	0.000	-0.190	-0.267	0.000	0.000
-C(O)OH	carboxylic acid	-0.012	-0.311	-0.308	0.255	0.243
>C(OH)C(OH) _i	1,2-diol	-0.043	0.052	0.000	0.100	0.000
	intercept	0.248	0.277	0.071	0.130	0.003

In contrast to the group additivity theory, the method of Platts uses a different regression scheme of the fragments. For example, in group additivity theory the carbonyl group is defined as a whole having one contribution for the carbon and oxygen atom, while for Platts fragments the carbonyl group is composed out of an sp- or sp²-hybridized carbon atom and a double bonded oxygen atom. A group is thus defined as an atom itself not taking into account the neighbors. In Table 2-5 only the carbon and oxygen platts fragments are listed, however, the original publication of Platts fragments does contain a much broader range of structural fragments, including halogen interactions, sulfur functionalities, nitrogen functionalities and multiple corrections for special non-neighbor interactions. The latter are of minor importance in this case, as oxidation processes of (oxygenated) hydrocarbons only requires knowledge

from the C, H, O-containing species. Since the aim of this thesis is to introduce the effect of a liquid phase in the in-house developed code, *i.e.* Genesys, this scheme forms an interesting starting point. In Figure 2-11, an example is given for the calculation of the solute descriptors for cyclohexanone. These estimated values can be compared with the tabulated values from experiments denoted in the bottom row (and also listed in Table 2-3. It can be concluded that these estimated values approach the experimental values rather well. If the Abraham equation is used with the tabulated solute descriptors and the Platts fragmental approach solute descriptors, the corresponding values for $\log P_{\text{gas-vapor}}$ is equal to 3.888 and 3.869 respectively and the error amounts to only 0.5 %. Note that by default for every solute descriptor, there exists a fixed value or so-called interception value on which the additional fragmental contributions are added to obtain the resulting descriptor value. These fixed values are equal for all molecules and are indicated in Table 2-5 with “intercept”, in Figure 2-11 this value is also referred to as “intercept”.



Fragment	'E'	'S'	'B'	'A'	'L'
<chem>>CH2</chem>	+ 0	+ 0	+ 0	+ 0	+ 0.499
<chem>>C=</chem>	+ 0.180	+ 0.101	+ 0	+ 0	+ 0.624
<chem>=O</chem>	- 0.041	+ 0.370	+ 0.334	+ 0	+ 0.495
Intercept	+ 0.248	+ 0.277	+ 0.071	+ 0.003	+ 0.130
<chem>C1CCCCC1=O</chem>	0.387	0.748	0.405	0.003	3.744
Tabulated values	0.403	0.895	0.530	0	3.759

Figure 2-11: Application of Platts fragment method for cyclohexanone with RMG fragment values and definitions.

2.4.5.2 Application of the Abraham equation in automatic model generators

The group of Prof. Green has published an article in which they test the accuracy and performance of the combined approach of Platts and Abraham methods for estimation of the free energy of solvation.

Results are obtained by comparing experimental values with values obtained from RMG, the automatic microkinetic model generator, in which the Abraham equation and extensions (Mintz theory and Platts fragments) have been implemented. Solvent descriptors were obtained from a database and the solute descriptors were determined with Platts method. The calculated value for K_{vs} is then converted by equation (2-22) into a value for ΔG_{solv}° . The comparison of estimated values and experimental values are depicted in the parity plot in Figure 2-12. The root mean squared error (RMS error) was found to amount to 1.96 kJ mol⁻¹. However, discrepancies up to 12 kJ mol⁻¹ were detected. The big advantage of this methodology is the low computational effort, only some seconds are required to estimate the Gibbs free energy of solvation, in contrast to several quantum mechanical approaches which require several up to hundreds of CPU-hours.

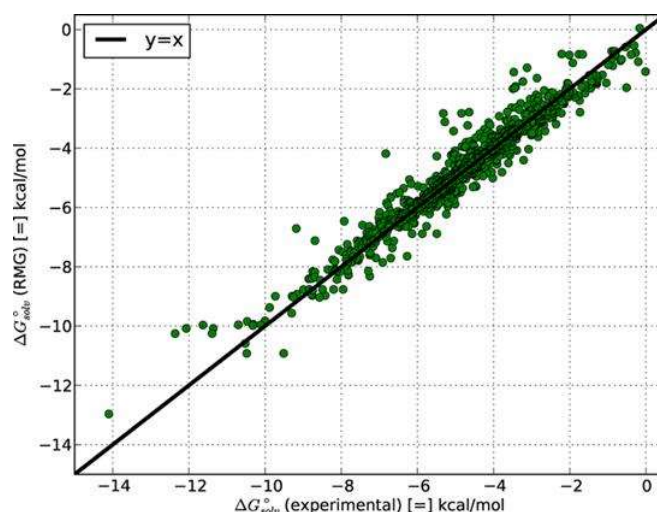


Figure 2-12: Comparison of estimated ΔG_{solv} calculated by RMG and experimental values [79].

In the method of Platts there are no structural fragments for radical compounds, which are necessary for calculation of the reaction rate coefficients. Neither is there a lot of experimental data available, although there have been attempts to determine Abraham descriptors for radical intermediates [79]. The approach adopted in RMG for radical intermediates is to use the descriptors for saturated compounds as a starting point. For carbon-centered radicals for example, the descriptors of the parent species were adopted without any modification. For oxygen-centered radicals, the descriptor 'A' which accounts for the acidity was corrected by subtraction of the contribution of the hydroxyl (-OH) group for the missing hydrogen atom on the radical site. Usage of this methodology for the hydroxyl and the hydroperoxyl radical solvation in water results in values of -19.6 and -31.4 kJ mol⁻¹ respectively, while the experimental values were determined to amount to -16.3 and 28.4 kJ mol⁻¹.

The performance of the approach was also tested on the SAMPL2 blind data set, of which the results are summarized for several other solvation models in Table 2-6 [79, 96]. Due to the presence of halogens and phosphorous compounds in the blind data set, it was required to extend the Abraham/Platts

methodology of RMG. From the results it can be observed that the Platts method is comparable and, in some cases, better than other more sophisticated models. The main outliers for the Abraham/Platts approach belong to the family of halo-uracil compounds with an RMS error close to 12.5 kJ mol⁻¹. Comparison of the data with the MNSOL database is not interesting, since it is expected that the Platts fragments were regressed from these data. As such, this test can only be considered as a kind of reproducibility.

Table 2-6: Performance of predictive solvation models on the SAMPL2 blind data set, All-atom Molecular Dynamics (MD), First-Shell Hydration (FiSH), Conductor like Screening Model for Real Solvents (COSMO-RS), ZAP and the Abraham/Platts approach in RMG [79, 96].

Method	RMS error [kJ mol ⁻¹]
All-atom MD (Mobley)	11.4
FiSH (Purisima)	9.4
COSMO-RS (Klamt)	6.5
ZAP (Nicholls and Ellingson)	9.1
Abraham/Platts (Green)	9.7

2.4.5.3 Machine learning approach

Jover et al. [84] performed research on the same topic by performing both multilinear regression analysis and machine learning techniques, for this case computational neural networks, on a large set of 470 experimental solute descriptors freely available with a broad range of functional groups, *e.g.* alcohols, ethers, halogens, esters, ketones, amines, nitro and amide [97]. Since the V descriptor can be calculated accurately by the McGowan characteristic volume, this descriptor was omitted for regression and thus only models for the four remaining descriptors were to be determined, *i.e.* ‘A’, ‘B’, ‘E’ and ‘S’. For each of these solute descriptors, the objective was to obtain a model with only properties related to the molecular structure. Each newly proposed model to estimate the solute descriptors ended up with 5 ‘molecule’ descriptors, filtered out of a set of 600 ‘molecule’ descriptors, calculated solely from the corresponding molecular structure, by a significance test. More than 5 ‘molecule’ descriptors lead to insignificant contributions and a more complex model. Thus, every solute descriptor is determined by a set of 5 descriptors of the molecular structure, but it is stressed that the descriptors of ‘A’ are not necessarily equal to the ones of descriptor ‘B’. For example, for the ‘S’ parameter the corresponding descriptors are the number of aromatic bonds, the total dipole of the molecule, the number of fluorine atoms, the average valency of the carbon atoms and a non-physical parameter called WNSA-3 related to partial charges.

Computational neural networks lend themselves very well for determination of nonlinear relationships between parameters. Although both machine learning and multi linear regression result in rather good outcomes, the best results were obtained from the neural network approach, as indicated in Figure 2-13 (analogue plots can be found in the work of Jover et al. for the ‘A’, ‘B’ and ‘S’ descriptors). This is also quantitatively proven by the standard deviation, *e.g.* for the ‘E’ descriptor which amounts to 0.52 and 0.10 respectively for the regression model and the neural network model. As drawback, it was observed that the physical meaning that was given to the parameters by Abraham is not completely maintained anymore, *e.g.* donor hydrogen interactions do contribute to the ‘E’ parameter as well as to the ‘S’ parameter. These facts can be explained taking into account the huge complexity of solute/solvent interactions and also the difficulty of isolating closely related physicochemical properties by performing experimental measurements. Thus, although the Abraham parameters reflect mainly the initially proposed physical meaning, the participation of other characteristics and/or capabilities of the solutes cannot be ruled out. The authors have not applied the Abraham equation itself for the determination of the solvation free energy, and thus, direct comparison of the models with experimentally determined solvation free energies is not performed.

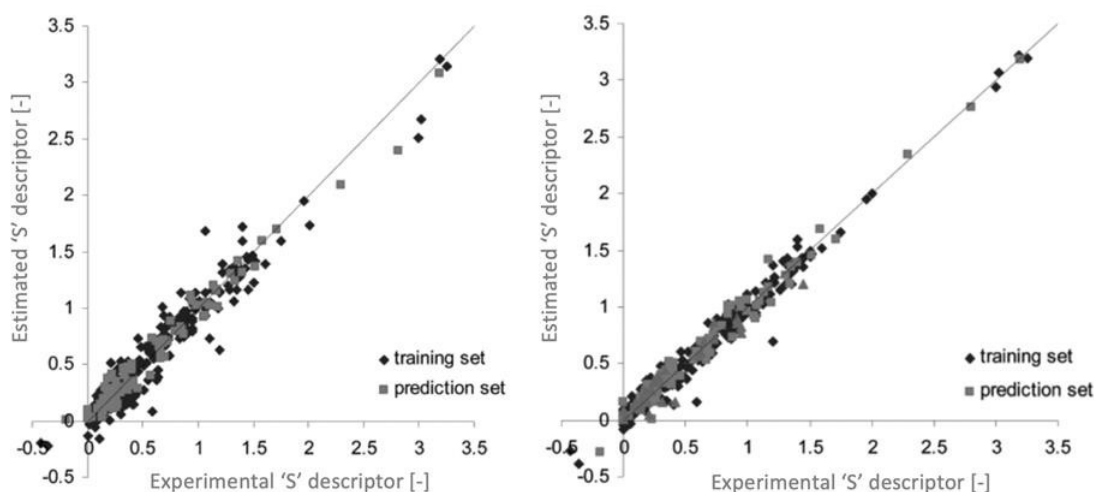


Figure 2-13: Plot of calculated vs experimental ‘E’ solute descriptor for multilinear regression analysis (left) and computational neural networks (right) [84].

Besides the availability of solute descriptors and methods to predict these descriptors, there is also the need for solvent descriptors. For the most important solvents these solvent descriptors have been determined experimentally. Though, similarly models have been developed to estimate solvent descriptors for species for which these descriptors are unavailable. The work of Bradley et al. [98] for example, allows estimation of solvent descriptors for organic solvents, with an open random forest model. Though, the used model omits the ‘c’ descriptor in the Abraham model.

2.4.6. Temperature dependency of the solvation free energy

The Abraham equation discussed so-far is focused on computing the ΔG_{solv} at room temperature for which many experimental data is available for validation. However, for microkinetic models to be useful, they should be valid over a broad range of operating conditions, *e.g.* temperature and pressure, to study the influence of these parameters on conversion and selectivity, and to finally optimize a given process. As such, it is necessary to obtain values for ΔG_{solv} at different temperatures. Modelling the temperature dependence of solvation thermodynamics is still identified as a major area for improvement [21].

The working principle used for the extension of the Abraham equation to other temperatures than for which the parameters are fitted, is based on first order estimation of the temperature dependence. The working equations for this category are given by both equation (2-29) and equation (2-30).

$$\Delta G_{solv}^{\circ}(T_0) = \Delta H_{solv}^{\circ} - T_0 \Delta S_{solv}^{\circ} \quad (2-29)$$

$$\Delta G_{solv}^{\circ}(T) = \Delta H_{solv}^{\circ} - T \Delta S_{solv}^{\circ} \quad (2-30)$$

For which it is assumed that the enthalpy and entropy of solvation are independent of temperature which results in a linear dependence of the free energy of solvation on temperature. The subscript ($^{\circ}$) remains to indicate that the standard state concentrations in both phases are the same at $T = T_0$. The quantities ΔH_{solv}° and ΔS_{solv}° can be obtained by splitting $\Delta G_{solv}^{\circ}(T_0)$ values estimated through empirical or quantum chemical methods. Mintz et al. [99-101] have correlated experimental data of ΔH_{solv}° for gas-solvent transfer for different solvents, with the Abraham parameters of the respective solutes, to determine equations of the form similar to equation (2-31).

$$\Delta H_{solv}^{\circ} = c + eE + sS + aA + bB + lL \quad (2-31)$$

The constants (c, e, s, a, b and l) have different values as the ones used in the Abraham equation, which are to be determined using multiple linear regression analysis. The tabulated solute descriptors can thus be used both for the Abraham and Mintz correlation. Solvent descriptors, to be used in equation (2-31), have been developed for several commonly used solvents: water, benzene, toluene, alkanes, alcohols, acetone, *etc.* [99-104]. This includes data for cyclohexane which is listed in Table 2-7 since this is the solvent of interest used in the liquid-phase oxidation of cyclohexane. Mintz et al. also developed correlations using the V solute descriptor for transfer of solute molecules between different solvents. Remark that for the determination of the solute descriptors, for most cases the temperature of the experiments was limited to the range of 283 to 318 K, in order to avoid errors from temperature dependence of ΔH_{solv}° . In essence this is thus also the range in which the Mintz extension is applicable, which is clearly not ideal for oxidation and pyrolysis processes.

Table 2-7: Mintz solvent descriptors for cyclohexane, solvent-solvent descriptors are used for the water-solvent partition coefficient [103].

Applicability	'c'	'e'	's'	'a'	'b'	'l'	'v'
Gas-Solvent	-6.507	3.375	0	0	0	-9.078	0
Solvent-Solvent	3.046	-8.735	-6.353	-1.264	-2.449	0	-33.550

The standard deviations of the linear fits given by the Mintz approach, in the temperature range between 283 and 318 K, lie in the range of 1.8 to 3.7 kJ mol⁻¹. The standard deviation for cyclohexane amounts in the same temperature range to 1.6 and 2.6 kJ mol⁻¹ for respectively the solvation enthalpy for gas-solvent and solvent-solvent [103]. In Figure 2-14, the corresponding parity plot of the predicted values and the experimental determined values are depicted, from which a rather good fit of the Mintz method is observed [103].

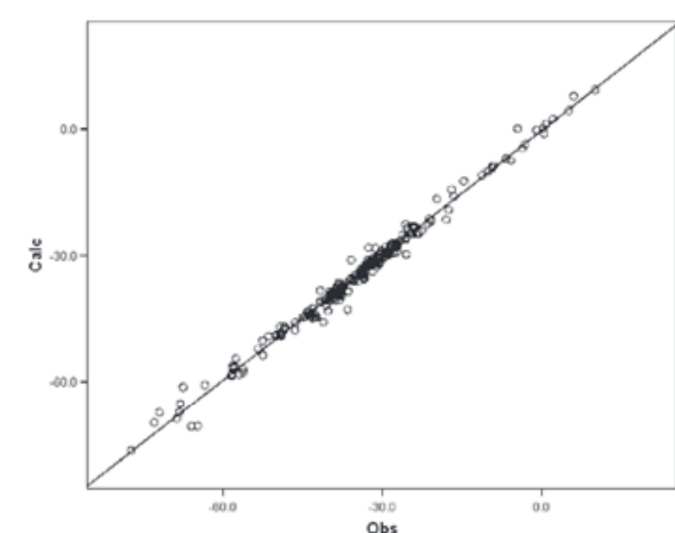


Figure 2-14: A parity plot of the observed and calculated values of the solvation enthalpy, *i.e.* ΔH_{solv} in [kJ mol⁻¹] [103].

Generally, it is stated that the correlations provide a robust and simple method for the estimation of ΔH_{solv}° from the Abraham parameters of the solute under consideration. Combining an estimate of ΔH_{solv}° with the conventional Abraham equation for $\Delta G_{solv}^\circ(T_0)$ allows one to estimate ΔS_{solv}° from equation (2-29). Using this value, it then becomes possible to predict $\Delta G_{solv}^\circ(T)$ using the linear equation (2-30) above.

The limitations of the approach proposed by Mintz et al. are essentially similar to the limitations of the Abraham model, they are extremely dependent on experimental data and hence cannot be applied to any solvent for which experimental data is lacking. Moreover, the Mintz correlations are applicable for only small deviations, approximately in the range 283 to 318 K, but should be used with caution at temperatures beyond the range of the experimental data fitted. This limitation, in general, will apply to all methods that assume the temperature independency of ΔH_{solv}° and ΔS_{solv}° .

Other methods to capture the temperature dependency of the free energy of solvation are out of the scope of this literature survey [64, 105-107], but some of them are briefly mentioned. Next to empirical models, there have also theoretical models been developed. For example, the scaled particle theory (SPT) provides a way to estimate the gas-to-liquid entropy change for a solute. Therefore, the implicit assumption is made that the entropy change related to the cavity formation counts for the bulk of the total entropy change. In the model, the equations of Pierotti [108] are used for the calculation of the Gibbs free energy of solvation, which are also used in many continuum solvation models to estimate the Gibbs free energy of cavitation formation, *i.e.* ΔG_{cav} [109]. Extension to higher order temperature dependence of the solvation free energy can expand the application domain by taking into account the heat capacity of solvation formalism (ΔC_p), *e.g.* equation (2-32) used by Chamberlin et al. in the SM6T model [110]. However, these methods are more complex since they require the availability of the entropy change and the heat capacity of solvation, which is to be determined by *e.g.* ab initio calculations.

$$\Delta G_{solv}^{\circ}(T) = \Delta G_{solv}^{\circ}(T_0) - \Delta S_{solv}^{\circ}(T_0)(T - T_0) + \Delta C_p^{\circ}(T_0) \left((T - T_0) - T \ln \left(\frac{T}{T_0} \right) \right) \quad (2-32)$$

2.5. Validation and accuracy of predicted thermodynamic parameters

Judging the accuracy of the predicted solvation energy for different models is still a difficult issue, mainly due to lack of experimental data. A helpful recent development is the introduction of the so-called blind challenges.

2.5.1. Accuracy tests

The discrete solvent models which use molecular mechanics can estimate free energies of hydrations with an average error of 4.2 kJ mol⁻¹ for small neutral solutes. Cramer and Truhlar introduced several solvation models, *e.g.* SM8 as one of the most universal and accurate continuum solvation methods with a mean absolute deviation (MAD) of only 2.5 kJ mol⁻¹ based on experimental data for 2346 solvation free energies of 940 neutral solutes to train the parameters [10]. Klamt et al. [19] reports a MAD of only 2.0 kJ mol⁻¹ for COSMO-RS for the same set of 940 neutral solutes.

The main conclusion appears to be that several continuum solvation models can yield quite accurate solvation free energies for neutral solutes, if care is taken by the user to the specific details of the parameterization for each of the methods. Predicting solvation energies of ions is more difficult and currently less accurate than neutral species.

The empirical Abraham LSER method is said to be quite accurate when tabulated values of both the solvent and solute descriptors are used. For a given solute with known Abraham descriptors tested in a range of common used solvents, errors in partitioning of about 0.8 to 1.3 kJ mol⁻¹ are typical. If the Abraham parameters for the solute are unknown these can be calculated using the Platts fragmental method [83], which will introduce an additional error. After proposing the fragmental approach, Platts applied it to predict water-solvent partition coefficients, amongst others for cyclohexane, octanol and chloroform, with a typical RMS error of 0.9 to 1.0 log units, corresponding to a $\Delta\Delta G_{solv}$ of 4.2 to 4.8 kJ mol⁻¹. Therefore, the estimated values were validated against an experimental database with partition coefficients, the so-called MedChem97 database [111]. Although, later research indicated that the results are much less accurate with a standard error of 15.1 kJ mol⁻¹ reported by Schüürmann et al. [112].

2.5.2. Blind challenges

The idea of blind challenges was introduced by Nicholls et al. [113] inspired by similar efforts in the protein folding and other communities. It describes an informal blind test in which molecules with available, but unpublished or inaccessible experimental solvation free energy data, *i.e.* ΔG_{solv} , is used

to test the accuracy of several computational solvation models. Making the data set inaccessible is preferred, to reduce the chance that the data is used to train the model parameters. A test set, called SAMPL0 (Statistical Assessment of the Modelling of Proteins and Ligands), was introduced in 2008 with 18 polyfunctional and highly polar molecules containing a wide range of functional groups, since these are important and challenging for solvation models. The successor of the blind challenge, called SAMPL1, proposed originally 63 drug-like molecules, but 7 of them were omitted due to errors with the data preparation. The objective of the blind challenges is to ease comparison of different solvation models and are valuable in helping researchers to identify and correct problems with their developed methods. Since 2008, the SAMPL database has attracted much attention from scientists engaged in the field of kinetic modelling around the world, and has resulted in well over 100 of publications with many of them highly cited.

The SAMPL1 challenge was taken up by different research groups to test the performance of the developed solvation models. Table 2-8 shows the root-mean-square (RMS) error as obtained for the SAMPL1 challenge set for the best version of each listed method. The Abraham method with descriptors as implemented in RMG was tested on its accuracy by Green et al. [10], which is noticeably worse than the other methods based on quantum mechanical or molecular dynamics calculations. Note the difference with the accuracy data tabulated in Table 4, from which results are more recent but less solvation models were tested. In that accuracy test, the Abraham equation did reach accuracies similar to higher computational effort methods. It is a general trend that the tabulated accuracies are really depending on the validation set used. The sulfonyl-urea family of compounds is detected as outliers in the evaluation of almost all participants and the experimental hydration energies have as such been called into question. Removing this type of compounds and 3 extra components which are outliers for some of the other methods as well, results in an RMS error of only 10.9 kJ mol⁻¹ for the Platts/Abraham method. This result would certainly be satisfying for a quick estimation of the solvation energy in an automatic microkinetic model generator. For the crucial reactions, where a more accurate prediction is necessary it is then recommended to perform more computational-expensive calculations.

Although larger and more complex molecules are used in the training sets for solvation models, the species in the SAMPL1 set are still smaller than most drugs and species of interest. Comparisons for drug-like molecules found that the uncertainty for the solvation models is proportional to the solvation energy, so the mean absolute deviation will be considerably higher for large drug-like molecules (and for ionic species) than for small neutral molecules.

More recently the SAMPL5 blind data set has been published which contains the water-cyclohexane distribution coefficient for 53 small, drug-like molecules [114]. This is the first SAMPL set which replaced the hydration solvation free energy which was previously compared by distribution coefficients. It is claimed that the experimental determination of distribution coefficients allows the

inclusion of more special groups in the SAMPL database. The most recent data set is the SAMPL6 which will be developed in two parts will include the comparison of estimated pKa values and includes new testing systems such as octa-acid, tetra-endo-methyl octa-acid, and a series of fragment-like small molecules [115]. This part has been released at the moment of writing, however the second part will contain octanol-water partition coefficients and will be released later this year.

Table 2-8: Performance of various methods on the SAMPL1 challenge set of 56 compounds [10].

Method	RMS error [kJ mol ⁻¹]
COSMO-RS	11.5
Continuum model with Boundary Element Method	15.2
Molecular Dynamics	14.8
Poisson Boltzmann	10.2
SMx	15.1
MST IEF-PCM	10.0
Absolv (Platts/Abraham)	19.0

2.5.3. Experimental data set: Minnesota Solvation database

Truhlar and coworkers [18] did not only develop several solvation models (SMx), they also performed experimental work to generate data to validate and fit parameters for their solvation models. The experimental values of the Gibbs free energy of solvation at 298 K for neutral species containing C, H and O have been published in the Minnesota solvation database, and can be obtained for free for non-commercial applications [65]. Older versions with only a limited amount of data points are also freely available on the internet [116]. The resulting, most recent database set contains 3037 data points spanning 790 solutes and 92 solvents. The solutes in the test set comprise many different types of species, *e.g.* alkanes, alcohols, acids, esters, ethers as well as cyclic and aromatic species. In the database, the Abraham parameters are not listed, but experimental solvation energies as well as the parameters for the SMx are tabulated. When trying new methods for solvation energy estimation, the given database is an interesting starting point to benchmark newly developed solvation models. More specifically for this thesis, the mentioned database can be of interest to perform multi regression and obtain new values for the Platts fragmental contribution approach.

2.5.4. Experimental data set: FreeSolv database

An open-source published experimental data set with Gibbs free energies of solvation is provided by Mobley and co-workers [117], the so-called Free Solvation database or FreeSolv database. In this data

set, only the solvation with water as solvent is considered and therefore only Gibbs free energies of hydration are reported in the data set. The total number of molecules amounts to 643 solutes which are obtained from prior literature publications. The list includes a wide range of hetero-atoms and functional groups, *e.g.* sulfur, nitrogen and halogen containing species, amines, amides, thiols, cyanides, alcohols, ethers, *etc.*

2.6. Liquid-phase effects on kinetics

Prediction of the chemical reaction rates for liquid-phase reactions, particularly in the absence of experimental data, is a very challenging task. Though, common families of reactions can occur in both the liquid phase and in the gas phase; for example, gas-phase combustion of hydrocarbon fuels are dominated by hydrogen abstraction reactions, which also counts for the low temperature, liquid-phase oxidation of fuels by exposure to air. Since this type of reactions also occurs in the gas phase and has been much more studied over the years, it can be advantageous to modify the known gas phase reaction rates accounting for the effect of the solvent rather than predicting the solution phase kinetics directly.

As already discussed earlier, *cf.* Figure 2-1, the presence of a solvent or generally a liquid phase has an influence on the barrier height of the reaction, which is a crucial aspect to account for during kinetic modelling. Moreover, in the liquid phase it is not only these intrinsic kinetics which has to be accounted for, but also the presence of the solvent can modify the reaction rate due to the solvent's physical diffusion limitation. However, these effects will be described by several examples and will not be discussed thoroughly, since the topic of liquid-phase kinetics is less covered in literature and there is a lack of general aspects.

2.6.1. Diffusional limitations

The consequence of the appearance of diffusional phenomena in the liquid phase can be that the intrinsic rate of reaction is not rate-determining since the reaction can be diffusion-limited. An effective reaction rate is then defined, k_{eff} , which is generally written as in equation (2-33).

$$\frac{1}{k_{eff}} = \frac{1}{k_{diff}} + \frac{1}{k_{int}} \quad (2-33)$$

with k_{diff} the diffusion rate and k_{int} the intrinsic reaction rate. For $k_{diff} \gg k_{int}$, the diffusion limitations can be safely neglected and only the intrinsic reaction rate will determine the effective reaction rate. On the other hand, for $k_{diff} \ll k_{int}$, the reaction is completely determined by diffusion and the intrinsic reaction rate is not of importance. For this latter case, it is thus important to accurately describe the rate of diffusion, which is proportional to the sum of the radii R and the sum of the diffusivities D of the reacting species, *cf.* equation (2-34).

$$k_{diff} = 4\pi RD \quad (2-34)$$

The diffusivity of a species A, *i.e.* D_A , can be estimated using the Stokes-Einstein relation, shown in equation (2-35).

$$D_A = \frac{k_b T}{6\pi\eta r_A} \quad (2-35)$$

with k_b the Boltzmann's constant, T the absolute temperature, η is the solvent dynamic viscosity, and r_A is the radius of chemical species A.

The presence of diffusion limitations is a well-known phenomenon when studying heterogeneously catalyzed processes and polymerizations. For example, also the liquid-phase oxidation of cyclohexane suffers from diffusion limitations. Hermans et al. [118] has performed research on this industrial project and discovered that under the experimental conditions (which resemble the industrial conditions) the oxygen addition is diffusion limited. Moreover, the cage-effect also affects several other reactions which lead in the end to by-product formation. More in detail discussion of the results of Hermans and co-workers will be given in the chapter on the liquid-phase microkinetic model.

2.6.2. Experimental determination of reaction rates

Experimental determination of reaction rate for reactions involving radicals in solutions is rather difficult due to the short-lived nature of some radicals; however, some methods have been developed over the last century and are commonly used for measuring the corresponding intrinsic kinetics. For example, the intermittent-illumination method (IIM) [119, 120], also known as the rotating sector method, or laser flash photolysis [121], both these techniques require photochemical initiation of radical chain reactions, with initiation and propagation steps. In the IIM method, a sample is exposed to a constant intensity of light for intermittent periods of time, such that the amount of time spent in light and in the dark remains constant from which the average reaction rate of propagation can be determined. This method is applicable for reactions in the gas phase or solution phase, but is only applicable for specific types of radical chain reactions which can be photochemically initiated. Flash photolysis is a pump-probe technique, in which the reaction environment is firstly excited by a strong pulse of light for a short timespan (nanosecond, picosecond or femtosecond). This pulse then initiates a chemical reaction or chemical reaction chain. By inducing several pulses with a fixed timespan between them, it is possible to determine the reaction rate coefficient of reactions by measuring the concentration of radical species (obtained from spectral absorption analysis). The developer of the laser flash photolysis technique even gained the Nobel prize for chemistry in 1967 for this invention. Nowadays, flash photolysis facilities are mainly applied in the domain of light-induced processes, *e.g.* in organic chemistry, polymers and photosynthesis in plants.

An indirect way of measuring the rate constants for radical-molecule reactions is the so-called radical clock method, which uses a known unimolecular reaction rate and a measured product distribution to determine an unknown radical-molecule reaction rate [122]. As an example, Roschek et al. developed

radical clocks for peroxy radical reactions using the competition between a unimolecular rearrangement of a peroxy radical ($R_1OO\cdot \rightarrow R_2OO\cdot$) and a bimolecular H-atom transfer ($ROO\cdot + AH \rightarrow ROOH + A\cdot$) [123]. This approach is more of interest for the studied process and is generally shown in Figure 2-15. Note that the quasi steady-state assumption (QSSA) is made for the concentration of R_2OO radicals, *i.e.* the reaction rate of R-4 is equal to the reaction rate of R-5.

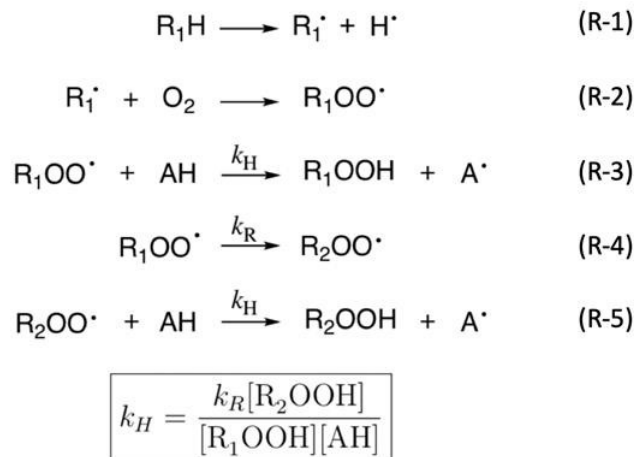


Figure 2-15: Concept of peroxy radical clock for predicting unknown k_H from known k_R [123].

As a drawback, persistency or highly stabilization of $R_1\cdot$ will require a large concentration of substrate to make sure the reactions take place. Furthermore, Jha and Pratt point out that modification of the radical clock method using peroxyesters, would make it possible to study a wider range of reactions [124].

2.6.3. Theoretical determination of reaction rates

An alternative approach for experiments to determine intrinsic kinetic solvent effects is computational chemistry. The reaction rate coefficient, *i.e.* $k(T)$, can be calculated directly using classical transition state theory via equation (2-36).

$$k(T) = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^\ddagger}{RT}\right) \quad (2-36)$$

with h Planck's constant, R is the ideal gas constant and ΔG^\ddagger is the difference in Gibbs free energy between the transition state and reactant. When the reaction is performed in the liquid phase, both reactant and transition state will experience an energetic change due to solvation. This differential solvation between reactants and transition states will change ΔG^\ddagger and thus affect the reaction rate, which was indicated in Figure 2-1. For this, the previously discussed solvation models come in handy.

2.6.4. Example 1: Hydrogen abstraction reactions

Most literature related to the effects of solvents on reaction rates is on bimolecular hydrogen abstraction reactions. Research by Das et al. [7] on the reaction of tert-butoxyl radicals with phenols in different solvents, proved that the rate decreased in polar solvents which is explained by the capability of the phenolic hydroxyl group to form hydrogen bonds with solvent molecules. Valgimigli et al. [6] found that the solvent effect on abstraction of the phenolic hydrogen from α -tocopherol by both tert-butoxy radical and 2,2-diphenyl-1-picrylhydrazyl (DPPH \cdot) does not influence the ratio of both reaction rates in almost every solvent that has been tested, which is depicted in Figure 2-16. This observation is especially surprising, since for the same radicals, the reaction rate in the same solvent differs by over a factor of 10^6 . The deviation in behavior is thought to be linked to diffusion limitations which are not negligible, as known for the reaction with tert-butyl alcohol. Furthermore, it was observed that a solvent with a more basicity character, *i.e.* higher β_2^H value, results in a decreased reaction rate of the hydrogen abstraction.

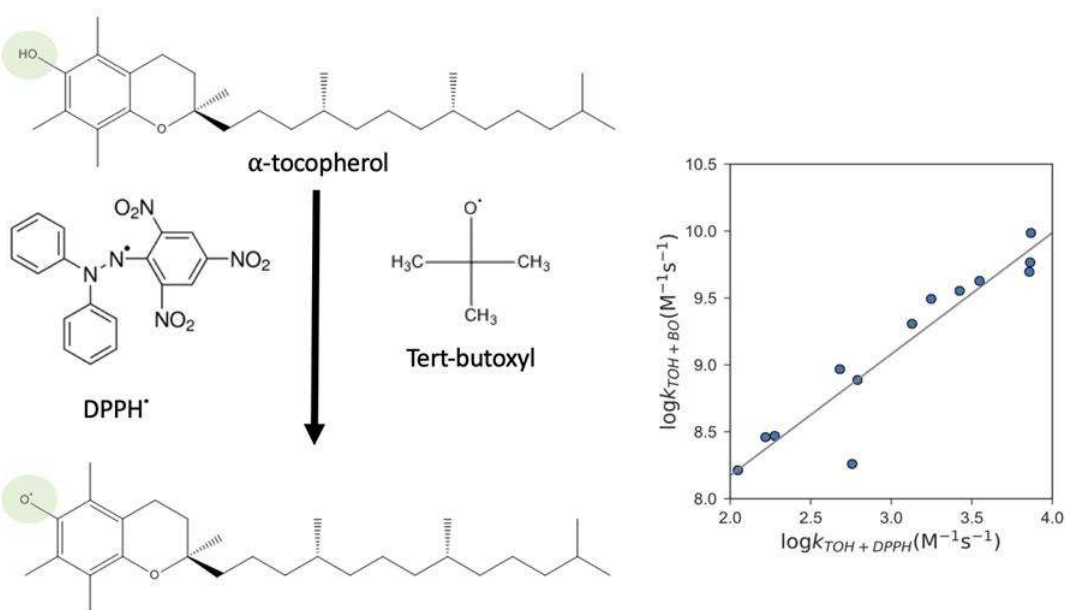
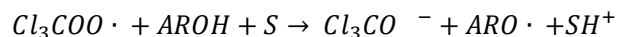
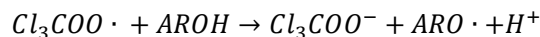


Figure 2-16: The solvent effect on hydrogen abstraction from α -tocopherol (TOH) by radical from 2,2-diphenyl-1-picrylhydrazyl (DPPH, x-axis) or tert-butyl alcohol (BO, y-axis). Each point represents another solvent [6].

In solvents with both high dielectric constant and basicity, a different behavior is observed, which was discovered in a study of the reaction between trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) and $Cl_3COO\cdot$. In these solvents, a mechanistic shift from hydrogen abstraction to electron transfer takes place. This electron transfer mechanism can also be accompanied by a solvent-assisted proton loss, which is also known as a sequential proton loss electron transfer (SPLET):



Thus, the electron transfer mechanism can account for rates which are higher than the rate one might expect by simply correlating rate with one solvent property. This same mechanism of fast electron transfer was also discovered between phenols and DPPH. While hydrogen atom transfer was dominant in nonpolar solvents most of the time, electron transfer still occurred if the radical was strongly oxidizing, as is the case for the Cl_3COO -radical. Reactions proceeding via the electron transfer mechanism should in general be faster in polar solvents. Another mechanism which has been observed in liquid-phase systems is proton-coupled electron transfer (PCET), where an electron and proton are transferred simultaneously but between different sets of orbitals. This mechanism differs from SPLET, as the electron and proton are transferred in a single elementary step rather than sequentially.

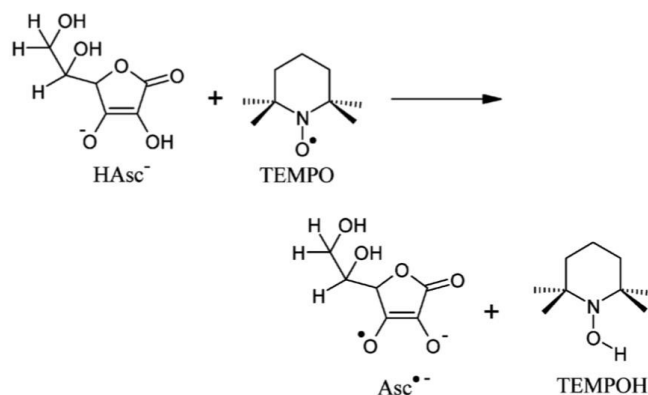


Figure 2-17: Example of the PCET mechanism for the hydrogen abstraction reaction between ascorbate and TEMPO radical [125].

From previous research results, it has generally been understood that solvent effects on hydrogen abstraction reactions are significant when looking at O-H bond abstraction, primarily because of the O-H bond's ability to participate in hydrogen bonding networks, but that the effect on C-H bond abstraction is less pronounced. Thus, the nature of the abstracting species has a large effect on the reaction rate in solvents. Moreover, it is still hypothesized that the reaction of a non-polar hydrogen donor and a non-polar abstracting radical will have little solvent effect.

2.6.5. Example 2: β -scission

Also, the β -scission, *i.e.* the reverse reaction of radical addition to multiple bond, has been studied in a solvent. The β -scission of tert-butoxyl radicals was investigated by Weber and Fischer [126] using spin resonance to measure the corresponding reaction rates. They found that at lower temperature, *i.e.* around

300 K, the rates in solution were at least ten times larger than in the gas phase. This observation is explained by the transition state effect, where the transition state is more polar than the radical and is thus stabilized by interactions with polar and polarizable solvents. Moreover, the reaction of cumyloxyl radicals using laser flash photolysis also showed a rate increase with increasing solvent polarity, which was generally observed for alkyloxyl radicals. The Dimroth-Reichardt parameter (E_T) correlates well with the increase in rate. This parameter is a representation for the charge-transfer absorption of the solvent in pyridinium N-phenolbetaine and serves as a different measure for solvent polarity instead of the commonly used dielectric constant. For example, methyl formamide has an extremely high dielectric constant, but the polarity is similar to methanol which is characterized by its E_T value.

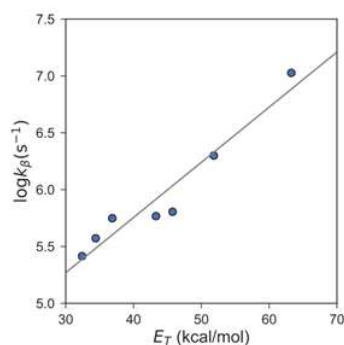


Figure 2-18: β -scission rates correlate with the Dimroth-Reichardt parameter E_T [127].

As general conclusion from the studies it follows that the rates of β -scission reactions increase with some measure of solvent polarity, *e.g.* E_T . However, since only reactions with cumyloxyl and alkyloxyl radicals have been investigated, and mostly performed in water, acetonitrile or mixtures of both, it is perhaps dangerous to generalize the kinetic solvent effects for the entire reaction family.

2.6.6. Example 3: Hydrolysis of formamide

The hydrolysis of formamide has been studied by Almerindo and Pliego [128] with *ab initio* calculations and with the PCM solvation model. In the quantum mechanical calculations, 1 to 4 explicit water molecules were considered, and both a stepwise with tetrahedral intermediate formation and a concerted mechanism was investigated, as indicated in Figure 2-19. Additionally, the level of theory used for the calculations had a large effect on the barriers.

For the stepwise mechanism with only one water molecule, the activation free energy barrier increased by 19.2 kJ mol^{-1} from $211.5 \text{ kJ mol}^{-1}$ in the gas phase to $230.7 \text{ kJ mol}^{-1}$ in the liquid phase. Calculations were performed at MP2 and CCSD(T) levels of theory with basis sets 6-31G(d) and 6-311+G(2df,2p). The stepwise mechanism with two water molecules, *i.e.* the catalyzed case, has in the gas phase an activation free energy of $177.7 \text{ kJ mol}^{-1}$ and in the liquid phase, by taking the solvation into account, an

activation free energy of $223.6 \text{ kJ mol}^{-1}$. However, differently than the first transition state, the level of theory does highly influence the calculated activation energy, *i.e.* differences of 33 kJ mol^{-1} were calculated (with 65.6 kJ mol^{-1} at MP2/6-31G(d) level compared to 99.5 kJ mol^{-1} at CCSD(T)/6-311+G(2df,2p) level). Adding additional water molecules to the transition state resulted in an increase of the activation energy due to the system which becomes entropically unfavorable.

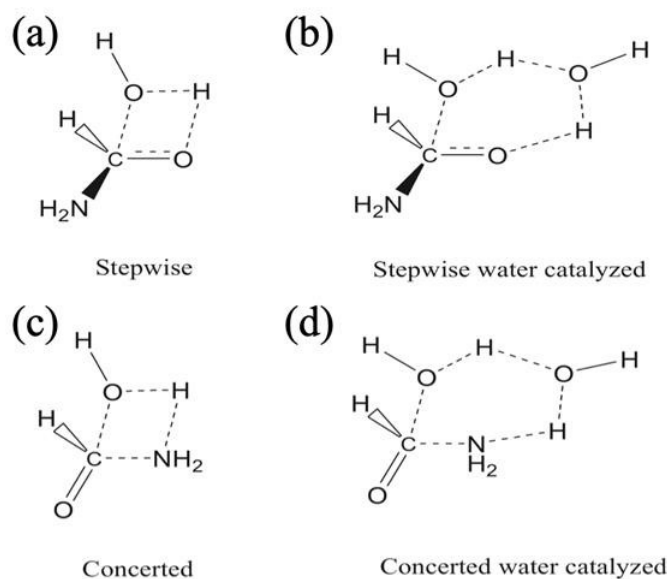


Figure 2-19: Figure 19: Mechanisms for the hydrolysis of formamide. (a) Stepwise, one water molecule; (b) Stepwise, two-water catalyzed; (c) Concerted, one water molecule, (d) Concerted, two-water catalyzed [128].

The study of the concerted case was limited to the case with two water molecules. For the catalyzed reaction, indicated in Figure 19 (d), in the gas phase the activation free energy amounted to $216.9 \text{ kJ mol}^{-1}$ and in the liquid phase it amounted to $227.8 \text{ kJ mol}^{-1}$. The activation energy for the concerted mechanism only increased by 26.8 kJ mol^{-1} , indicating that the transition state is more stabilized by the solvent than in the stepwise mechanism. Also, in this case it was observed that the activation energy highly depends upon the level of theory considered, with values differing 45.6 kJ mol^{-1} (94.9 kJ mol^{-1} at MP2/6-31G(d) level and $140.4 \text{ kJ mol}^{-1}$ at CCSD(T)/6-311+G(2df,2p) level).

2.7. Conclusion

The present literature study has focused on kinetic modelling of processes in the liquid phase. The presence of a liquid phase cannot simply be neglected for modelling purposes. First, the optimal configurational geometry of molecules is highly influenced by the presence of a solvent, especially the polarity of the solvent molecules is of importance. Moreover, several examples have proven that kinetics and thermochemistry can be drastically influenced by the presence of a liquid phase. Differences in reaction rates between gas and liquid phase with a factor thousand are not an exception. The reactants, transition states and products all have a change in Gibbs free energy due to solvation which directly reflects in the corresponding reaction rate coefficients and equilibrium coefficients. An appropriate determination of the Gibbs free energy of solvation is thus crucial. Therefore, several solvation models have been developed to be able to predict the effect of the liquid phase.

Solvation models are classified in empirical methods and *ab initio*, *i.e.* first-principles, based methods. As empirical method the Abraham equation is well suited for potential extension of Genesys to account for the presence of a liquid phase. This linear free-energy relationship makes use of solvent and solute descriptors, which are related to physical properties of the species, *e.g.* the hydrogen bond donor and acceptor capacity. The corresponding descriptors are obtained by experimental data for a huge set of solvents and solutes, which have been regressed with the Abraham model. Temperature dependency of the Gibbs free energy of solvation can be taken into account by an extension of the model which allows prediction of the enthalpy of solvation by Mintz theory. Another extension is the use of Platts fragments for estimation of unknown solute descriptors for species based on the molecular structure, a methodology similar to group additivity theory. Drawback is the lack of data for radicals and also peroxide linkages are missing. This low computational cost method allows a fast estimation of the Gibbs free energy of solvation with a RMS of 10 to 20 kJ mol⁻¹ (depending on the source and validation set used) which is still reasonable for the purpose of automatic microkinetic model generation. However, more computational expensive calculations are required to obtain more accurate results.

A lot of theoretical, so-called *ab initio* or first principle, methods have been developed based on quantum mechanical theory, molecular mechanics theory or a combination of both theories. A first category which is extremely popular are the continuum solvation models, in which the solvent is represented by a continuum with given properties rather than individual molecules. Important solvation models which are well-known in literature are the solvation models (SMx) developed by Truhlar and Cramer and COSMO-RS developed by Klamt. Although that the SMx model is classified under the category of *ab initio* methods, it still required experimental input for regression of several model parameters, therefore the term semi-empirical quantum mechanical solvation model may be more correct. A second group of models are the so-called discrete solvation models in which all molecules are treated separately (both

solvent and solute) after which statistical averaging is applied, with for example Monte Carlo techniques, to obtain physical meaningful properties. Due to the increased molecular detail, these problems are mostly solved at a lower level of theory, *i.e.* solving force fields by using molecular mechanics instead of quantum mechanical theory. Moreover, there are also techniques which combine both quantum mechanical and molecular mechanics techniques, *e.g.* by solving the solute descriptors with high-level quantum mechanical theory and the solvent molecules with molecular mechanics. For the more sensitive reactions, *i.e.* reactions requiring a higher accuracy, it seems most interesting to use the continuum solvation models such as COSMO-RS for which the computational effort remains limited and higher accuracies are reached than the Abraham method. In addition, the laboratory has already gained experience with the COSMOtherm software, which is used for these calculations.

Generally spoken, small and simple species with functionalities adequately represented in training sets, all solvation methods can predict the solvation free energy within about 4.2 kJ mol⁻¹, and nowadays some within 2.1 kJ mol⁻¹. Though, using inappropriate settings or estimating properties for larger molecules, these same methods can give quite poor results. The computational effort however, varies from milliseconds for *e.g.* the Platts and Abraham LSER, to hundreds of CPU-hours for *e.g.* the discrete solvent QM/MM. Typically the more expensive methods will give much more information as outcome than just solvation energies, such as activity coefficients, surface charge densities and optimized geometries. A quick method such as Platts may be a reasonable choice for applications like predictive kinetics where one needs to estimate free energies for hundreds or thousands of small species.

For bigger molecules with multiple functional groups, such as pesticides, the method accuracy decreases to around 12-16 kJ mol⁻¹ and the empirical methods start to fail. Accuracies get even worse for larger molecules. Ions are most difficult to estimate, and the errors are rarely below 16 kJ mol⁻¹. As general remark stated in multiple publications found in the open literature, a plea for more, accurate, experimental data to test the theoretical methods is made. A lot of the data in the discussed blind challenges have estimation uncertainties of about 8 kJ mol⁻¹, which is the same order of magnitude as the errors in the predicted solvation free energies.

When considering processes in a liquid phase, new phenomena appear which can be safely neglected in the gas phase, but which have to be taken into account for kinetic modelling of liquid-phase systems. An important transport phenomenon is the appearance of diffusion limitations *e.g.* by the appearance of a solvent cage. When the reaction is diffusion-limited accurately knowing the intrinsic kinetics is of minor importance and the prediction of diffusion rates becomes more important. Extension of Genesys to account for diffusion limitations will thus also be necessary by *e.g.* the Stokes-Einstein equation, to model the apparent kinetics.

With the given time-frame of the thesis taken into account, the empirical method of Abraham seems the best option to make the transition from a gas-phase model to a liquid-phase model for the oxidation of cyclohexane. Experimental descriptors are available for the main components of interest, *i.e.* as solvent cyclohexane and as solutes cyclohexanol and cyclohexanone, and Platts fragments can be used for unknown species, although that an important functionality, *i.e.* the peroxide linkages, and data for radicals are missing.

2.8. References

1. S.B. Halligudi, B.M. Devassay, A. Ghosh, and V. Ravikumar, *Kinetic study of vapor phase hydrodechlorination of halons by Pd supported catalysts*. Journal of molecular catalysis, 2002. **184**(1): p. 175-181.
2. R. Van de Vijver, N.M. Vandewiele, P.L. Bhoorasingh, B.L. Slakman, F.S. Khanshan, H.H. Carstensen, M.F. Reyniers, G.B. Marin, R.H. West, and K.M. Van Geem, *Automatic Mechanism and Kinetic Model Generation for Gas- and Solution-Phase Processes: A Perspective on Best Practices, Recent Advances, and Future Challenges*. International Journal of Chemical Kinetics, 2015. **47**(4): p. 199-231.
3. M.G. Evans and M. Polanyi, *Further considerations on the thermodynamics of chemical equilibria and reaction rates*. Transactions of the Faraday Society, 1936. **32**(0): p. 1333-1360.
4. P.P. Plehiers, G.B. Marin, C.V. Stevens, and K.M. Van Geem, *Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics*. Journal of Cheminformatics, 2018. **10**(1): p. 11.
5. M. Bietti, M. Salamone, G.A. DiLabio, S. Jockusch, and N.J. Turro, *Kinetic Solvent Effects on Hydrogen Abstraction from Phenol by the Cumyloxyl Radical. Toward an Understanding of the Role of Protic Solvents*. Journal of Organic Chemistry, 2012. **77**(3): p. 1267-1272.
6. L. Valgimigli, J.T. Banks, K.U. Ingold, and J. Lusztyk, *Kinetic Solvent Effects on Hydroxylic Hydrogen Atom Abstractions Are Independent of the Nature of the Abstracting Radical. Two Extreme Tests Using Vitamin E and Phenol*. Journal of the American Chemical Society, 1995. **117**(40): p. 9966-9971.
7. P.K. Das, M.V. Encinas, S. Steenken, and J.C. Scaiano, *Reaction of tert-butoxy radicals with phenols - comparison with the reactions of carbonyl triplets*. Journal of the American Chemical Society, 1981. **103**(14): p. 4162-4166.
8. B.L. Slakman and R.H. West, *Kinetic solvent effects in organic reactions*. Journal of Physical Organic Chemistry, 2019. **32**(3): p. 12.
9. T.F. Redmond and B.B. Wayland, *Dimerization of nitrogen dioxide in solution: a comparison of solution thermodynamics with the gas phase*. The Journal of Physical Chemistry, 1968. **72**(5): p. 1626-1629.
10. A. Jalan, R.W. Ashcraft, R.H. West, and W.H. Green, *Predicting solvation energies for kinetic modeling*. Annual Reports Section "C" (Physical Chemistry), 2010. **106**(0): p. 211-258.
11. C.F. Fischer, *General Hartree-Fock program*. Computer Physics Communications, 1987. **43**(3): p. 355-365.
12. A.D. Becke, *Density-Functional thermochemistry. 3. The role of exact exchange*. Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.
13. L.A. Curtiss, K. Raghavachari, P.C. Redfern, and J.A. Pople, *Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation*. Journal of Chemical Physics, 1997. **106**(3): p. 1063-1079.
14. R.G. Mullen and E.J. Maginn, *Reaction Ensemble Monte Carlo Simulation of Xylene Isomerization in Bulk Phases and under Confinement*. Journal of Chemical Theory and Computation, 2017. **13**(9): p. 4054-4062.
15. O. Acevedo and W.L. Jorgensen, *Quantum and molecular mechanical Monte Carlo techniques for modeling condensed-phase reactions*. Wiley Interdisciplinary Reviews-Computational Molecular Science, 2014. **4**(5): p. 422-435.
16. J.P. Ryckaert and A. Bellemans, *Molecular-dynamics of liquid alkanes*. Faraday Discussions, 1978. **66**: p. 95-106.

17. D.E. Alexander, *Bioaccumulation, bioconcentration, biomagnification*, in *Environmental Geology*. 1999, Springer Netherlands: Dordrecht. p. 43-44, https://doi.org/10.1007/1-4020-4494-1_31.
18. C.J. Cramer and D.G. Truhlar, *A universal approach to solvation modeling*. Accounts of Chemical Research, 2008. **41**(6): p. 760-768.
19. A. Klamt, *Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena*. The Journal of Physical Chemistry, 1995. **99**(7): p. 2224-2235.
20. A. Klamt, *The COSMO and COSMO-RS solvation models*. Wiley Interdisciplinary Reviews-Computational Molecular Science, 2018. **8**(1): p. 11.
21. C.J. Cramer and D.G. Truhlar, *Implicit solvation models: Equilibria, structure, spectra, and dynamics*. Chemical Reviews, 1999. **99**(8): p. 2161-2200.
22. D.S. Abrams and J.M. Prausnitz, *Statistical thermodynamics of liquid-mixtures - New expression for excess Gibbs energy of partly or completely miscible systems*. Aiche Journal, 1975. **21**(1): p. 116-128.
23. H. Renon and J.M. Prausnitz, *Local compositions in thermodynamic excess functions for liquid mixtures*. Aiche Journal, 1968. **14**(1): p. 135-+.
24. A. Fredenslund, R.L. Jones, and J.M. Prausnitz, *Group-contribution estimation of activity-coefficients in nonideal liquid-mixtures*. Aiche Journal, 1975. **21**(6): p. 1086-1099.
25. A. Klamt, F. Eckert, and W. Arlt, *COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures*, in *Annual Review of Chemical and Biomolecular Engineering, Vol 1*, J.M. Prausnitz, M.F. Doherty, and R.A. Segalman, Editors. 2010, Annual Reviews: Palo Alto. p. 101-122
26. M.H. Abraham, A. Ibrahim, and A.M. Zissimos, *Determination of sets of solute descriptors from chromatographic measurements*. Journal of Chromatography A, 2004. **1037**(1-2): p. 29-47.
27. M.H. Abraham, C.F. Poole, and S.K. Poole, *Classification of stationary phases and other materials by gas chromatography*. Journal of Chromatography A, 1999. **842**(1): p. 79-114.
28. C.W. Gao, J.W. Allen, W.H. Green, and R.H. West, *Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms*. Computer Physics Communications, 2016. **203**: p. 212-225.
29. R.H.W. William H. Green, and the RMG Team. *Liquid Phase Systems*. 2019; Available from: <https://reactionmechanismgenerator.github.io/RMG-Py/users/rmg/liquids.html>.
30. M. Orozco and F.J. Luque, *Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems*. Chemical Reviews, 2000. **100**(11): p. 4187-4226.
31. M. Orozco, C. Alhambra, X. Barril, J.M. Lopez, M.A. Busquets, and F.J. Luque, *Theoretical methods for the representation of solvent*. Journal of Molecular Modeling, 1996. **2**(1): p. 1-15.
32. A.M. Ferrenberg and R.H. Swendsen, *New Monte-Carlo technique for studying phase-transitions*. Physical Review Letters, 1988. **61**(23): p. 2635-2638.
33. B.J. Alder and T.E. Wainwright, *Studies in molecular dynamics. I. General method*. Journal of Chemical Physics, 1959. **31**(2): p. 459-466.
34. R. Car and M. Parrinello, *Unified approach for molecular-dynamics and density-functional theory*. Physical Review Letters, 1985. **55**(22): p. 2471-2474.
35. P.C. Aeberhard, J.S. Arey, I.C. Lin, and U. Rothlisberger, *Accurate DFT Descriptions for Weak Interactions of Molecules Containing Sulfur*. Journal of Chemical Theory and Computation, 2009. **5**(1): p. 23-28.

-
36. E.R. Johnson, I.D. Mackie, and G.A. DiLabio, *Dispersion interactions in density-functional theory*. Journal of Physical Organic Chemistry, 2009. **22**(12): p. 1127-1135.
37. W.L. Jorgensen and L.L. Thomas, *Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria*. Journal of chemical theory and computation, 2008. **4**(6): p. 869-876.
38. J. Kastner, H.M. Senn, S. Thiel, N. Otte, and W. Thiel, *QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction*. Journal of Chemical Theory and Computation, 2006. **2**(2): p. 452-461.
39. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, *GROMACS: Fast, flexible, and free*. Journal of Computational Chemistry, 2005. **26**(16): p. 1701-1718.
40. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM - A program for macromolecular energy, minimization, and dynamics calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
41. M.S. Gordon, M.A. Freitag, P. Bandyopadhyay, J.H. Jensen, V. Kairys, and W.J. Stevens, *The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry*. Journal of Physical Chemistry A, 2001. **105**(2): p. 293-307.
42. G.B. Desmet, N. De Rybel, P.H.M. Van Steenberge, D.R. D'hooge, M.-F. Reyniers, and G.B. Marin, *Ab-Initio-Based Kinetic Modeling to Understand RAFT Exchange: The Case of 2-Cyano-2-Propyl Dodecyl Trithiocarbonate and Styrene*. Macromolecular Rapid Communications, 2018. **39**(2): p. 1700403.
43. G.B. Desmet, M.K. Sabbe, D.R. D'Hooge, P. Espeel, S. Celasun, G.B. Marin, F.E. Du Prez, and M.-F. Reyniers, *Thiol-Michael addition in polar aprotic solvents: nucleophilic initiation or base catalysis?* Polymer Chemistry, 2017. **8**(8): p. 1341-1352.
44. G.B. Desmet, D.R. D'hooge, M.K. Sabbe, M.-F. Reyniers, and G.B. Marin, *Computational Investigation of the Aminolysis of RAFT Macromolecules*. The Journal of Organic Chemistry, 2016. **81**(23): p. 11626-11634.
45. G.B. Desmet, D.R. D'hooge, M.K. Sabbe, G.B. Marin, F.E. Du Prez, P. Espeel, and M.-F. Reyniers, *Computational Study and Kinetic Analysis of the Aminolysis of Thiolactones*. The Journal of Organic Chemistry, 2015. **80**(17): p. 8520-8529.
46. G.W.T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016., *Gaussian 09*.
47. Gaussian. *SCRF*. Available from: <http://gaussian.com/scrf/>.
48. S. Malanowski, R. Patz, M.T. Ratzsch, and C. Wohlfarth, *Dielectric, electro-optical and thermodynamic investigation of binary-mixtures formed by puridine with hydrocarbons and chlorobenzene*. Fluid Phase Equilibria, 1979. **3**(4): p. 291-312.
49. T.N. Truong and E.V. Stefanovich, *A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity*. Chemical Physics Letters, 1995. **240**(4): p. 253-260.
-

50. J.L. Pascualahir, E. Silla, and I. Tunon, *GEPOL - AN IMPROVED DESCRIPTION OF MOLECULAR-SURFACES .3. A NEW ALGORITHM FOR THE COMPUTATION OF A SOLVENT-EXCLUDING SURFACE*. Journal of Computational Chemistry, 1994. **15**(10): p. 1127-1138.
51. 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, 1981. **55**(1): p. 117-129.
52. J. Foresman, *Exploring Chemistry with Electronic Structure Methods*. Vol. 10. 1996.
53. J.B. Foresman, T.A. Keith, K.B. Wiberg, J. Snoonian, and M.J. Frisch, *Solvent Effects. 5. Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation on ab Initio Reaction Field Calculations*. The Journal of Physical Chemistry, 1996. **100**(40): p. 16098-16104.
54. A. Klamt, *From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. 2005. 246.
55. COSMOlogic GmbH & Co. KG. *COSMOtherm: Predicting Solutions since 1999*. Available from: <http://www.cosmologic.de/products/cosmotherm.html>.
56. A. Hellweg and F. Eckert, *Brick by Brick Computation of the Gibbs Free Energy of Reaction in Solution Using Quantum Chemistry and COSMO-RS*. Aiche Journal, 2017. **63**(9): p. 3944-3954.
57. 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, 1991. **113**(22): p. 8305-8311.
58. C.J. Cramer and D.G. Truhlar, *PM3-SM3: A general parameterization for including aqueous solvation effects in the PM3 molecular orbital model*. Journal of Computational Chemistry, 1992. **13**(9): p. 1089-1097.
59. C.P. Kelly, C.J. Cramer, and D.G. Truhlar, *SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters*. Journal of Chemical Theory and Computation, 2005. **1**(6): p. 1133-1152.
60. A.V. Marenich, C.J. Cramer, and D.G. Truhlar, *Generalized Born Solvation Model SM12*. Journal of Chemical Theory and Computation, 2013. **9**(1): p. 609-620.
61. A.V. 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 the Bulk Dielectric Constant and Atomic Surface Tensions*. The Journal of Physical Chemistry B, 2009. **113**(18): p. 6378-6396.
62. J. Liu, C.P. Kelly, A.C. Goren, A.V. Marenich, C.J. Cramer, D.G. Truhlar, and C.-G. Zhan, *Free Energies of Solvation with Surface, Volume, and Local Electrostatic Effects and Atomic Surface Tensions to Represent the First Solvation Shell*. Journal of Chemical Theory and Computation, 2010. **6**(4): p. 1109-1117.
63. A.V. Marenich, R.M. Olson, C.P. Kelly, C.J. Cramer, and D.G. Truhlar, *Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges*. Journal of Chemical Theory and Computation, 2007. **3**(6): p. 2011-2033.
64. A.C. Chamberlin, C.J. Cramer, and D.G. Truhlar, *Predicting Aqueous Free Energies of Solvation as Functions of Temperature*. The Journal of Physical Chemistry B, 2006. **110**(11): p. 5665-5675.
65. A.V.K. Marenich, C. P.; Thompson, J. D.; Hawkins, G. D.; and C.C.G. Chambers, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., *Minnesota Solvation Database – version 2012*, M. University of Minnesota, Editor. 2012.

-
66. D.J. Giesen, C.C. Chambers, C.J. Cramer, and D.G. Truhlar, *Solvation Model for Chloroform Based on Class IV Atomic Charges*. The Journal of Physical Chemistry B, 1997. **101**(11): p. 2061-2069.
67. A.G. Loudon, *Advances in linear free energy relationships edited by N. B. Chapman and J. Shorter*, Plenum Press, London, 1972. pp. XIV + 486. £10.00, \$28.00. Organic Mass Spectrometry, 1973. **7**(3): p. 365-365.
68. L.P. Hammett, *The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives*. Journal of the American Chemical Society, 1937. **59**(1): p. 96-103.
69. R.W. Taft, *Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters*. Journal of the American Chemical Society, 1952. **74**(11): p. 2729-2732.
70. J.N. Brönsted, *Einige Bemerkungen über den Begriff der Säuren und Basen*. Recueil des Travaux Chimiques des Pays-Bas, 1923. **42**(8): p. 718-728.
71. M.J. Kamlet, J.L. Abboud, and R.W. Taft, *The solvatochromic comparison method. 6. The .pi.* scale of solvent polarities*. Journal of the American Chemical Society, 1977. **99**(18): p. 6027-6038.
72. T. Yokoyama, R.W. Taft, and M.J. Kamlet, *The solvatochromic comparison method. 3. Hydrogen bonding by some 2-nitroaniline derivatives*. Journal of the American Chemical Society, 1976. **98**(11): p. 3233-3237.
73. M.J. Kamlet and R.W. Taft, *The solvatochromic comparison method. I. The .beta.-scale of solvent hydrogen-bond acceptor (HBA) basicities*. Journal of the American Chemical Society, 1976. **98**(2): p. 377-383.
74. J. Li, Y. Zhang, A.J. Dallas, and P.W. Carr, *Measurement of solute dipolarity/polarizability and hydrogen bond acidity by inverse gas chromatography*. Journal of Chromatography A, 1991. **550**: p. 101-134.
75. J. Li, Y. Zhang, H. Ouyang, and P.W. Carr, *Gas chromatographic study of solute hydrogen bond basicity*. Journal of the American Chemical Society, 1992. **114**(25): p. 9813-9828.
76. P. Laffort and F. Patte, *Solubility factors in gas-liquid chromatography: Comparison between two approaches and application to some biological studies*. Journal of Chromatography A, 1976. **126**: p. 625-639.
77. L. Rohrschneider, *Eine methode zur chrakterisierung von gaschromatographischen trennflüssigkeiten*. Journal of Chromatography A, 1966. **22**: p. 6-22.
78. L. Rohrschneider, *Chromatographic Characterization of Liquid Phases and Solutes for Column Selection and Identification*. Journal of Chromatographic Science, 1973. **11**(3): p. 160-166.
79. A. Jalan, R.H. West, and W.H. Green, *An Extensible Framework for Capturing Solvent Effects in Computer Generated Kinetic Models*. Journal of Physical Chemistry B, 2013. **117**(10): p. 2955-2970.
80. C.F. Poole, S.N. Atapattu, S.K. Poole, and A.K. Bell, *Determination of solute descriptors by chromatographic methods*. Analytica Chimica Acta, 2009. **652**(1): p. 32-53.
81. M. Vitha and P.W. Carr, *The chemical interpretation and practice of linear solvation energy relationships in chromatography*. Journal of Chromatography A, 2006. **1126**(1): p. 143-194.
82. J.A. Platts, D. Butina, M.H. Abraham, and A. Hersey, *Estimation of molecular linear free energy relation descriptors using a group contribution approach*. Journal of Chemical Information and Computer Sciences, 1999. **39**(5): p. 835-845.
83. J.A. Platts, M.H. Abraham, D. Butina, and A. Hersey, *Estimation of Molecular Linear Free Energy Relationship Descriptors by a Group Contribution Approach. 2. Prediction of Partition Coefficients*. Journal of Chemical Information and Computer Sciences, 2000. **40**(1): p. 71-80.
-

84. J. Jover, R. Bosque, and J. Sales, *Determination of Abraham Solute Parameters from Molecular Structure*. Journal of Chemical Information and Computer Sciences, 2004. **44**(3): p. 1098-1106.
85. M.H. Abraham, *Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes*. Chemical Society Reviews, 1993. **22**(2): p. 73-83.
86. M.H. Abraham and J.C. McGowan, *The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography*. Chromatographia, 1987. **23**(4): p. 243-246.
87. D. Svozil, J.G.K. Ševčík, and V. Kvasnička, *Neural Network Prediction of the Solvatochromic Polarity/Polarizability Parameter*. Journal of Chemical Information and Computer Sciences, 1997. **37**(2): p. 338-342.
88. O. Lamarche, J.A. Platts, and A. Hersey, *Theoretical prediction of the polarity/polarizability parameter π_2H* . Physical Chemistry Chemical Physics, 2001. **3**(14): p. 2747-2753.
89. M.J.S. Dewar, E.G. Zebisch, E.F. Healy, and J.J.P. Stewart, *The development and use of quantum-mechanical molecular models. 76. AM1 - A new general purpose quantum-mechanical molecular model*. Journal of the American Chemical Society, 1985. **107**(13): p. 3902-3909.
90. J.J.P. Stewart, *Optimization of parameters for semiempirical methods. 1. Method*. Journal of Computational Chemistry, 1989. **10**(2): p. 209-220.
91. I. Cacelli, S. Campanile, A. Giolitti, and D. Molin, *Theoretical Prediction of the Abraham Hydrogen Bond Acidity and Basicity Factors from a Reaction Field Method*. Journal of Chemical Information and Modeling, 2005. **45**(2): p. 327-333.
92. K. Morokuma, *Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity*. Accounts of Chemical Research, 1977. **10**(8): p. 294-300.
93. J. Schwöbel, R.-U. Ebert, R. Kühne, and G. Schüürmann, *Modeling the H bond donor strength of $\square OH$, $\square NH$, and $\square CH$ sites by local molecular parameters*. Journal of Computational Chemistry, 2009. **30**(9): p. 1454-1464.
94. A.M. Zissimos, M.H. Abraham, M.C. Barker, K.J. Box, and K.Y. Tam, *Calculation of Abraham descriptors from solvent-water partition coefficients in four different systems; evaluation of different methods of calculation*. Journal of the Chemical Society, Perkin Transactions 2, 2002(3): p. 470-477.
95. M.H. Abraham and W.E. Acree, *Descriptors for the Prediction of Partition Coefficients and Solubilities of Organophosphorus Compounds*. Separation Science and Technology, 2013. **48**(6): p. 884-897.
96. A. Nicholls, S. Wlodek, and J.A. Grant, *SAMPL2 and continuum modeling*. Journal of Computer-Aided Molecular Design, 2010. **24**(4): p. 293-306.
97. A.M. Zissimos, M.H. Abraham, A. Klamt, F. Eckert, and J. Wood, *A Comparison between the Two General Sets of Linear Free Energy Descriptors of Abraham and Klamt*. Journal of Chemical Information and Computer Sciences, 2002. **42**(6): p. 1320-1331.
98. J.-C. Bradley, M. H. Abraham, W. Acree, and A. Lang, *Predicting Abraham model solvent coefficients*. Vol. 9. 2015. 12.
99. C. Mintz, J. Gibbs, W.E. Acree, and M.H. Abraham, *Enthalpy of solvation correlations for organic solutes and gases dissolved in acetonitrile and acetone*. Thermochemica Acta, 2009. **484**(1): p. 65-69.
100. C. Mintz, W.E. Acree Jr., and M.H. Abraham, *Correlation of Minimum Inhibitory Concentrations Toward Oral Bacterial Growth Based on the Abraham Model*. QSAR & Combinatorial Science, 2006. **25**(10): p. 912-920.

101. C. Mintz, W.E. Acree, and M.H. Abraham, *Comments on "Solvation Parameters. 2. A Simplified Molecular Topology To Generate Easily Optimized Values"*. Journal of Chemical Information and Modeling, 2006. **46**(5): p. 1879-1881.
102. C. Mintz, K. Burton, W.E. Acree, and M.H. Abraham, *Enthalpy of solvation correlations for gaseous solutes dissolved in dimethyl sulfoxide and propylene carbonate based on the Abraham model*. Thermochimica Acta, 2007. **459**(1): p. 17-25.
103. C. Mintz, M. Clark, K. Burton, W.E. Acree Jr., and M.H. Abraham, *Enthalpy of Solvation Correlations for Gaseous Solutes Dissolved in Benzene and in Alkane Solvents Based on the Abraham Model*. QSAR & Combinatorial Science, 2007. **26**(8): p. 881-888.
104. C. Mintz, M. Clark, K. Burton, W.E. Acree, and M.H. Abraham, *Enthalpy of Solvation Correlations for Gaseous Solutes Dissolved in Toluene and Carbon Tetrachloride Based on the Abraham Model*. Journal of Solution Chemistry, 2007. **36**(8): p. 947-966.
105. Z.A. Fekete, K. Héberger, Z. Király, and M. Görgényi, *Temperature dependence of solvation heat capacities by gas chromatography*. Analytica Chimica Acta, 2005. **549**(1): p. 134-139.
106. F.R. González, *Considerations on the temperature dependence of the gas-liquid chromatographic retention*. Journal of Chromatography A, 2002. **942**(1): p. 211-221.
107. S. Höfinger and F. Zerbetto, *Introducing temperature dependence in an enhanced Poisson-Boltzmann approach*. Chemical Physics Letters, 2009. **480**(4): p. 313-317.
108. R.A. Pierotti, *A scaled particle theory of aqueous and nonaqueous solutions*. Chemical Reviews, 1976. **76**(6): p. 717-726.
109. J. Tomasi, B. Mennucci, and R. Cammi, *Quantum Mechanical Continuum Solvation Models*. Chemical Reviews, 2005. **105**(8): p. 2999-3094.
110. A. C. Chamberlin, C. Cramer, and D. Truhlar, *SM6T: A model to predict aqueous free energies of solvation as a function of temperature*. 2019.
111. G.D.H. C J Cramer, G C Lynch, D J Giesen, I Rossi, J W Storer, D G Truhlar, D A Liotard, Bull, *Masterfile (1994) from MedChem Software*. 1995, BioByte Corp., P.O. 517.
112. G. Schüürmann, R.-U. Ebert, and R. Kühne, *Prediction of Physicochemical Properties of Organic Compounds from 2D Molecular Structure; Fragment Methods vs. LFER Models*. CHIMIA International Journal for Chemistry, 2006. **60**(10): p. 691-698.
113. A. Nicholls, S. Wlodek, and J.A. Grant, *The SAMPL1 Solvation Challenge: Further Lessons Regarding the Pitfalls of Parametrization*. The Journal of Physical Chemistry B, 2009. **113**(14): p. 4521-4532.
114. A.S. Rustenburg, J. Dancer, B. Lin, J.A. Feng, D.F. Ortwine, D.L. Mobley, and J.D. Chodera, *Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge*. Journal of Computer-Aided Molecular Design, 2016. **30**(11): p. 945-958.
115. Drug Design Data Resource. *SAMPL6*. 2019; Available from: <https://drugdesigndata.org/about/sampl6>.
116. D.M.D. Paul Winget, David J. Giesen, Christopher J. Cramer, Donald G. Truhlar. *Minnesota Solvent Descriptor Database*. 1999; Available from: <https://comp.chem.umn.edu/solvation/mnsddb.pdf>.
117. D.L. Mobley and J.P. Guthrie, *FreeSolv: a database of experimental and calculated hydration free energies, with input files*. Journal of Computer-Aided Molecular Design, 2014. **28**(7): p. 711-720.
118. I. Hermans, P. Jacobs, and J. Peeters, *The Formation of Byproducts in the Autoxidation of Cyclohexane*. Chemistry – A European Journal, 2007. **13**(3): p. 754-761.

119. F. Briers, D.L. Chapman, and E. Walters, *LXXVII.—The influence of the intensity of illumination on the velocity of photochemical changes. The determination of the mean life of a hypothetical catalyst*. Journal of the Chemical Society (Resumed), 1926. **129**(0): p. 562-569.
120. F. Briers and D.L. Chapman, *CCXXXI.—The influence of the intensity of illumination on the velocity of the photochemical union of bromine and hydrogen, and the determination of the mean life of a postulated catalyst*. Journal of the Chemical Society (Resumed), 1928(0): p. 1802-1811.
121. G. Porter, *Flash Photolysis and Some of Its Applications*. Science, 1968. **160**(3834): p. 1299-1307.
122. A. Shepp, *Rate of Recombination of Radicals. I. A General Sector Theory; A Correction to the Methyl Radical Recombination Rate*. The Journal of Chemical Physics, 1956. **24**(5): p. 939-943.
123. D.H. Geske and A.H. Maki, *Electrochemical Generation of Free Radicals and Their Study by Electron Spin Resonance Spectroscopy; the Nitrobenzene Anion Radical*. Journal of the American Chemical Society, 1960. **82**(11): p. 2671-2676.
124. B. Roschek, K.A. Tallman, C.L. Rector, J.G. Gillmore, D.A. Pratt, C. Punta, and N.A. Porter, *Peroxyl Radical Clocks*. The Journal of Organic Chemistry, 2006. **71**(9): p. 3527-3532.
125. I. Sajenko, V. Pilepić, C. Jakobušić Brala, and S. Uršić, *Solvent Dependence of the Kinetic Isotope Effect in the Reaction of Ascorbate with the 2,2,6,6-Tetramethylpiperidine-1-oxyl Radical: Tunnelling in a Small Molecule Reaction*. The Journal of Physical Chemistry A, 2010. **114**(10): p. 3423-3430.
126. M. Weber and H. Fischer, *Absolute Rate Constants for the β -Scission and Hydrogen Abstraction Reactions of the tert-Butoxyl Radical and for Several Radical Rearrangements: Evaluating Delayed Radical Formations by Time-Resolved Electron Spin Resonance*. Journal of the American Chemical Society, 1999. **121**(32): p. 7381-7388.
127. D.V. Avila, C.E. Brown, K.U. Ingold, and J. Luszyk, *Solvent effects on the competitive .beta.-scission and hydrogen atom abstraction reactions of the cumyloxyl radical. Resolution of a long-standing problem*. Journal of the American Chemical Society, 1993. **115**(2): p. 466-470.
128. G.I. Almetindo and J.R. Pliego, *Ab initio investigation of the kinetics and mechanism of the neutral hydrolysis of formamide in aqueous solution*. Journal of the Brazilian Chemical Society, 2007. **18**(4): p. 696-702.

3.

Methodology - Genesys and ab initio calculations

3 . METHODOLOGY - GENESYS AND AB INITIO CALCULATIONS	79
3.1. GENESYS: AUTOMATIC MICROKINETIC MODEL GENERATOR.....	80
3.1.1. User-defined input	80
3.1.2. Termination criteria	83
3.1.3. Reaction generation	84
3.1.4. Assigning thermodynamic properties	85
3.1.4.1 Group additivity theory.....	86
3.1.4.2 Group additivity algorithm	87
3.1.5. Assigning kinetic parameters	89
3.1.6. Output and post-processing.....	90
3.1.6.1 CHEMKIN.....	90
3.1.6.2 The AramcoMech base mechanism	91
3.2. AB INITIO CALCULATIONS	93
3.3. CONCLUSIONS	97
3.4. REFERENCES.....	98

3.1. Genesys: automatic microkinetic model generator

Genesys is an automatic microkinetic model builder of which the development has started in 2012 in the Laboratory for Chemical Technology (LCT) at Ghent University. The code is developed to generate microkinetic models for important gas-phase free radical processes, covering the important pyrolysis and oxidation processes. Genesys is written in the programming language JAVA and is integrated with the open-source chemo-informatics library “Chemical Development Kit” (CDK) [1-4] for internal representation of the species by a weighted graph. The latter representation allows a general usage of Genesys regardless of the involved chemical elements and chemistry. Moreover, with CDK come a lot of different algorithms used in Genesys, *e.g.* the conversion of SMILES [5] and InChI [6] identifiers to the internal representation, graph isomorphism, subgraph recognition with SMARTS, *etc.*

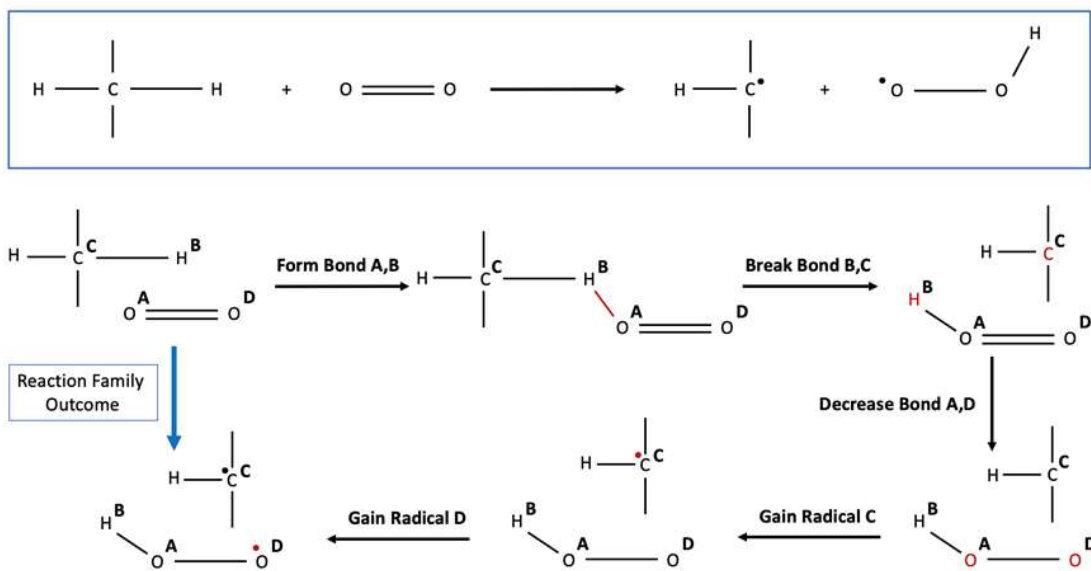
3.1.1. User-defined input

Before generating a microkinetic model, the initial pool of reactants must be specified in an input file in either InChI or SMILES notation. Furthermore, a second file is required which defines all the reaction families which are to be considered in the reaction network. Many kinetic model generators rely on the concept of reaction families [7]. Each reaction family, either unimolecular or in the presence of another species, represents a particular type of elementary chemical reaction. Besides defining the structural requirements of the reactants and the reactive moiety, the reaction family also specifies the atom rearrangements occurring, by elementary transformations, when proceeding from the reactants to the products. To automatically generate the complete reaction network, a species must be checked whether the molecular structure of the species meets the requirements for a certain reaction family. Therefore, a subgraph recognition query is performed which checks if the molecular structure, defined by a SMARTS identifier, is present in a certain species. SMARTS form also an important instrument used for the thermodynamic properties calculation scheme discussed later. When a species is eligible for a certain reaction family, the reaction recipe, which defines the associated elementary transformations, is executed, *e.g.* bond formation and breakage. After the execution of the recipe, product species are formed, which can subsequently be subjected to a reaction family themselves giving birth to second-generation, third-generation, and so on products. The kinetic model generation iteratively continues until the reaction network is considered complete by user-defined criteria.

Reaction families are defined with an Extendable Markup Language (XML) input file. One reaction family starts with defining the name and mentioning whether or not the reaction family considers a unimolecular or bimolecular reaction. An example of the molecular rearrangements and the reaction family template is depicted in Figure 3-1 for the hydrogen abstraction reaction by molecular oxygen on a secondary carbon atom. Next follow three elements which have to be defined for every reaction family:

1. A recipe, which defines the molecular rearrangement which is to be carried out when the elementary reaction is executed with certain reactants. Examples of such transformations are breaking and forming of bonds (BREAK_BOND or FORM_BOND), losing or gaining a radical (LOSE_RADICAL or GAIN_RADICAL), *etc.* A series of elementary transformations that are representative for the hydrogen abstraction on a secondary carbon atom by molecular oxygen are visible in the global scheme in Figure 3-1, together with the associated code in frame number **1**.
2. A description of the reactant in SMARTS language, this defines the specific molecular structure in the reactants required to execute the reaction family. Thereafter follows the definition of the reactive centers in the reacting molecule which allows the code to apply the recipe to the defined atoms. Therefore, each reactive center, *i.e.* an atom, in the reactants related to the elementary transformation is given a symbol which is used in the recipe. Both unimolecular and bimolecular reactions are possible. In case of bimolecular reactions, two reactants have to be defined. Constraints can be defined on the reacting species to limit the reaction network *e.g.* only molecules with less than 2 oxygen atoms or with at most one single bond are allowed to react in the specified reaction family. It is also possible to set constraints in the SMARTS language on the atoms of the reactant. The latter is for example applied for reactions where the difference is made between primary, secondary and tertiary carbon atoms by putting constraints on the number of hydrogen atoms connected to an atom. With the code framed and indicated with number **2** in Figure 3-1, the two reacting molecules and the associated reactive centers are defined via SMARTS, *i.e.* both oxygen atoms from molecular oxygen, a secondary carbon atom and an attached hydrogen atom. One constraint is set on the carbon reactive center, that it should be a secondary carbon atom, *i.e.* having two hydrogen neighbors, and one constraint is set on the associated molecule that it cannot already contain a radical, *i.e.* the number of single electrons should equal zero to be eligible.
3. A kinetic method and associated input parameters which must be used to calculate the kinetic parameters for all reactions of the reaction family. Several options are included in Genesys, for example a(n) (modified) Arrhenius relationship, rate rules, group additivity theory and on-the-fly ab initio calculations. Each option requires different input, *e.g.* the Arrhenius relationship requires the corresponding Arrhenius parameters and group additivity theory requires a reference to the database where the group additive values can be found. For example, the code in frame number **3** in Figure 3-1 indicates that the kinetics are calculated via a rate rule with specified modified Arrhenius parameters for all of the reactions generated by this reaction family.

Optionally, the reaction family file can be ended with constraints on the products formed. For example, a maximum amount of heavy atoms or double bonds, a minimal amount of carbon atoms, or limitations on the presence of certain molecular structures.



```

<inp-reaction-family
  name="Hydrogen abstraction by oxygen - secondary carbon">
    <bimolecular>true</bimolecular>
    1 <inp-recipe>
      <inp-transformation type="FORM_BOND" centers="A,B" />
      <inp-transformation type="BREAK_BOND" centers="B,C" />
      <inp-transformation type="DECREASE_BOND" centers="A,D" />
      <inp-transformation type="GAIN_RADICAL" centers="D" />
      <inp-transformation type="GAIN_RADICAL" centers="C" />
    </inp-recipe>
    2 <inp-reactant value="1" smarts="[O]=[O]">
      <inp-reactive-center symbol="A" smarts="[$([O]=[O])]" />
      <inp-reactive-center symbol="D" neighbour="A" smarts="[$([O]=[O])]" />
    </inp-reactant>
    <inp-reactant value="2" smarts="[C;X4v4H2]-[#1]">
      <inp-reactive-center symbol="B" smarts="[$([#1]-[C;X4v4H2])]" />
      <inp-reactive-center symbol="C" neighbour="B" smarts="[$([C;X4v4H2]-[#1])]" />
      <inp-molecule-constraint type="SINGLEELECTRONCOUNT" limit="EQ" value="0" />
    </inp-reactant>
    3 <inp-kinetics type="RATE_RULE">
      <rate-coeff A="1.338E07" N="0.0" Ea="207.8"/>
      <!-- Cai 2016 -->
    </inp-kinetics>
  </inp-reaction-family>
  
```

Figure 3-1: Example of the structural rearrangements defined for the hydrogen abstraction by molecular oxygen on a secondary carbon atom and the associated reaction family template, (1) is the recipe, (2) the reactant specification and (3) the methodology to calculate the reaction rate, letters above the atoms refer to the recipe.

3.1.2. Termination criteria

Automatically generated models can easily attain thousands of reactions by carrying out reaction families, it is therefore important to include termination criteria. These criteria or rules determine when to stop the generation of new reactants and thereby excludes species which do not significantly affect the kinetic model. The latter allows one to attain a manageable size of the microkinetic model to perform simulations which still contains the essential chemistry. The user-defined input of termination criteria remains a challenging and crucial task for automatic generated models which is highly influenced by the user's experience [8]. Three types of termination criteria exist, these are the so-called rank-based, rule-based and rate-based criteria. The latter two termination criteria are implemented in Genesys.

The rate-based approach requires no user-based knowledge or rules of the chemistry since on-the-fly reactor simulations are carried out to include the essential chemistry [9, 10], for example with the software package CHEMKIN. All possible products are generated originating from the initial pool of reactants. However, the products are only added to the reaction mechanism if the calculated rate of production is higher than a pre-defined threshold. The main drawback is that the generated kinetic model and the inclusion of species is highly dependent on the thermodynamic and kinetic properties. Inaccurate thermodynamics and kinetics result in the inclusion of different species and reactions in the kinetic model as the reactor simulations result in different rates of production. Other drawbacks are a higher computational cost and that the microkinetic model only applies for certain reactor conditions, *e.g.* steam cracking of ethane will not include isobutane, this model can hence not be applied for isobutane steam cracking purposes.

The rule-based approach is the method of choice for model generation in this master dissertation. It requires chemical knowledge of the user to decide whether or not a certain reacting species can undergo the structural rearrangement defined by a reaction family. Constraints can be defined in multiple manners, *i.e.* on the reaction family, on the reactants of a reaction family and on the products formed during reaction network generation. The latter kind of constraint makes it possible to restrict product species which do not meet these constraints from the kinetic model. Examples of constraints are limitations on the number of heavy atoms, *i.e.* the number of atoms except hydrogen atoms in a molecule, the number of certain atoms, *e.g.* carbon or oxygen atoms, the number of double bonds, the presence or amount of certain molecular structures, *e.g.* hydroxyl or carbonyl groups, limitations on the smallest ring count, *etc.* Similarly, the same kind of constraints can be set on each individual reactant of a reaction family. Examples of both reaction family and product constraints, which can highly influence the generated reaction network, can be found in Figure 3-2.

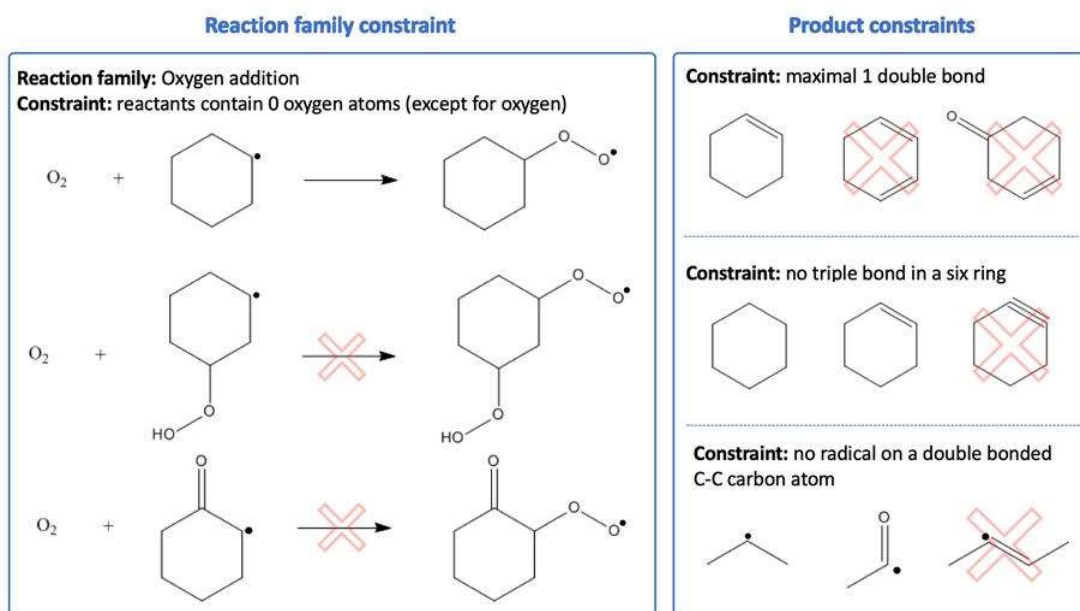


Figure 3-2: Examples of a reaction family constraint and several product constraints.

3.1.3. Reaction generation

The reaction network is created by iterating over the available reactant species and the possible reaction families as defined in the input files. A schematic representation of the reaction network generation algorithm is depicted in Figure 3-3.

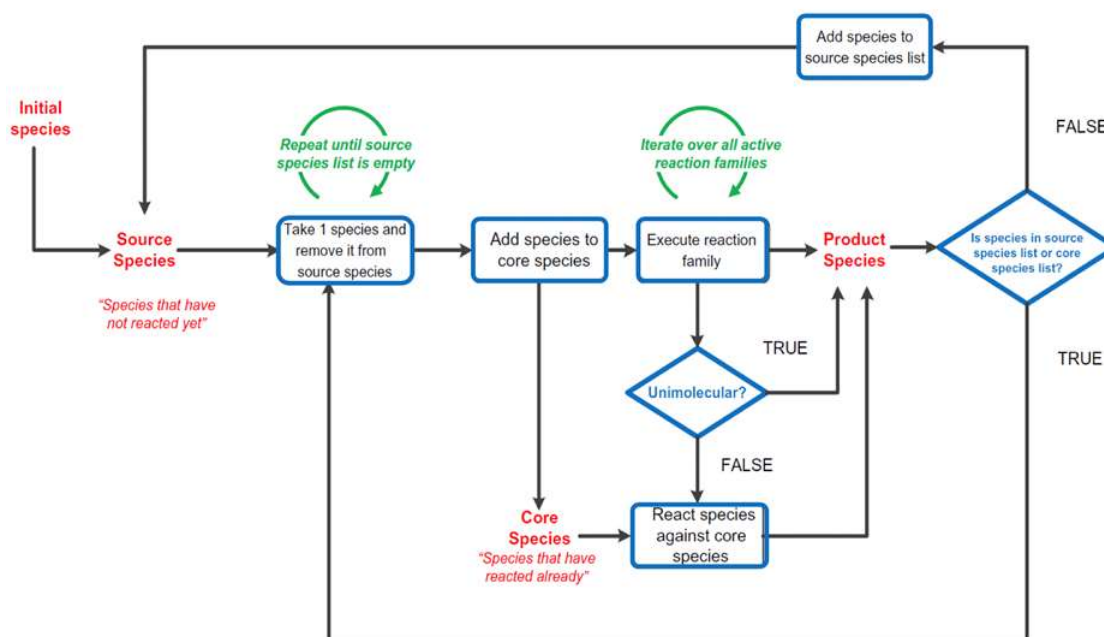


Figure 3-3: Schematic representation of the reaction network generation algorithm.

The reaction network generation starts by reading all species defined with an InChI or SMILES molecule identifier from the input file. A molecule identifier is a textual identifier for chemical substances, designed to provide a standard way to encode molecular structural information and to facilitate the search for such information in databases and in the open literature. The initial species are added to the pool of source species, which are species which have not yet reacted. Subsequently one species is taken from this pool and it is checked whether the species is eligible to undergo one or multiple of the defined reaction families. Meanwhile the species is removed from the source species pool and placed in the core species pool, which contains the species which have already reacted. After all eligible reaction families have been carried out, newly formed species are checked whether they are already present in the source or core species pool. Newly formed species are added to the source species pool. For a bimolecular reaction, the second reacting species eligible to execute the reaction are searched for in the pool of core species. The listed procedure is iterated until the source pool is empty.

3.1.4. Assigning thermodynamic properties

When the reaction network generation has finished, it is important to assign accurate thermodynamic properties, *i.e.* the enthalpy of formation ($\Delta_f H(T)$), the total entropy ($S(T)$) and heat capacities ($C_p(T)$ at constant pressure), to all species. The latter are important for amongst others calculation of equilibrium coefficients and the enthalpy of reaction. The assigned thermodynamic parameters are reported by Genesys in NASA polynomial format [11] which describes the temperature dependency of all parameters as a function of a 14 coefficient polynomial, *cf.* equations (3-1), (3-2) and (3-3).

$$\frac{C_p(T)}{R} = a_1 + a_2 \cdot T + a_3 \cdot T^2 + a_4 \cdot T^3 + a_5 \cdot T^4 \quad (3-1)$$

$$\frac{\Delta_f H(T)}{R} = a_1 \cdot T + a_2 \cdot T^2 + a_3 \cdot T^3 + a_4 \cdot T^4 + a_5 \cdot T^5 + a_6 \quad (3-2)$$

$$\frac{S(T)}{R} = a_1 \cdot \ln(T) + a_2 \cdot T + \frac{a_3 \cdot T^2}{2} + \frac{a_4 \cdot T^3}{3} + \frac{a_5 \cdot T^4}{4} + a_7 \quad (3-3)$$

In which a_x (with x a number) represents the NASA polynomial coefficient, T the absolute temperature and R the universal gas constant. Of the 14 coefficients, the first 7 are for the high-temperature range, *i.e.* typically above 1000 K, and the other 7 are for the low-temperature range. The boundary temperature to change from region is specified on the first line of the NASA polynomial definition. Care should be taken in this region since the polynomial does not always form a continuous function when the transition is made.

A prescribed procedure is followed to assign thermodynamic properties to a species. First, the databases are checked whether the parameters are not available from high-level ab initio calculations. If accurate

thermodynamic data is lacking for a certain species, a calculation methodology is applied, *i.e.* Benson group additivity theory [12].

3.1.4.1 Group additivity theory

The Benson group additivity theory is a calculation scheme to calculate thermodynamic properties which assumes that every group from which a molecule is built contributes to the thermodynamic properties of that molecule. A group is defined as a polyvalent atom in the molecule together with its immediate neighbors. The corresponding contributions, *i.e.* group additivity values (GAVs), are regressed from a dataset of species (the training dataset) of which their thermochemistry is accurately known from experiments or high-level ab initio calculations. The in-house developed GAVs are regressed solely on ab initio calculations, more specifically on the CBS-QB3 or G4 level of theory. An example illustrating the principle of group additivity is depicted in Figure 3-4. The thermodynamic properties of 2-methylnonane are calculated based on the sum of the GAVs for a primary, secondary and tertiary carbon atom. In this work the GAVs generated by Sabbe et al. [13, 14] for hydrocarbons and Paraskevas et al. [15] for oxygenates are used. For alkanes, GAVs agree well with experimental values within 2 kJ mol^{-1} for the enthalpy of formation, with the deviation increasing with increasing size of the molecule. The GAVs regressed by Paraskevas predict standard enthalpies of formation, entropies and heat capacities with a chemical accuracy of , respectively, 4 kJ mol^{-1} and $4 \text{ J mol}^{-1} \text{ K}^{-1}$ compared to ab initio calculations.

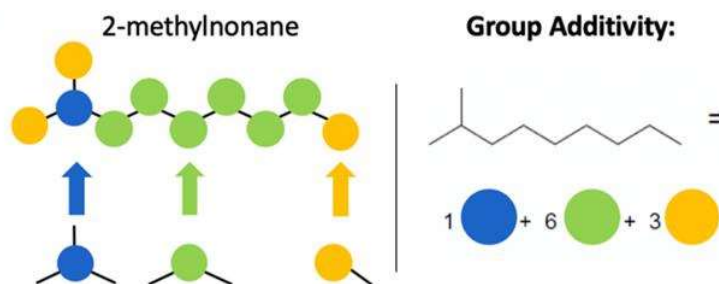


Figure 3-4: Principle of group additivity theory, an important calculation method for gas-phase thermochemistry.

Benson group additivity theory does not only account for the latter introduced GAVs which are the primary contributions. Three additional corrections are considered. Ring strain corrections (RSC) account for the presence of a ring structure creating strain which is not present in the acyclic counterparts. Non-neighbor interactions (NNI) consider interactions, which can appear between intermolecular fragments which are not directly bonded to each other, an example are gauche and eclipse interactions. And finally, resonance corrections (RES) account for the stabilization resulting from electron interactions in a molecule. The thermodynamic properties are accordingly calculated from equation (3-4).

$$\frac{C_p(T)}{S_{GAV}(T)} = \sum_{i=1}^m GAV(X_i) + \sum_{i=1}^n RSC_i + \sum_{i=1}^o NNI_i + \sum_{i=1}^p RES_i \quad (3-4)$$

The heat capacity is calculated at seven different temperatures (300, 400, 500, 600, 800, 1000 and 1500 K) and the enthalpy of formation and entropy are both calculated at 298 K. Moreover, the entropy calculated with GAVs still has to be corrected for the molecular symmetry in the molecule by equation (3-5).

$$S(T) = S_{GAV}(T) - R \cdot \ln\left(\frac{\sigma}{n_{opt}}\right) \quad (3-5)$$

In which σ resembles the total symmetry number and n_{opt} the number of optical isomers. The total symmetry number is the number of identical configurations obtained by rotating the molecule around internal symmetry axes as well as rotations around internal bonds. Optical isomers are defined as species with the same molecular structure but being non-superimposable, *e.g.* by the presence of a chiral center. Both σ and n_{opt} can be calculated automatically by Genesys using the concept of graph automorphism and stereoisomer enumeration and do not require a three-dimensional representation of the molecule. Automatically calculated symmetry numbers comply with the expected value in almost 90 % of the time, in 96 % of the time the symmetry number deviates from the expected value with only a factor of 2 or less [16]. The Benson group additivity scheme applies well for implementation in computer codes and is therefore an excellent approach to quickly calculate thermodynamic properties in case of lacking data. However, the accuracy and applicability of Benson group additivity theory is dependent on the availability of GAVs, which are for acyclic hydrocarbons readily available. If it is not possible to assign a value to every non-hydrogen atom, no thermodynamic properties are calculated and another approach should be applied, *e.g.* high-level ab initio calculations.

For the calculation of thermodynamic properties for radicals, other GAVs have been regressed which contain contributions for radical centered atoms. An alternative approach that can be used is introduced by Lay et al. [17], *i.e.* the hydrogen bond increment (HBI) method. The latter calculates the thermochemistry of a radical species $R\bullet$ as the sum of the GAVs of the associated species $R-H$ to which a contribution is added for the missing hydrogen atom.

3.1.4.2 Group additivity algorithm

The structural fragments of the different groups in SMARTS with the associated GAVs are stored in databases. To calculate thermodynamic properties via the algorithm, it starts with the input of a species in InChI or SMILES format, as indicated in Figure 3-5. Before calculating the thermodynamic properties with the group additivity scheme, it is first checked if the species is not yet available in the high-level

ab initio databases of Genesys. When present, the thermodynamic properties found in the database are assigned to the species. If the species are unavailable in the database, the thermodynamic properties are calculated via the GAV algorithm. To start the algorithm, the species is read and a check is made if every non-hydrogen atom has contributed to the thermodynamic property assignment or if the database has run through, since it is a new species none of the atoms has been visited yet and the database has not been run through. Subsequently, the SMARTS of one structural fragment is taken from the database. A matching SMARTS query is performed by CDK to check if the structural fragment is present in the molecule by subgraph recognition. When present, the corresponding atom(s) of the molecule are indicated as 'visited' and the thermodynamic properties of the species is updated with the values of the matched group. This procedure is iterated until all non-hydrogen atoms in the molecule have been visited or when the database with GAVs has been completely run through. The different GAVs are checked in the order as they are listed in the database. Once a certain molecular group has been matched with the SMARTS of a GAV, it is flagged and not searched anymore even if a more specific group can be matched. It is therefore important that GAVs are well-defined and that more general GAVs are listed last. For radical-centered groups, radical-centered GAVs exist, as well as for groups which have radical neighbors. Radicals are characterized by the valency of an atom, *e.g.* a radical carbon has valency 3, which can be indicated via the SMARTS identifier. After the assignment of primary contributions to all non-hydrogen atoms in a molecule, the three additional corrections are applied, *i.e.* RSCs, NNIs and RESs. Finally, the intrinsic entropy calculated by GAVs is corrected for symmetry and optical isomers

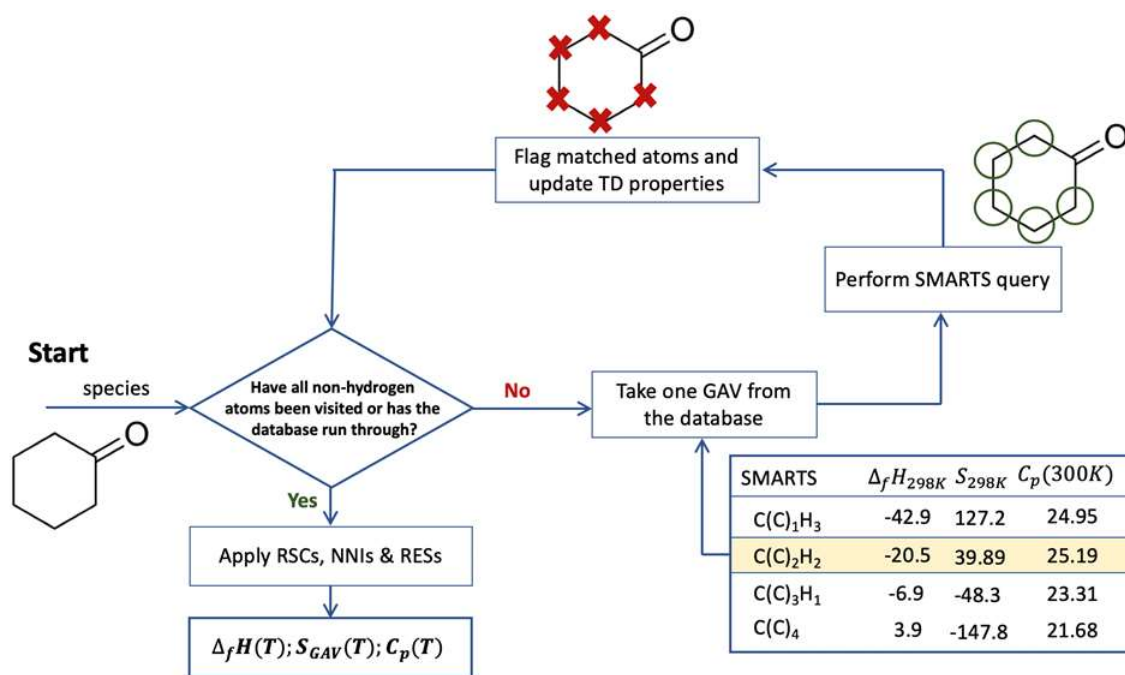


Figure 3-5: Benson group additivity algorithm implemented in Genesys to assign primary group additivity values to candidate species.

to acquire a value for the total entropy. A NASA polynomial can be obtained by fitting a smooth curve with given format, *cf.* equations (3-1), (3-2) and (3-3), through the calculated thermodynamic values at different temperatures, *i.e.* 300, 400, 500, 600, 800, 1000 and 1500 K. In case that not all non-hydrogen atoms have been visited, the thermodynamic properties cannot be calculated.

3.1.5. Assigning kinetic parameters

Besides obtaining thermodynamic properties for the species, it is also important to assign accurate kinetic parameters to all reactions present in the reaction network. The temperature dependency of reaction rate coefficients is typically described by an Arrhenius equation or modified Arrhenius equation, written respectively in equations (3-6) and (3-7).

$$k(T) = A \cdot \exp\left(-\frac{E_a}{RT}\right) \quad (3-6)$$

$$k(T) = A \cdot \left(\frac{T}{1\text{ K}}\right)^n \cdot \exp\left(-\frac{B}{RT}\right) \quad (3-7)$$

In which $k(T)$ is the reaction rate coefficient, A the pre-exponential factor and E_a the activation energy. In the modified Arrhenius expression, B is representative for the activation energy, while the new constant A and n , an exponent, account for the temperature dependency of the pre-exponential factor from the simple Arrhenius expression. Genesys typically expresses reaction rate coefficients in the units $(\text{mol m}^{-3})^m \text{ s}^{-1}$, with m depending on the reaction order and equal to 0 for unimolecular, -1 for bimolecular reactions, *etc.*

Several options are available to assign values to the Arrhenius parameters for a certain elementary reaction, these include performing experiments or regression to experimental data, performing ab initio calculations or applying reactivity-based calculation methods. Ab initio determined parameters, mostly at the CBS-QB3 level of theory, are stored in databases and are searched before applying calculation methods. Similarly to the Benson method for calculation of thermodynamic data for unknown species, a group additivity method can be applied for the calculation of reaction rate coefficients as proposed by Saeys et al. [18]. The latter assumes that the reaction rate coefficient is determined by the moiety of the reactive center rather than the molecule in its entirety. A new formula for the Arrhenius expression was introduced by Saeys et al. which is used to calculate the reaction rate via GAVs, *cf.* equation (3-8) with \tilde{A} and E_a the single-event kinetic parameters calculated by GAVs.

$$k(T) = n_e \kappa \tilde{A} \exp\left(-\frac{E_a}{RT}\right) \quad (3-8)$$

A reference reaction is considered with reference values for $\log \tilde{A}$ and E_a whereby GAVs are summed up to account for the different neighbors compared to the reference reaction, as written in equations (3-9) and (3-10).

$$\log \tilde{A} = \log \tilde{A}_{Ref} + \sum_{i=1}^n \Delta GAV_{\log \tilde{A}}^{\circ}(X_i) \quad (3-9)$$

$$E_a = E_{a,Ref} + \sum_{i=1}^n \Delta GAV_{E_a}^{\circ}(X_i) \quad (3-10)$$

With $\Delta GAV_{\log \tilde{A}}^{\circ}$ and $\Delta GAV_{E_a}^{\circ}$ the group additive values of a group within the reactive center for calculation of the pre-exponential factor and activation energy respectively. κ is the tunneling coefficient which accounts for quantum mechanical tunneling of atoms through the reaction barrier, typical values at pyrolysis temperatures are close to, but always higher than, 1. An empirical power law is used [19, 20] which correlates the tunneling coefficient with the temperature and the activation energy in the exothermic direction which is a fast and accurate method to obtain adequate values for κ . These power laws are determined from regression of the Eckart tunneling coefficient for several reactions that were calculated ab initio. n_e represents the number of single events of the reaction, which is often different from the number of single events of the reference reaction, and hence needs to be accounted for. Note that a variation of the simple Arrhenius expression is used rather than the modified Arrhenius expression. The temperature dependency of the pre-exponential is included by different GAVs which are valid in different temperature ranges.

3.1.6. Output and post-processing

The output of Genesys is written in several files which are human readable. One of the files contains the cooresponding microkinetic model which can be directly used for processing in reactor simulation software. The latter is written in the CHEMKIN format which is directly usable in the CHEMKIN software [21, 22], a commonly known reactor simulation software package. Cantera, other reactor simulation software, has a built-in option to convert the CHEMKIN format to a format usable for Cantera.

3.1.6.1 CHEMKIN

The CHEMKIN and CHEMKIN PRO software tools are used to process the generated microkinetic model and perform reactor simulations. CHEMKIN is typically used for gas-phase systems, but can also be applied to catalysis and solvation systems. The main applications are within combustion, *e.g.* in a jet-stirred or tubular reactor, flames, shock tubes *etc.* The results of reactor simulations with the newly developed kinetic model allow comparison of experimental determined data with model predictions and

indicate if the model is able to describe experimental trends or not. Moreover, CHEMKIN PRO allows the user to visualize the chemical reaction networks and analyze reaction pathways as well as indicating the sensitivity of reaction rates on the product formation rates. Besides the kinetic model, also the reactor conditions (pressure and temperature), inlet flows and reactor configuration have to be specified. An alternative open-source reactor simulation package is Cantera, which has similar functionalities as CHEMKIN, but with the possibility to adjust the source code.

3.1.6.2 *The AramcoMech base mechanism*

Over the past years, Genesys has been applied for the construction of microkinetic models for several important processes including pyrolysis and oxidation of hydrocarbon and hetero-atomic species. A base mechanism for the pyrolysis of small compounds, containing C, H, N, S and O atoms, has thereby been constructed at the LCT and is included in Genesys. The usage of a base model is necessary since thermodynamic and kinetic data obtained from GAVs are not accurate enough for small species. For a lot of small species and reactions between these species in pyrolysis, the kinetic and thermodynamic data are included in the high-level ab initio databases of Genesys. In case of combustion, no base mechanism is available in Genesys. Moreover, for typical combustion conditions, kinetics of reactions between the smallest species are often pressure dependent. This feature is not yet implemented in Genesys and highlights the current need for a separate base mechanism when modelling combustion reactions.

When a kinetic model is generated with Genesys, it can subsequently be merged with such a base mechanism to introduce the essential chemistry for smaller compounds. As a base mechanism, AramcoMech is used, as this is a well validated and updated kinetic model that uses universal species identifiers. Several other base mechanisms are available in literature, *e.g.* GRI [23] and JetSurf [24]. The model characterizes the kinetic and thermodynamic properties for a large set of hydrocarbons and oxygenates, validated over a wide set of experimental conditions including shock tubes, rapid compression machines, jet-stirred and tubular reactors. Three models are freely available: AramcoMech 1.3 (2013) [25], AramcoMech 2.0 (2016) [26-29] and AramcoMech 3.0 (2018) [30]. The more recent models contain more species, up to 8 carbon atoms for AramcoMech 3.0, and associated more reactions. The most up-to-date version, AramcoMech 3.0, is only recently published and the required format for merging with a Genesys model is not yet available, whereby it will not be used in this work. The kinetic model generated by Genesys in this master dissertation merged with AramcoMech 2.0 before carrying out reactor simulations. Future work could encompass merging the developed Genesys model with AramcoMech 1.3 and AramcoMech 3.0 to assess the performance of the different versions. Note that none of the base models contains cyclohexane or was validated for the oxidation of cyclohexane.

When merging, a master kinetic model is defined from which all species and reactions are taken as is; from the other mechanism, *i.e.* the slave mechanism, only the species and reactions which do not appear in the master kinetic model are added to the final merged model. AramcoMech is used as master kinetic model, as the thermodynamic properties and kinetics for small species and the reactions between these species specific to combustion are often not present in the Genesys high-level ab initio databases. Moreover, the AramcoMech mechanism is regressed against experiments based on the complete kinetic model.

3.2. Ab initio calculations

Accurate thermodynamic and kinetic parameters are a main and essential constituent to obtain a good microkinetic model. To obtain these values for several possible reaction pathways, quantum mechanical calculations are carried out on the high-performance supercomputer at Ghent University at the CBS-QB3 [31] level of theory as implemented in Gaussian 16 [32]. The outcome allows determination of heat capacities at different temperatures, the standard entropy and the standard enthalpy of formation for both reactants, products and transition states and the subsequent calculation of (modified) Arrhenius parameters.

Geometry optimization of transition states and stable products and the search for the lowest energy conformer are performed on the B3LYP level of theory with the 3-21G basis set, extended to the 6-31G(d,p) basis set if required to obtain convergence. Subsequently, the optimized geometry is used for calculations at the CBS-QB3 level of theory. For some transition states, intrinsic reaction coordinate (IRC) calculations have been carried out to identify the reactants and products when this was not clear from the imaginary frequency. For the calculation of thermodynamic properties, most internal modes are treated as harmonic oscillators. Modes resembling rotations around a single bond are treated as one-dimensional internal rotors, as long as the rotational barrier does not exceed 50 kJ mol⁻¹, for reactants, products and transition states. When the energy exceeds the latter boundary value, the corresponding mode is also represented by a harmonic oscillator. The hindered rotor profiles are scanned in steps of 10 degrees at the B3LYP/6-31G(d) level of theory with relaxed surface scans in which all coordinates, except for the dihedral angle which is varied, are re-optimized at each scanned angle. These hindered rotor profiles additionally served as an indication that the lowest energy conformer was found.

The Fourier series expansion of the hindered rotor profile together with the reduced moment of inertia calculated at the I^(2,3) level, as defined by East and Radom [33], are used to construct the time-independent Schrödinger equation for the one-dimensional internal rotation. The eigenvalues of the solution of the time-independent Schrödinger equation are used to determine the partition function as a function of temperature. The total partition function describes the statistical properties of the thermodynamic system in equilibrium and consists out of contributions for the translational, rotational, vibrational and electronic partition function, *cf.* equation (3-11).

$$q_{tot} = q_{elec} q_{trans} q_{rot} q_{vib} \quad (3-11)$$

The standard thermodynamic formulas from statistical physics are then applied to obtain enthalpies, intrinsic entropies and heat capacities as a function of temperature, *cf.* equations (3-12), (3-13) and (3-14). To calculate the enthalpy of formation at standard conditions, *i.e.* $\Delta_f H^\circ$, the atomization method is applied.

$$H(T) = RT^2 \left(\frac{\partial \ln q_{tot}(T)}{\partial T} \right) + RT \quad (3-12)$$

$$S(T) = R \left(1 + \ln \left(\frac{q_{tot}(T)}{N_A} \right) + T \frac{\partial \ln q_{tot}(T)}{\partial T} \right) \quad (3-13)$$

$$C_p(T) = R \left(1 + 2T \frac{\partial \ln q_{tot}(T)}{\partial T} + T \frac{\partial^2 \ln q_{tot}(T)}{\partial T^2} \right) \quad (3-14)$$

Calculated enthalpies of formation at the CBS-QB3 level of theory are corrected with spin-orbit corrections (SOC) and with empirically determined bond additive corrections (BACs) before usage in the microkinetic model. The latter corrections result from a systematic error introduced by ab initio calculations when these are compared to experimentally obtained enthalpies of formation. BACs employ the underlying assumption that the errors in calculated bond energies are constant for each type of bond and that these errors are additive. Isodesmic bond additive corrections were introduced by Petersson et al. [34] for accurate calculation of thermodynamic parameters by comparing reactants and products from isodesmic reactions. An isodesmic reaction conserves the number of bonds of each order between each pair of atom types. If the error in the calculated bond energy is constant for each type of bond and is additive, then the calculated enthalpy change for an isodesmic reaction would be exact. The latter allows the accurate calculation of thermodynamic parameters for one species of an isodesmic reaction if the thermodynamics of the other species are known. This concept has been replaced by BACs for which the corrections are determined by regression based on comparison of ab initio results with experimental data. The uncertainty on the bond energy is therefore replaced by the concept of BACs. Sabbe et al. observed a mean absolute deviation (MAD) from experiments of 8.2 kJ mol⁻¹ and 1.3 kJ mol⁻¹ for respectively calculations without and with BACs for a set of hydrocarbon species [14]. For transition states BACs are not defined, as no experimental data is available, and thus the enthalpies of formation for stable species and transition states, used for construction of the potential energy surface, are not corrected with SOC or BACs. The same is valid for the determination of kinetic parameters. The used BAC corrections for the structural bonds in this master dissertation are not those as implemented in Genesys, as these are not yet optimized, but those listed in Table 3-1.

After determination of the partition functions to determine thermodynamic properties, the reaction rate coefficients for elementary reactions are calculated via the transition state theory (TST) [35]. The transition state theory assumes that there exists a first order saddle point, *i.e.* the transition state, which divides the phase space in a reactant and product region. To carry out a reaction, the reactants have to follow a trajectory along the reaction coordinate which passes via the transition state to the products over the PES. The transition state is identified as the point with the highest energy on the minimum energy path and is characterized by one imaginary frequency. TST assumes a chemical equilibrium

Table 3-1: Structural bonds and corresponding BAC values.

Structural bond	BAC Value
C – H	-0.188
C – C	-0.003
C = C	-0.476
C ≡ C	-0.558
c – c	-0.045 (aromatic)
C – O	1.178
C = O	0.099
O – H	-0.180
O – O	1.003

between the reactants and the transition state. The latter allows the calculation of the reaction rate coefficients with information of only two points on the PES, *i.e.* of the stable reactant species and the transition state. Equation (3-15) can be used for calculation of the reaction rate coefficient.

$$k(T) = \frac{k_B T}{h} \left(\frac{RT}{P} \right)^{-\Delta^\ddagger n} \exp \left(- \frac{\Delta^\ddagger H - T \Delta^\ddagger S}{k_B T} \right) \quad (3-15)$$

In which k_B is the Boltzmann constant, T the absolute temperature, h Planck's constant, R the gas constant, P the pressure, $\Delta^\ddagger n$ the change in number of molecules upon formation of the transition state, $\Delta^\ddagger H$ the standard enthalpy of activation and $\Delta^\ddagger S$ the standard activation entropy. Both the energy and partition function of the transition state are obtained by computational chemistry methods.

To end the part on ab initio calculations a remark is made concerning the different configurations possible when working with substituted cyclic species. Considering cyclohexane, two orientations are possible of the hydrogen atoms or substituents on the cyclohexane ring, *i.e.* equatorial and axial, as indicated in Figure 3-6.

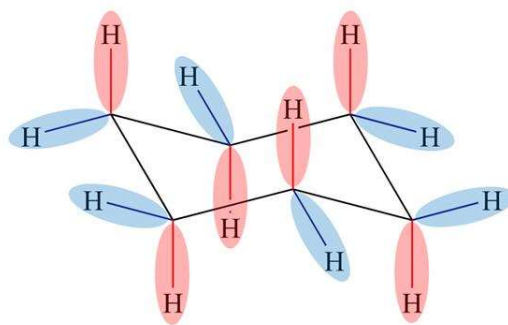


Figure 3-6: Equatorial (blue) and Axial (red) positioned hydrogen atoms for the chair conformer of cyclohexane.

Different energies and accompanied different reactivities are expected for equatorial and axial substituents due to steric hindrance. These differences are not accounted for in a similar manner as open-chain molecules, where the different conformers are accounted for by a one-dimensional hindered rotor approximation. A similar approach would require accounting for flipping of the cyclic structure from the chair conformer, to the boat conformer and back to the chair conformer. However, as this approach is very complex, no algorithm is available and the difference in energy between axial and equatorial substituents are often omitted in literature. In this master dissertation, in most cases the energy and reactivity of both the axial and equatorial conformer are accounted for and compared. For example, hydrogen abstraction reactions from cyclohexane are calculated for an axial and equatorial hydrogen atom. Also for the case of substituted cyclohexane present during the low temperature oxidation channels of cyclohexane, both conformers are accounted for. The differences in energy are discussed in the following chapter.

In the final kinetic model however, both conformers are lumped and the respective entropy is the one calculated for the lowest energy conformer, corrected by $R\ln(2)$, *cf.* equation (3-5), to account for the lumping of both conformers. Similar, for reactions, both transition states are lumped and the fastest reaction is multiplied by a factor of 2.

3.3. Conclusions

In this methodology Chapter, the automatic kinetic model generator developed at the Laboratory for Chemical Technology, *i.e.* Genesys, has been introduced. A second part covered the procedures followed to carry out and post-process the results of high-level ab initio calculations.

Genesys requires as input the initial pool of reactants and a reaction family file which contains all the elementary reactions which are considered in the generated reaction network. Termination criteria are applied to limit the size of the reaction network while maintaining the essential chemistry. Another crucial aspect is the assignment of accurate thermodynamic and kinetic parameters, which can be searched in a database or can be calculated via calculations schemes such as the Benson group additivity theory. Ab initio calculations can be performed if the latter methods are not sufficient. To obtain the final developed model, it is first merged with a base mechanism before reactor simulations are done to include the detailed chemical reactivity of small compounds which are not present in Genesys.

In order to obtain accurate values for the thermodynamic properties of species and (modified) Arrhenius parameters for chemical reactions, ab initio calculations have been carried out. Therefore, first a geometry optimization is performed on the CBS-QB3 level of theory followed by hindered internal rotor calculations. Conformational search is employed to obtain the lowest energy conformer of species and transition states. Post-processing of these results is performed automatically by Genesys via the conventional theories derived in statistical physics, *e.g.* transition state theory. Care should be taken with the outcome, *i.e.* symmetry numbers, number of single events, *etc.* should be verified as these are not always correctly calculated via the computer algorithms.

3.4. References

1. E.L. Willighagen, J.W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliaskova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C.T. Evelo, R. Guha, and C. Steinbeck, *The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching*. Journal of Cheminformatics, 2017. **9**(1): p. 33.
2. J.W. May and C. Steinbeck, *Efficient ring perception for the Chemistry Development Kit*. Journal of Cheminformatics, 2014. **6**(1): p. 3.
3. C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, and E. Willighagen, *The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics*. Journal of Chemical Information and Computer Sciences, 2003. **43**(2): p. 493-500.
4. C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, and E.L. Willighagen, *Recent developments of the Chemistry Development Kit (CDK) - An open-source Java library for chemo- and bioinformatics*. Current Pharmaceutical Design, 2006. **12**(17): p. 2111-2120.
5. D. Weininger, *SMILES, a chemical language and information-system.1. Introduction to methodology and encoding rules*. Journal of Chemical Information and Computer Sciences, 1988. **28**(1): p. 31-36.
6. S.E. Stein, S.R. Heller, and D. Tchekhovski, *An open standard for chemical structure representation: The IUPAC Chemical Identifier*. Proceedings of the 2003 International Chemical Information Conference, ed. H. Collier. 2003, Glos: Infonortics Ltd. 131-143.
7. R. Van de Vijver, N.M. Vandewiele, P.L. Bhoorasingh, B.L. Slakman, F.S. Khanshan, H.H. Carstensen, M.F. Reyniers, G.B. Marin, R.H. West, and K.M. Van Geem, *Automatic Mechanism and Kinetic Model Generation for Gas- and Solution-Phase Processes: A Perspective on Best Practices, Recent Advances, and Future Challenges*. International Journal of Chemical Kinetics, 2015. **47**(4): p. 199-231.
8. W.H. Green, P.I. Barton, B. Bhattacharjee, D.M. Matheu, D.A. Schwer, J. Song, R. Sumathi, H.H. Carstensen, A.M. Dean, and J.M. Grenda, *Computer construction of detailed chemical kinetic models for gas-phase reactors*. Industrial & Engineering Chemistry Research, 2001. **40**(23): p. 5362-5370.
9. W. Sun, Z. Chen, X. Gou, and Y. Ju, *A path flux analysis method for the reduction of detailed chemical kinetic mechanisms*. Combustion and Flame, 2010. **157**(7): p. 1298-1307.
10. R.G. Susnow, A.M. Dean, W.H. Green, P. Peczak, and L.J. Broadbelt, *Rate-based construction of kinetic models for complex systems*. Journal of Physical Chemistry A, 1997. **101**(20): p. 3731-3740.
11. A. Burcat, *Thermochemical Data for Combustion Calculations*. 1984. p. 455-473
12. S.W. Benson and J.H. Buss, *Additivity rules for the estimation of molecular properties - thermodynamic properties*. Journal of Chemical Physics, 1958. **29**(3): p. 546-572.
13. M.K. Sabbe, M. Saeys, M.F. Reyniers, G.B. Marin, V. Van Speybroeck, and M. Waroquier, *Group additive values for the gas phase standard enthalpy of formation of hydrocarbons and hydrocarbon radicals*. Journal of Physical Chemistry A, 2005. **109**(33): p. 7466-7480.
14. M.K. Sabbe, F. De Vleeschouwer, M.F. Reyniers, M. Waroquier, and G.B. Marin, *First Principles Based Group Additive Values for the Gas Phase Standard Entropy and Heat Capacity of Hydrocarbons and Hydrocarbon Radicals*. Journal of Physical Chemistry A, 2008. **112**(47): p. 12235-12251.
15. P.D. Paraskevas, M.K. Sabbe, M.-F. Reyniers, N. Papayannakos, and G.B. Marin, *Group Additive Values for the Gas-Phase Standard Enthalpy of Formation, Entropy and Heat Capacity of Oxygenates*. Chemistry – A European Journal, 2013. **19**(48): p. 16431-16452.

16. N.M. Vandewiele, R. Van de Vijver, K.M. Van Geem, M.-F. Reyniers, and G.B. Marin, *Symmetry calculation for molecules and transition states*. Journal of Computational Chemistry, 2015. **36**(3): p. 181-192.
17. T.H. Lay, J.W. Bozzelli, A.M. Dean, and E.R. Ritter, *Hydrogen-atom bond increments for calculation of thermodynamic properties of hydrocarbon radical species*. Journal of Physical Chemistry, 1995. **99**(39): p. 14514-14527.
18. M. Saeys, M.F. Reyniers, G.B. Marin, V. Van Speybroeck, and M. Waroquier, *Ab initio group contribution method for activation energies for radical additions*. Aiche Journal, 2004. **50**(2): p. 426-444.
19. M.K. Sabbe, A.G. Vandeputte, M.-F. Reyniers, M. Waroquier, and G.B. Marin, *Modeling the influence of resonance stabilization on the kinetics of hydrogen abstractions*. Physical Chemistry Chemical Physics, 2010. **12**(6): p. 1278-1298.
20. A.G. Vandeputte, M.K. Sabbe, M.-F. Reyniers, and G.B. Marin, *Kinetics of a hydrogen abstractions from thiols, sulfides and thiocarbonyl compounds*. Physical Chemistry Chemical Physics, 2012. **14**(37): p. 12773-12793.
21. D.T. Daly and P. Nag, *Combustion Modeling of Soot Reduction in Diesel and Alternate Fuels using CHEMKIN®*. 2001, SAE International.
22. CHEMKIN 2007: San Diego, CA, USA: Reaction Design, Inc.
23. D.M.G. Gregory P. Smith, Michael Frenklach, Nigel W. Moriarty, Boris Eiteneer, Mikhail Goldenberg, C. Thomas Bowman, Ronald K. Hanson, Soonho Song, William C. Gardiner, Jr., Vitali V. Lissianski, and Zhiwei Qin. Available from: http://www.me.berkeley.edu/gri_mech/.
24. E.D. H. Wang, B. Sirjean, D. A. Sheen, R. Tango, A. Violi, J. Y. W. Lai, F. N. Egolfopoulos, D. F. Davidson, R. K. Hanson, C. T. Bowman, C. K. Law, W. Tsang, N. P. Cernansky, D. L. Miller, R. P. Lindstedt, *A high-temperature chemical kinetic model of n-alkane (up to n-dodecane), cyclohexane, and methyl-, ethyl-, n-propyl and n-butyl-cyclohexane oxidation at high temperatures*. September 19, 2010
25. W.K. Metcalfe, S.M. Burke, S.S. Ahmed, and H.J. Curran, *A Hierarchical and Comparative Kinetic Modeling Study of C1 – C2 Hydrocarbon and Oxygenated Fuels*. International Journal of Chemical Kinetics, 2013. **45**(10): p. 638-675.
26. Y. Li, C.-W. Zhou, K.P. Somers, K. Zhang, and H.J. Curran, *The oxidation of 2-butene: A high pressure ignition delay, kinetic modeling study and reactivity comparison with isobutene and 1-butene*. Proceedings of the Combustion Institute, 2017. **36**(1): p. 403-411.
27. C.-W. Zhou, Y. Li, E. O'Connor, K.P. Somers, S. Thion, C. Keese, O. Mathieu, E.L. Petersen, T.A. DeVerter, M.A. Oehlschlaeger, G. Kukkadapu, C.-J. Sung, M. Alrefae, F. Khaled, A. Farooq, P. Dirrenberger, P.-A. Glaude, F. Battin-Leclerc, J. Santner, Y. Ju, T. Held, F.M. Haas, F.L. Dryer, and H.J. Curran, *A comprehensive experimental and modeling study of isobutene oxidation*. Combustion and Flame, 2016. **167**: p. 353-379.
28. U. Burke, W.K. Metcalfe, S.M. Burke, K.A. Heufer, P. Dagaut, and H.J. Curran, *A detailed chemical kinetic modeling, ignition delay time and jet-stirred reactor study of methanol oxidation*. Combustion and Flame, 2016. **165**: p. 125-136.
29. S.M. Burke, U. Burke, R. Mc Donagh, O. Mathieu, I. Osorio, C. Keese, A. Morones, E.L. Petersen, W. Wang, T.A. DeVerter, M.A. Oehlschlaeger, B. Rhodes, R.K. Hanson, D.F. Davidson, B.W. Weber, C.-J. Sung, J. Santner, Y. Ju, F.M. Haas, F.L. Dryer, E.N. Volkov, E.J.K. Nilsson, A.A. Konnov, M. Alrefae, F. Khaled, A. Farooq, P. Dirrenberger, P.-A. Glaude, F. Battin-Leclerc, and H.J. Curran, *An experimental and modeling study of propene oxidation. Part 2: Ignition delay time and flame speed measurements*. Combustion and Flame, 2015. **162**(2): p. 296-314.

30. C.-W. Zhou, Y. Li, U. Burke, C. Banyon, K.P. Somers, S. Ding, S. Khan, J.W. Hargis, T. Sikes, O. Mathieu, E.L. Petersen, M. AlAbbad, A. Farooq, Y. Pan, Y. Zhang, Z. Huang, J. Lopez, Z. Loparo, S.S. Vasu, and H.J. Curran, *An experimental and chemical kinetic modeling study of 1,3-butadiene combustion: Ignition delay time and laminar flame speed measurements*. Combustion and Flame, 2018. **197**: p. 423-438.
31. J.A.M. Jr., M.J. Frisch, J.W. Ochterski, and G.A. Petersson, *A complete basis set model chemistry. VI. Use of density functional geometries and frequencies*. The Journal of Chemical Physics, 1999. **110**(6): p. 2822-2827.
32. G.W.T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16*, R. B.01, Editor. 2016: Wallingford CT.
33. A.L.L. East and L. Radom, *Ab initio statistical thermodynamical models for the computation of third-law entropies*. Journal of Chemical Physics, 1997. **106**(16): p. 6655-6674.
34. G.A. Petersson, D.K. Malick, W.G. Wilson, J.W. Ochterski, J.A. Montgomery, and M.J. Frisch, *Calibration and comparison of the Gaussian-2, complete basis set, and density functional methods for computational thermochemistry*. Journal of Chemical Physics, 1998. **109**(24): p. 10570-10579.
35. K.J. Laidler and M.C. King, *Development of transition-state theory*. The Journal of Physical Chemistry, 1983. **87**(15): p. 2657-2664.

4.

Gas-phase model for cyclohexane oxidation

4 . GAS-PHASE MODEL FOR CYCLOHEXANE OXIDATION	101
4.1. INTRODUCTION	102
4.2. PUBLISHED EXPERIMENTAL AND KINETIC MODELLING STUDIES	104
4.3. KINETIC MODEL CONSTRUCTION	107
4.3.1. Kinetic model generation with Genesys	107
4.3.2. New introduced GAVs for -OOH and -OO•	110
4.4. AB INITIO CALCULATIONS RESULTS	112
4.4.1. Thermodynamic parameters	112
4.4.1.1 Single substituted six-membered ring structures	112
4.4.1.2 Double substituted six-membered ring structures	115
4.4.2. Hydrogen abstractions	118
4.4.3. Energy scans	119
4.4.3.1 Addition of molecular oxygen to the cyclohexyl radical	119
4.4.3.2 Cyclohexyl hydroperoxide oxygen-oxygen scission	120
4.4.3.3 Cyclohexane conformer search	121
4.4.4. Potential energy surface	122
4.5. PERFORMANCE OF THE DEVELOPED KINETIC MODEL	128
4.6. CONCLUSIONS	131
4.7. REFERENCES	133

4.1. Introduction

Cyclic components, so-called naphthenes, form a significant fraction of energy fuels, up to 3-10% of gasoline and up to 35% of diesel fuels [1, 2]. Since it is impossible to study the chemical reactivity of all compounds present in energy fuels, it is more typical to use a reference component to represent a certain mixture, e.g. iso-octane and n-dodecane for diesel fuels and typically cyclohexane for naphthenes [3]. Moreover, cycloalkanes are key components for the formation of unwanted aromatic species. Therefore, knowledge about the fundamental chemistry of the cyclohexane oxidation with a detailed microkinetic model allows optimization of the oxidation conditions, e.g. maximal heat production with minimal carbon monoxide and other byproducts formation. In this master dissertation, the microkinetic model will be used for other purposes, namely to gather insight concerning the elementary chemistry between cyclohexane and molecular oxygen. In addition, several elements of the gas-phase model will be reused for the development of a liquid-phase model, especially since literature indicates that in the low temperature gas-phase oxidation zone similar reactions occur as during the liquid-phase oxidation.

In this chapter, a microkinetic model is developed which gathers insight in the product formation pathways and the rate in which products are formed. The latter allows prediction of conversion rates and product yields for a variety of reaction conditions, e.g. temperature, pressure, reactor length for tubular reactors, residence time for CSTRs, *etc.* Detailed kinetic models however require the inclusion of a lot of compounds and associated a lot of possible reactions which makes the manual construction of it sometimes impossible and certainly error-prone. Genesys will be used to construct the microkinetic model solely based on elementary reactions taking place during the gas-phase oxidation of cyclohexane. As described in literature, the elementary steps included in models for the oxidation of cyclic alkanes are similar to those describing the oxidation of acyclic alkanes [4], which are used as inspiration source. The model will eventually be validated which will confirm the well-functioning of the automatic kinetic model generation framework for the gas phase.

First, in the literature it is searched which experimental and modelling work has already been performed concerning the gas-phase oxidation of cyclohexane. The developed kinetic model can be compared with other models proposed to investigate the performance. Moreover, experimental data obtained from literature is used to validate the model.

Kinetic model development via Genesys requires defining the different elementary reactions which are considered in the reaction network via a reaction family input file. Furthermore, assignment of kinetic and thermodynamic parameters for the species and reactions is required via group additivity calculation schemes or *ab initio* calculations. A potential energy surface is made for the important reaction intermediates.

The final model is compared with the models described in literature and the agreement of model predictions with experimental measurements is used to validate the performance. Moreover, the most important reaction pathways can be tracked associated with the main product formation rates.

During this chapter a convention is adapted to indicate the position of radicals and substituents on the six-ring. The main functional group, which will be clear from the context, is attached to the 0-labelled carbon-atom and counting is started there. The neighbor carbon-atoms are labelled 1, their neighbors are labelled 2 and the final carbon atom is labelled 3, as indicated in Figure 4-1. Molecules containing double bonds are labelled with the two numbers of the carbon atoms between which the double bond is formed. Molecules with three or more substituents are not covered in this chapter.

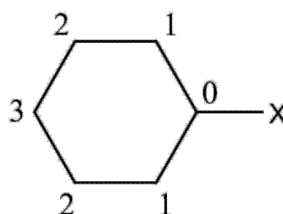


Figure 4-1: Labelling convention of the carbon atoms with X the main functional group where counting is started.

4.2. Published experimental and kinetic modelling studies

Cyclohexane is an important model compound for naphthenes and is due to its high symmetry easier to model than for example cyclopentane or methylcyclohexane. A lot of research has already been published in the open literature concerning the oxidation (and pyrolysis) of cyclohexane which is covered in this section.

A low-temperature mechanism to reproduce oxidation experiments with cyclohexane performed in a static reactor at 635 K and low pressure, i.e. 6 kPa, with an equivalence ratio of 9 is proposed by Klai and Baronnet [5, 6]. However, the latter conditions are rather far off from those observed in actual combustion engines. The research group from Orléans developed a high-temperature mechanism to model experimental results in a jet-stirred reactor with temperatures ranging from 750K to 1200K, pressures from 1 to 10 atm. and equivalence ratios between 0.5 and 1.5 [7, 8]. Moreover, the model was validated at higher temperatures, but it is not able to predict the cyclohexane conversion in the negative temperature coefficient (NTC) zone at lower temperatures. The latter indicates that important elementary reactions are missing to describe the lower temperature chemistry. Combustion chemistry is commonly divided in three temperature regimes: the low temperature region with temperatures approximately between 300 – 550 K, the NTC region between 550 - 700 K and the high temperature typically above 1000 K. In each region different reaction pathways dominate. To obtain a detailed kinetic mechanism it is important that reaction pathways for all three regimes are included [9]. The initial increase in reactivity in the NTC zone is owing to chain reactions increasing the number of radicals. With increasing temperature other reactions take over which do consume radicals and as a result no conversion is observed between the NTC zone and the high temperature region.

More recently Buda et al. [1] have presented a modelling study concerning the oxidation of cyclohexane covering the low to intermediate temperature region (650 to 1050 K), including prediction of the negative temperature coefficient zone. The model was a reinvestigation of an earlier model developed by Granata et al. [10]. EXGAS, an alternative automatic kinetic model generator for Genesys, was used to construct the model. EXGAS provides a reaction base for C₀-C₂ including all reactions for radicals and molecules containing less than three carbon atoms as well as a coupled reaction base for C₃-C₄ unsaturated hydrocarbons, including formation of benzene. The model was updated by Sirjean and coworkers who did modelling work concerning cyclic alkanes [11].

Silke et al. proposed a detailed mechanism validated against data from Lille [12] and Orléans [7, 8]. A potential energy diagram was made based on the thermodynamic properties of the species involved which can be used as comparison with the developed energy diagram in this work.

The same group of the model of Buda et al., published a newer model which now allows model predictions for a temperature range between 500 K and 1100 K also generated with EXGAS [13]. The model was validated with experiments performed on a jet-stirred reactor (JSR). The same experimental data will be used to validate the model generated with Genesys. The experiments were performed at a constant pressure of 1.07 bar, with a residence time of $2 (\pm 0.04)$ s, temperatures in the range 500 to 1100 K, an inlet mole fraction of cyclohexane equal to 0.0067 and equivalence ratios from 0.5, 1 and 2.

Experimental data is available for multiple reactor configurations including static vessel, shock tube, rapid compression machine, jet-stirred reactor and flame reactor (laminar burning, premixed laminar flat flame, ...) covering a wide range of oxidation conditions. An overview of several experimental works is given in Table 4-1.

Table 4-1: Overview of available research concerning cyclohexane oxidation experiments.

Reactor configuration	Temperature [K]	Pressure	Equivalence Ratio [-]	Reference
Static vessel	533-673	0.067-0.13 bar	Cyclohexane/air (1:2 – 2:1)	Zeelenberg et al. [14]
Static vessel	503-623	0.027-0.27 bar	Cyclohexane/O ₂ (1:1)	Bonner et al. [15]
Static vessel	635	0.047 bar	$\phi=9.0$	Klai et al. [5, 6]
Static vessel	673-773	0.02 bar	$0.9 < \phi < 5.1$	Handford et al. [16]
Shock tube	1300-1500	1.5-5.0 atm	$\phi=0.5, 1.0$	Hong et al. [17]
Shock tube	850-1380	15-50 atm	$\phi=0.25, 0.5, 1.0$	Daley et al. [18]
Shock tube	1300-1500	7.3-9.5 atm	$\phi=0.25, 0.5, 1.0$	Sierjean et al. [19]
Rapid compression machine	600-900	7-14 bar	$\phi=1.0$	Lemaire et al. [12]
Rapid compression machine	680-910	12-40 bar	$\phi=0.25, 0.5, 1.0$	Vranckx et al. [20]
Jet-stirred reactor	750-1100	10 atm	$0.5 < \phi < 1.5$	Voisin et al. [7]
Jet-stirred reactor	850-1100	1, 2 and 10 atm	$0.5 < \phi < 1.5$	El Bakali et al. [8]
Jet-stirred reactor	500-1100	1.07 bar	$\phi=0.5, 1.0, 2.0$	Serinyel et al. [13]
Laminar burning velocity	353	1, 2, 5 and 10 atm	$0.6 < \phi < 1.6$	Wu et al. [21]

Since cyclohexane is frequently used as surrogate, it has also been introduced in several combustion mechanisms for gasoline, diesel and jet fuels. Such models can also be used to compare the developed model although they are not specifically made to predict the oxidation of cyclohexane and may thereby lack (important) reactions. Especially reactions which form cyclohexanone and cyclohexanol are missing since these are of lesser importance during gas-phase oxidation. The models used in this work are JetSurf [22], the model developed by the CRECK modelling group under supervision of Ranzi [23] and the combustion model developed by the group of Westbrook [24].



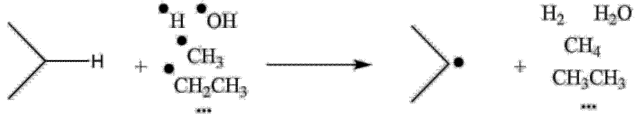

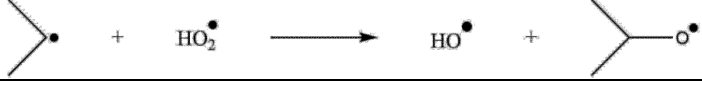
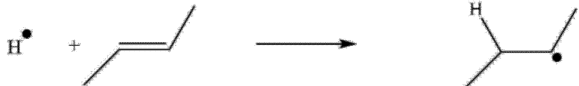
4.3. Kinetic model construction

The kinetic model for cyclohexane is developed based on reactions important for the oxidation of acyclic alkanes. The latter are defined as elementary reaction families accompanied with kinetic parameters from literature (rate rules, Arrhenius parameters, GAVs) and supplemented with data from self-performed ab initio calculations for the important reaction pathways. Thermodynamic properties are assigned via group additivity schemes or self-performed ab initio calculations for the important species. The ab initio results are covered in the next section.

4.3.1. Kinetic model generation with Genesys

The reaction families used to construct the reaction network are listed up in Table 4-2 including the references for kinetic parameters. For the reaction pathways which are considered as important, ab initio calculations have been carried out from which the modified Arrhenius parameters have been calculated. For the other reactions, rate rules, GAVs or literature data is taken.

Table 4-2: Overview of the general oxidation reaction families included in the cyclohexane oxidation model including references, grouped in bimolecular and unimolecular reactions.

BIMOLECULAR REACTION FAMILIES	
Hydrogen abstraction by oxygen - Initiation reaction	
	[25, 26]
Oxygen addition	
	[25]
Hydrogen abstraction by •H, HO2•, •OH, RO•, RO2•, alkyl radicals, ...	
	[25, 27-31]
Hydrogen abstraction next to radical by O2 with alkene formation	
	[1]
Addition of O by HO2•	
	[25]
H• addition	
	[32, 33]

UNIMOLECULAR REACTION FAMILIES

ROOH scission to RO• and •OH



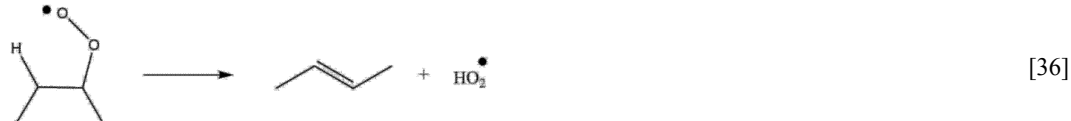
HOOH scission to •OH



Cyclohexane ring-opening and isomerization



Alkene formation from ROO•



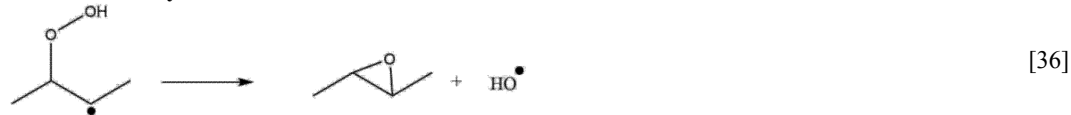
Intramolecular hydrogen abstractions by ROO•



Isomerization of R-O• to R=O



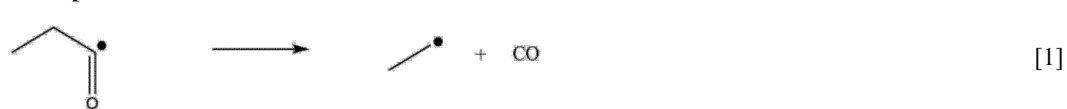
Formation of cyclic ether and •OH from QOOH



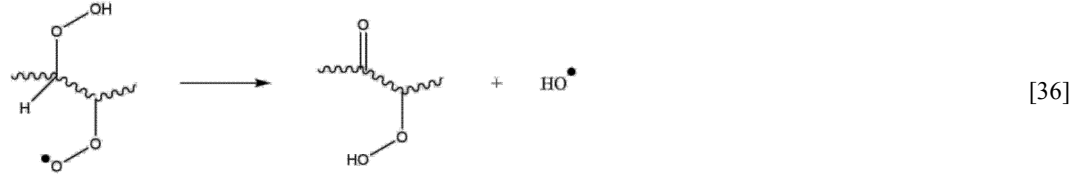
Carbon-centered β-scission



CO alpha-scission



Chain branching •OOQOOH to carbonyl hydroperoxide and •OH



Several hydrogen abstraction reactions are included with oxygen, hydroxyl, hydroperoxyl, alkyl and other radicals. For the kinetics, a distinction is made for abstraction from a primary, secondary or tertiary carbon atom. For several cases, the rate rule is available for hydrogen abstraction next to a double bond, i.e. an allylic structure is formed, which results in a resonance stabilized structure. The latter is included for abstraction with oxygen and the hydroxyl, hydroperoxyl and hydrogen radical. For hydrogen

abstraction with oxygen on tertiary carbon atoms the rate rule is missing. The kinetic parameters are therefore calculated by extrapolation of the value from primary and secondary carbon atoms.

For isomerization reactions which invoke an internal hydrogen abstraction, different reaction families are defined to allow usage of different kinetic parameters depending on the number of atoms which form a cyclic ring in the transition state and the type of carbon atom, *i.e.* primary, secondary or tertiary. Similarly, different kinetic parameters are defined for formation of a cyclic ether with a 3-ring, a 4-ring and a 5-ring.

Reactions to form cyclohexanone and cyclohexanol are in general not important in the gas phase and most of them are therefore neglected. However, this are the main components of interest in this work and thus the most important reactions are defined which form cyclohexanone, cyclohexanol and derived species. For the gas phase, it will have only minor influence on the performance of the global model, but it can result in new insight on the production of these compounds on the molecular level and even new production routes for the future. Moreover, the availability of data for these reactions will facilitate the transition from gas phase to liquid phase since kinetic data is already available for the gas-phase reactions. Note that only some of the cyclohexanone or cyclohexanol formation reactions are calculated *ab initio* due to the lesser importance for the gas-phase oxidation.

An important reaction in the liquid-phase oxidation of cyclohexane is the combination of two cyclohexyl peroxide radicals with formation of cyclohexanol, cyclohexanone and molecular oxygen as depicted in Figure 4-2. The reaction is included in the model with kinetic parameters which have been taken over from Buda et al. [1] for the gas phase.

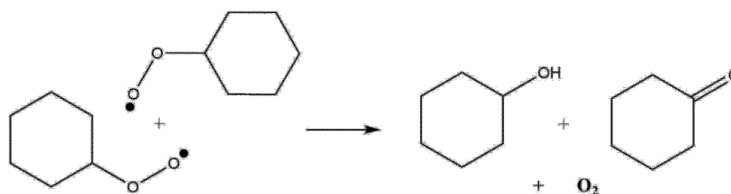


Figure 4-2: Cyclohexanone and cyclohexanol formation reaction important in the liquid phase.

Though, the importance of the latter reaction is questioned since two cyclohexyl peroxide radicals have to find each other, which is a rather rare event. Other ketone formation reactions include all kinds of abstractions or shifts of the hydrogen atom, connected to the carbon atom on which the hydroperoxide group is substituted. An example is depicted in Figure 4-3 where the 1,3-hydrogen shift takes place to the peroxide radical. For the liquid phase, hydrogen abstraction from the cyclohexyl peroxide radical, *e.g.* by molecular oxygen, leading to ketone formation is in our mindset more important for cyclohexanone formation, since molecular oxygen is readily available in the reaction mixture. Further research is necessary to investigate this theory.

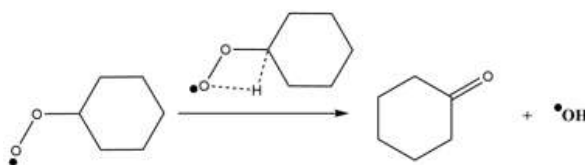


Figure 4-3: Cyclohexanone formation by a 1,3-hydrogen shift to form cyclohexanone and a hydroxyl radical.

To limit the size of the generated reaction network, severe conditions are implemented of which only the most prominent ones are stated. For the products, only molecules with maximum 8 heavy atoms are allowed in the reaction pool and the maximum amount of oxygen atoms in a molecule amounts to 2. The latter constraint prohibits propagation of the chain branching reactions which are therefore separately included in the final model. The maximal amount of double carbon-carbon bonds amounts to 1 which does not allow benzene formation. Future work encompasses addition of benzene formation pathways since these are important for the high-temperature oxidation region. As discussed in the methodology chapter, section 3.1.6.2, the generated Genesys model is merged with the AramcoMech 2.0 base model to obtain the final model which is validated.

4.3.2. New introduced GAVs for -OOH and -OO•

Group additive values for the Benson calculation scheme are calculated for the hydroperoxide and peroxide radical group, *i.e.* -OOH and -OO•, based on the thermodynamic properties of the cyclohexyl hydroperoxide and the cyclohexyl peroxide radical, the most stable geometry, *i.e.* the equatorial positioned groups, are taken (see section 4.4.1). The values are calculated considering the GAVs which are already available for the carbon atoms and with the ring strain corrections tabulated for a six ring, the outcome is tabulated in Table 4-3. Though, this approach is not optimal since GAVs should consider a broad set of training data from which the GAV values for a functional group is regressed. Future work encompasses a regression based on more molecules with the named functionalities for which accurate thermodynamic data is available.

Table 4-3: Calculated GAV values for the hydrogen peroxide functional group and the cyclohexyl peroxide radical.

STRUCTURE	$\Delta_f H^\circ$ (298 K)	S° (298 K)	C_p (300 K)	C_p (400 K)	C_p (500 K)	C_p (600 K)	C_p (800 K)	C_p (1000 K)	C_p (1500 K)
-OOH	-82.03	142.21	38.53	43.75	47.38	49.83	53.11	55.22	58.25
-OO•	56.58	138.18	29.13	32.5	34.55	35.83	37.59	38.74	40.20

The newly regressed GAVs are implemented in the Benson GAV database of Paraskevas. These are now used to calculate the thermodynamic properties of molecules containing a hydroperoxide or peroxide radical group when the species thermodynamic parameters are not available in the *ab initio* database. Several other non-hydrogen atoms could also not be assigned a GAV due to missing groups

in the database. The latter was resolved by approximating the contributions for these groups by similar groups. For example, the group with a carbon atom connected to two oxygen atoms with valence two, another carbon atom with valence four and a hydrogen atom is missing, which is approximated with the GAVs of the group with one oxygen atom with valence two, two carbon atoms with valence four and a hydrogen atom, as depicted in . A total of 19 “new” GAVs were necessary to allow proper calculation of thermodynamic parameters for all species. These approximations will introduce an error in the thermodynamic parameter calculation for the associated species, but since thermodynamic parameters are determined via ab initio calculations for most important species, it is considered to be of lesser importance for the performance of the final model.

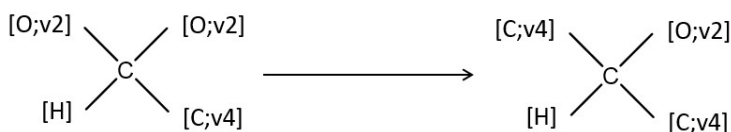


Figure 4-4: Missing GAVs for the structural group left replaced with GAVs for the structural group right.

4.4. Ab initio calculations results

Ab initio calculations have been performed at the CBS-QB3 level of theory to determine both thermodynamic and kinetic parameters. The results obtained from these calculations are discussed in this section.

4.4.1. Thermodynamic parameters

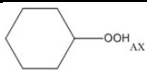
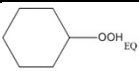
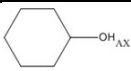
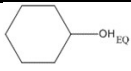
For cyclohexane both axial and equatorial positions are possible for hydrogen atoms and for single-bonded substituents. In case of double bonded substituents, the carbon atom is sp^2 -hybridized causing the spatial arrangement to be planar which makes axial and equatorial position occupation impossible. Triple bonded substituents are impossible for cyclic compounds. Only the cyclic compounds with axial and equatorial positions are discussed in this section.

Several important species important during the oxidation of cyclohexane, all possible positions are optimized, and thermodynamic parameters have been calculated by Genesys. These are discussed next. All ab initio calculated thermodynamic parameters, *i.e.* $\Delta_f H_{298 K}$, $S_{298 K}$, C_p at 300, 400, 500, 600, 800, 1000 and 1500 K, are tabulated in Appendix B. Conformers with equatorial and axial positioned substituents are tabulated as different molecules. For all cases, the thermodynamic parameters of the most stable geometry have been considered for the microkinetic model with correction of the entropy to account for the presence of the other geometries.

4.4.1.1 Single substituted six-membered ring structures

Several single substituted six-membered ring structures have been optimized, *e.g.* cyclohexanone, cyclohexanol, *etc.* Only when it concerns a single-bonded constituent, it is possible to have both axial and equatorial positions, consider for example cyclohexyl hydroperoxide depicted in Table 4-4.

Table 4-4: Enthalpy of formation and entropy at 298K for cyclohexyl hydroperoxide and cyclohexanol

	Cyclohexyl hydroperoxide		Cyclohexanol	
				
$\Delta_f H_{298 K} [\text{kJ mol}^{-1}]$	-212.83	-213.12	-267.58	-268.72
$S_{298 K} [\text{J mol}^{-1} \text{K}^{-1}]$	373.53	378.41	345.08	347.99

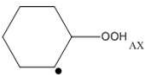
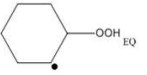
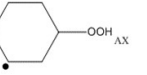
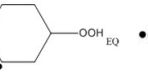
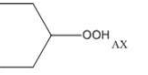
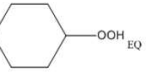
For both molecules, the equatorial position has a slightly more negative $\Delta_f H_{298 K}$, respectively 0.29 kJ mol^{-1} and 1.14 kJ mol^{-1} for cyclohexyl hydroperoxide and cyclohexanol. The equatorial position is enthalpically more favorable due to less interactions with surrounding hydrogen atoms. Entropically,

also the equatorial positions are preferred for the same spatial reason. In the equatorial position the hydroperoxide and hydroxyl groups have more possibilities to position themselves, *i.e.* more randomness, without hindrance from the hydrogen atoms compared to the axial position.

The latter observation is generally valid for all single substituted six-ring species which have been optimized. The equatorial position is enthalpically and entropically more favorable compared to the axial position for a six-membered ring which does not contain double bonds and radical positions. The equatorial position results in less steric hindrance with the neighboring hydrogen atoms. However, the difference for $\Delta_f H_{298 K}$ is limited to only 0 - 2 kJ mol⁻¹. The difference in entropy can reach up to 5 J mol⁻¹ K⁻¹ for the calculated species and for the heat capacities almost identical values are observed. In the future, for this case the need to calculate both species can be omitted and only the equatorial positioned geometry should be considered for the single substituted species without radicals and double bonds in the ring structure. Since the difference in thermodynamic parameters remains within a margin of 4 kJ mol⁻¹ for the enthalpy of formation, and the entropy and heat capacities do not deviate much neither it would not lead to big errors if the value of the axial positioned geometry is used for kinetic model construction.

When double bonds or radical species are present in the six-membered ring, the previous described trend is not valid anymore. For example, consider the different isomers of the cyclohexyl peroxide radical listed in Table 4-5. For CHHP_RAD1, it is calculated that the enthalpy of formation of the axial position is 2.52 kJ mol⁻¹ more favorable compared to the equatorial position. The hydroperoxide group is positioned in a way such that the hydrogen atom and oxygen atoms of the group do have as less influence of the radical and neighboring hydrogen atoms as possible. For the CHHP_RAD1 isomer, it is the axial position which is most suited to position the hydroperoxide group in such a way. When the CHHP_RAD2 isomer is considered, it is observed that the equatorial position is enthalpically the most favorable one. For the CHHP_RAD3 isomer, the influence of the radical on the difference in enthalpy of both geometries is less pronounced since the radical is positioned further away of the substituent. A difference in enthalpy of 0.53 kJ mol⁻¹ is obtained compared to 2.52 kJ mol⁻¹ and 1.98 kJ mol⁻¹ for the other isomers, while the difference amounts to 0.31 kJ mol⁻¹ for the non-radical species, *i.e.* cyclohexyl hydroperoxide. The relative energy difference between the different isomers is again limited within 4 kJ mol⁻¹ in this case. Entropically it turns out that for all isomers the equatorial position is still most favorable. Entropy and enthalpy thus do not always favor the same geometry.

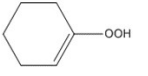
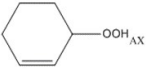
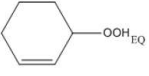
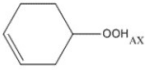
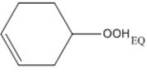
Table 4-5: Enthalpy of formation at 298K for several isomers of the cyclohexyl peroxide radical.

	CHHP_RAD1		CHHP_RAD2		CHHP_RAD3	
						
$\Delta_f H_{298 K}$ [kJ mol ⁻¹]	-10.34	-7.82	-12.90	-14.88	-13.63	-13.10

Though, the observations described for the cyclohexyl peroxide isomers are not generally valid. If the isomers of the cyclohexyl oxy radical are considered, it is always the equatorial position which is enthalpically most stable, but the enthalpy differences are much less pronounced, *i.e.* only 0.02 kJ mol⁻¹, 0.88 kJ mol⁻¹ and 0.17 kJ mol⁻¹ for position of the radical on the first, second and third labelled carbon atom. Enthalpically there is no trend visible in which position is most favorable, but the equatorial position has always the largest value for the entropy. The enthalpy difference of both axial and equatorial positioned geometries remains within 4 kJ mol⁻¹. Note that the amount of species with radicals in the six-membered ring for which thermodynamic parameters have been calculated is limited and thus general conclusions would be unreliable.

The presence of double bonds in the six-ring is also considered, *e.g.* the cyclohexene molecules substituted with a hydroperoxide group listed in Table 4-6. The first isomer where the double bond is present between the carbon atoms labelled with 0 and 1, is the most stable of the considered isomers. For the 12 position, the axial geometry is 4.49 kJ mol⁻¹ more stable enthalpically than the equatorial, which is a significant difference in energy compared to the previous cases. For the 23 position, it is observed that the enthalpy difference is again less pronounced since the double bond is positioned further off and the equatorial position is enthalpically more stable. Entropically it turns out that again the equatorial position is always favorable.

Table 4-6: Enthalpy of formation at 298 K for different isomers of cyclohexene molecules substituted with a hydroperoxide group.

	01 Position	12 Position		23 Position	
					
$\Delta_f H_{298 K}$ [kJ mol ⁻¹]	-95.08	-89.51	-85.02	-91.43	-91.87

A general trend in the enthalpically most favorable position of single substituted six ring molecules is not observed. However, when a six-membered ring is considered with no radical carbon atoms and only single carbon-carbon bonds in the ring, the equatorial position is enthalpically most favorable due to less interaction with neighboring hydrogen atoms. The energy difference in this case remains limited

for the axial and equatorial position within 1 to 2 kJ mol⁻¹. When radicals or double bonds are present in the six ring this trend is not valid anymore and there is also no trend visible that a certain position of the radical/double bond is always the most stable one. Most of the time the difference in energy between the equatorial and axial position is limited, but several exceptions exist where it can reach beyond the boundary limit of accuracy, *i.e.* 4 kJ mol⁻¹. Entropically the equatorial position is in almost all cases the most favorable position.

4.4.1.2 Double substituted six-membered ring structures

Thermodynamic parameters have been calculated for double substituted six-membered ring structures, which are for example important for the chain branching reactions in the low temperature oxidation zone. Examples of such molecules are cyclohexyl hydroperoxide molecules with an extra substituted peroxide radical group, only the results of these isomers are discussed although several others are listed in Appendix B. When multiple single-bonded functional groups are present, these can both be positioned equatorial and axial, which makes that the number of possible geometries increases. Note that for these molecules, enantiomers originate when a chiral carbon atom is present in the molecule.

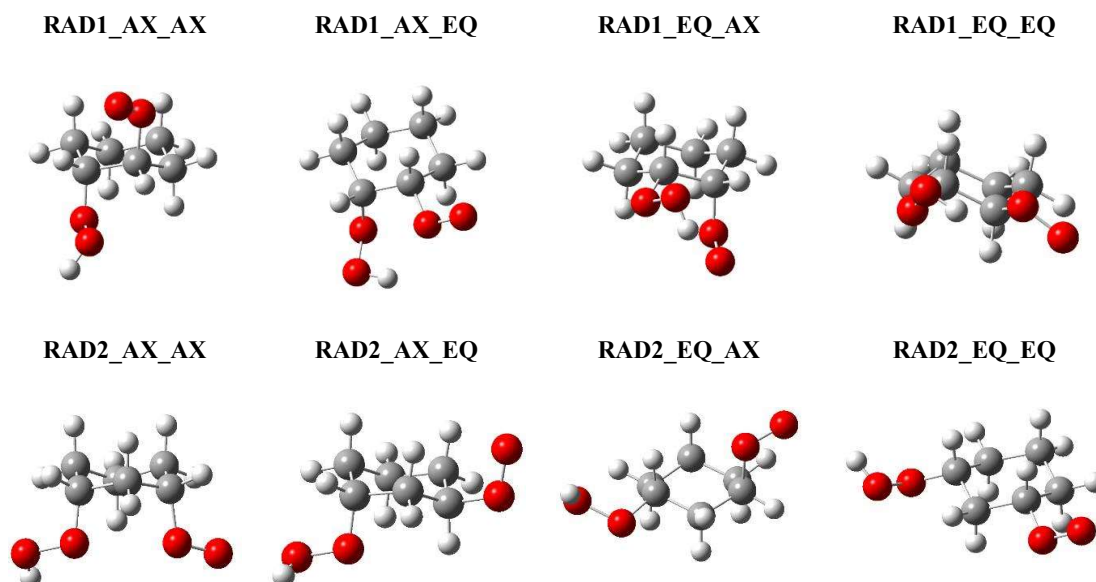
In Table 4-7, the different isomers of the cyclohexyl peroxide molecule with a substituted peroxide radical group are listed with the corresponding enthalpy of formation and entropy at 298 K. For this group of species, it is possible to have interactions between the functional groups which makes that positioning the groups most far away from each other does not automatically result in the most stable geometry.

Table 4-7: Enthalpy of formation and entropy at 298 K for different geometries of cyclohexyl peroxide with a substituted peroxide radical group (chemical formula: C₆H₁₁O₄) on the 1, 2 and 3 labelled carbon position.

	RAD1_AX_AX	RAD1_AX_EQ	RAD1_EQ_AX	RAD1_EQ_EQ
$\Delta_f H_{298\text{ K}}$ [kJ mol ⁻¹]	-157.74	-155.36	-155.92	-155.82
$S_{298\text{ K}}$ [J mol ⁻¹ K ⁻¹]	436.79	422.64	422.19	429.41
	RAD2_AX_AX	RAD2_AX_EQ	RAD2_EQ_AX	RAD2_EQ_EQ
$\Delta_f H_{298\text{ K}}$ [kJ mol ⁻¹]	-170.79	-164.48	-164.01	-161.85
$S_{298\text{ K}}$ [J mol ⁻¹ K ⁻¹]	437.59	455.09	454.76	448.83

	RAD3_AX_AX	RAD3_AX_EQ	RAD3_EQ_AX	RAD3_EQ_EQ
$\Delta_f H_{298\text{ K}}$ [kJ mol ⁻¹]	-	-165.01	-163.29	-160.81
$S_{298\text{ K}}$ [J mol ⁻¹ K ⁻¹]	-	434.95	439.30	443.30

For the first isomer with two functional groups, the geometry with both groups positioned axial is enthalpically and entropically the most favorable, but the difference in enthalpy of formation with the other geometries is limited within 2.38 kJ mol⁻¹. The maximal difference in entropy amounts to 14.60 J mol⁻¹ K⁻¹. The spatial arrangements of the different geometries are depicted in Figure 4-5. For RAD1_AX_AX, the hydrogen atom of the hydroperoxide group is positioned away from the oxygen radical, while the radical oxygen atom is positioned in such a way to interact with the oxygen atom of the hydroperoxide group and the neighboring hydrogen atoms. Although, RAD1_AX_EQ and RAD1_EQ_AX or not equal geometries, the thermodynamic properties are very similar. The difference in geometry is visible in the spatial arrangement where for the RAD1_AX_EQ geometry the hydrogen atom interacts with the non-radical oxygen atom and in the RAD1_EQ_AX case it interacts with the radical oxygen atom. Due to the chiral center two enantiomers are present for this isomer. This makes that RAD1_AX_AX and RAD1_EQ_EQ can go over in each other when the six-ring flips, similar for RAD1_AX_EQ and RAD1_EQ_AX, but *e.g.* RAD1_AX_AX cannot flip into RAD1_AX_EQ or the other way around.



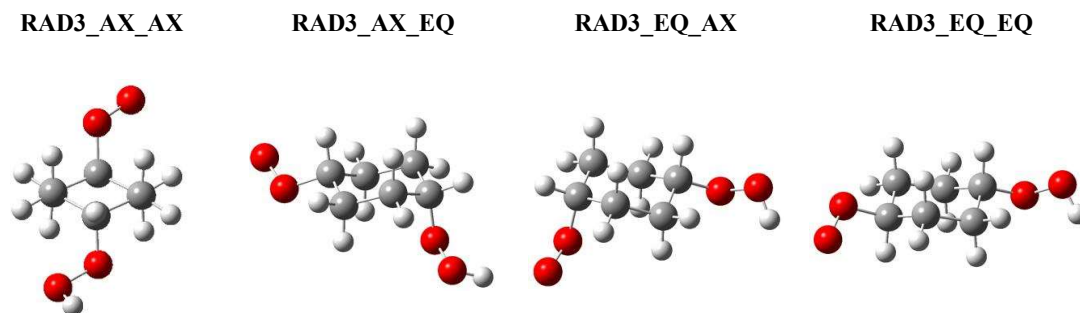


Figure 4-5: Spatial arrangement of the different geometries of the isomers of $C_6H_{11}O_4$.

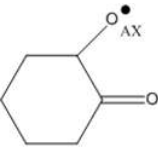
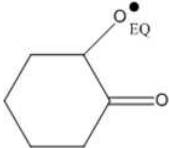
For the fourth geometry, *i.e.* RAD1_EQ_EQ, the hydrogen atom forms again a hydrogen bond with the non-radical oxygen atom and the radical oxygen atom is positioned away. For the second isomer, *i.e.* the RAD2 geometries, the difference in the enthalpy of formation is maximal between the RAD2_AX_AX, *i.e.* the most stable geometry, and RAD2_EQ_EQ which amounts to 8.94 kJ mol^{-1} . This is already a significant difference, which can already result to inaccuracy if only the RAD2_EQ_EQ geometry would be considered since it reaches above the accuracy limit of 4 kJ mol^{-1} . Entropically the combination of an equatorial and axial positioned group is most favorable. From the spatial rearrangement it is observed that both groups do still influence each other. For both groups positioned axial or equatorial, the hydrogen atom is positioned away from the peroxide group, while for the other two geometries the hydrogen atom is positioned towards the peroxide group.

For the final isomer, with groups positioned on the carbon atoms labelled 0 and 3, it was not possible to gain convergence of the RAD3_AX_AX geometry on the CBS-QB3 level of theory, but it converges on the B3LYP level of theory with the 6-31G(d) basis set from which the spatial arrangement is shown. Due to missing data it is difficult to state which geometry is most stable, but for the known data it is the AX_EX geometry. Entropically the EQ_EQ geometry is most favorable, although it can be expected that the AX_AX will have a higher entropy since the groups have for this case more options to move due to less interaction with neighbor atoms.

In Appendix B several cases are observed for which the difference in enthalpy of formation reaches beyond 10 kJ mol^{-1} , however these are characterized by attractive or repulsive interactions which appear between the substituted groups. For example, consider the different axial and equatorial positioned geometries of 2-oxycyclohexan-1-olate in Table 4-8 for which a significant difference of 18.9 kJ mol^{-1} is detected. The axial position is favored since it creates less repulsive interactions between the carbonyl and radical oxy group. Similarly, the equatorial or axial position of a group can favor the formation of attractive interactions such as hydrogen bonds. Hydrogen bonds decrease the enthalpy of formation, but in addition the entropy also decreases due to more order. The latter is for example observed for the axial

and equatorial geometry of 2-hydroperoxycyclohexan-1-one which forms a hydrogen bond in the axial position.

Table 4-8: Enthalpy of formation at 298 K for an axial and equatorial positioned oxy radical group with significant energy differences.

	Axial	Equatorial
		
$\Delta_f H_{298\text{ K}} [\text{kJ mol}^{-1}]$	-164.74	-145.87

The thermodynamic parameters calculated for the different geometries of double substituted do not show systematic trends to predict beforehand which geometry is enthalpically most favorable. In addition, the difference in energy can be quite big, which requires that all geometries are optimized to secure determination of the lowest energy conformer essential for thermodynamic parameter assignment.

4.4.2. Hydrogen abstractions

In Table 4-9, the modified Arrhenius parameters are listed for the hydrogen abstraction by several species, i.e. molecular oxygen and the hydrogen, hydroperoxyl, formyl and methyl radical, on both the axial and equatorial position of cyclohexane. All rates are listed as if it is the only hydrogen abstraction which could occur, i.e. the number of single events amounts to 12 (except for the case with oxygen and the methyl radical where it amounts to 24). These are compared with the hydrogen abstraction rate as if the abstraction would be modelled with the reaction rate obtained from rate rules, numeric values obtained from the references in Table 4-1. This way it is possible to investigate the influence of the ring structure on the kinetics. The number of single events is thus also taken to be 12 to allow a good comparison. No transition state was found for hydrogen abstraction by a hydroxyl radical, unfortunately, since it forms an important reaction pathway, especially in the low-temperature region. Note that several other hydrogen abstraction reactions have been tried, *e.g.* hydrogen abstraction by the cyclohexyl peroxide radical which is thought to be important in the liquid phase. However, the latter transition state contains 14 heavy atoms and does not converge at the CBS-QB3 level of theory within the maximal possible wall time of the Ghent University high-performance computing infrastructure. The reaction rate for the initiation reaction, *i.e.* hydrogen abstraction by molecular oxygen, is currently still calculated with kinetic parameters from a rate rule since the acquired *ab initio* reaction rate parameters are wrongly calculated with a singlet state instead of a triplet.

Table 4-9: Modified Arrhenius parameters for hydrogen abstractions from cyclohexane and propane.

Reacting species		Log A	n	B	reaction rate coefficient (1000 K) [$\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$]
O₂	Rate Rule	-	-	-	$1.87 \cdot 10^{-4}$
H	Cyclohexane Axial	-1.346	3.017	17.661	$6.69 \cdot 10^6$
	Cyclohexane Equatorial	-1.478	3.065	19.128	$5.82 \cdot 10^6$
	Rate Rule	-	-	-	$7.53 \cdot 10^7$
HO₂	Cyclohexane Axial	-14.097	6.490	26.681	$1.32 \cdot 10^4$
	Cyclohexane Equatorial	-13.631	6.554	28.841	$4.67 \cdot 10^4$
	Rate Rule	-	-	-	$6.82 \cdot 10^3$
CHO	Cyclohexane Axial	-9.572	5.388	52.265	$5.91 \cdot 10^3$
	Cyclohexane Equatorial	-9.189	5.412	53.456	$2.23 \cdot 10^4$
	Rate Rule	-	-	-	$1.29 \cdot 10^3$
CH₃	Cyclohexane Axial	-10.914	5.765	16.950	$4.17 \cdot 10^5$
	Cyclohexane Equatorial	-12.096	5.803	17.690	$3.27 \cdot 10^4$
	Rate Rule	-	-	-	$4.95 \cdot 10^5$

A general trend which favors the abstraction from the axial or equatorial position is not observed. However, steric hindrance is expected for axial abstraction which is therefore less favorable for bigger molecules, *e.g.* the formyl radical and the hydroperoxyl radical. For smaller molecules where steric hindrance is less important, abstraction from the axial position is faster and thus more favorable since the planar structure of the cyclohexyl radical is immediately formed, *e.g.* visible for the hydrogen and methyl radical. Comparing reaction rates from *ab initio* results with the ones obtained via rate rules indicates that the magnitude is well, but deviations even with a factor 10 are observed for the abstraction by a hydrogen radical. For Genesys, the modified Arrhenius parameters of the fastest reaction is always included.

4.4.3. Energy scans

For several reactions it is not possible to obtain convergence of the transition state, typically for transition states characterized with none or a very small activation energy. For this reason, an energy scan is carried out in which the bond between two atoms is systematically broken by increasing the interatomic distance in small steps. Moreover, an energy profile is constructed for cyclohexane which shows the difference in energy when the six-membered ring variates between the different conformers.

4.4.3.1 Addition of molecular oxygen to the cyclohexyl radical

Calculations for the transition state of the addition of molecular oxygen to the cyclohexyl radical do not converge. This is a common problem for several oxygen addition reactions since it has no barrier and as such there exists no transition state. Therefore, a scan has been performed at the B3LYP level of theory with the 6-31G(D) basis set, depicted in Figure 4-6, which systematically increases the bond length

between the carbon and oxygen atom both for axial and equatorial cyclohexyl peroxide. The axial and equatorial high-level optimized geometry have initially a bond length of respectively 1.48 Å and 1.47 Å with the equatorial species being 0.7 kJ mol⁻¹ more stable. When the length is increased it is observed that the species with equatorial substituent increases faster in comparison with the axial substituent. When the C-O length of the axial substituent increases, the planar structure of the cyclohexyl radical is already formed which makes it more favorable. Also notice the drop in energy around 1.88 Å which is due to rotation of the oxygen molecule. In the end both evolve to the same energy state with two species, *i.e.* the cyclohexyl radical and molecular oxygen. It is thus clear that there is no energy barrier during the addition and a transition state thus cannot be found.

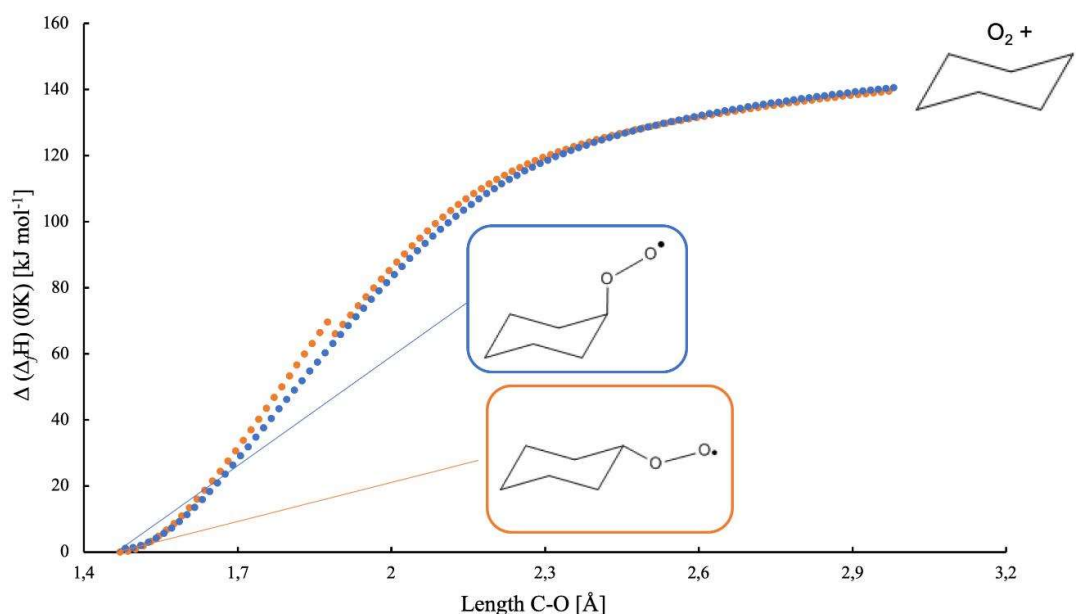


Figure 4-6: Energy profile at the B3LYP level of theory with the 6-31G(d) basis set for the reverse reaction of addition of molecular oxygen to the cyclohexyl radical with an axial (blue) and equatorial (orange) positioned peroxide group, values are relative to the enthalpy of formation at 0 K of the equatorial positioned cyclohexyl peroxide.

4.4.3.2 Cyclohexyl hydroperoxide oxygen-oxygen scission

Similar to the addition of molecular oxygen on the cyclohexyl radical, the cyclohexyl hydroperoxide decomposition is difficult to calculate due to lack of a transition state. An energy scan is carried out with both an axial and equatorial positioned hydroperoxide group which decomposes from which no activation energy is observed. The difference in energy of the optimized axial and equatorial cyclohexyl hydroperoxide only amounts 0.3 kJ mol⁻¹ in which the equatorial is more stable due to less steric hindrance. Similar for the decomposed products, the equatorial positioned group is 2.75 kJ mol⁻¹ more stable. The total energy difference between cyclohexyl peroxide and the decomposed products, *i.e.* the cyclohexyloxy and hydroxyl radical, amounts to respectively 190 kJ mol⁻¹ and 193 kJ mol⁻¹ for the axial and equatorial geometry.

When no restrictions are performed, the just released hydroxyl radical immediately reacts further and abstracts a hydrogen atom of the neighbor carbon atom and simultaneously the six-membered ring breaks open and forms a five-membered ring with an attached carbonyl group. The final products formed are water and cyclopentane carboxaldehyde as depicted in Figure 4-7 and the corresponding heat release amounts to $-214.2 \text{ kJ mol}^{-1}$. Currently this reaction is not further investigated, but when experiments would indicate the presence of five-membered ring structures this could be the origin.

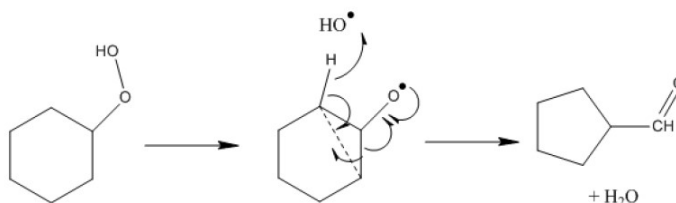


Figure 4-7: Rearrangement of the six ring to form a five ring starting from scission of the cyclohexyl hydroperoxide species.

4.4.3.3 Cyclohexane conformer search

Cyclohexane is known to have different conformers of which the boat and chair conformer are the most-known. An energy profile, at B3LYP level of theory with basis set 6-31G(D) in which the dihedral angle between the carbon atoms is changed, is constructed to investigate the difference in energy between the different conformers, *cf.* Figure 4-8. The chair conformer is the most stable conformer and by default most cyclohexane species with substituents are present in this state, however in transition states it happens that the boat conformer is more stable, *e.g.* for the formation of a 5-ring cyclic ether. A difference in energy of 30.6 kJ mol^{-1} is calculated boat versus chair. In this work the interchange between boat and chair conformers is not considered for the thermodynamic properties for any species. Observe that the chair conformer is a kind of transition state situated between two twisted boat conformers which are 3.5 kJ mol^{-1} more stable. Thence, it is better to state that the most-common conformers of cyclohexane are the chair and twisted-boat conformer rather than the chair and boat. When evolving from chair to twisted boat or vice versa the half chair has to be passed which has an energy of 47.8 kJ mol^{-1} . Calculation of the energy of the twisted boat structure at the CBS-QB3 level of theory confirms the energy difference of 27.1 kJ mol^{-1} , which is also reported by Sirjean et al. [35].

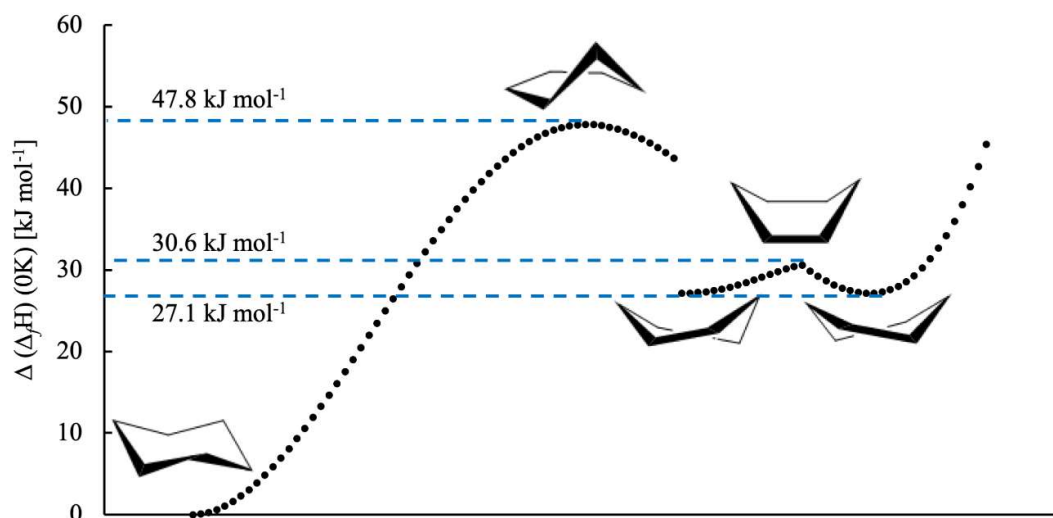


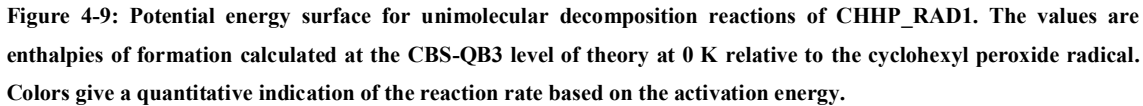
Figure 4-8: Energy profile with the different conformers of cyclohexane from left to right: chair, half chair, twisted boat, boat and twisted boat, values are enthalpies of formation at 0 K at the B3LYP level of theory with the 6-31G(d) basis set relative to the cyclohexane chair conformer.

4.4.4. Potential energy surface

The oxidation of cyclohexane is initiated by the hydrogen abstraction by molecular oxygen forming a cyclohexyl radical and a hydroperoxyl radical. Subsequently, molecular oxygen adds to the cyclohexyl radical to form the cyclohexyl peroxide radical associated with a heat of reaction equal to 156 kJ mol⁻¹. Several decomposition reactions are investigated for this radical, but especially the isomerization reactions by hydrogen shifts are of interest for this radical. Three radicals can be formed depending on the number of atoms in the transition state, CHHP_RAD1, CHHP_RAD2 and CHHP_RAD3 are formed for respectively the 1,4-hydrogen shift, the 1,5-hydrogen shift and the 1,6-hydrogen shift, nomenclature indicated in . The consecutive chemistry is investigated for each of these radicals by construction of a potential energy surface (PES) relative to the enthalpy of formation of the cyclohexyl peroxide radical at 0 K. For CHHP_RAD1 this is depicted in Figure 4-9. Colors give a quantitative indication on the reaction rate based on the activation energy.

Table 4-10: Nomenclature used for the different isomers of the cyclohexyl peroxide radical.

CHHP_RAD1	CHHP_RAD2	CHHP_RAD3



As observed from the PES the 1,4-hydrogen shift to form the CHHP_RAD1 has an activation energy of 7 kJ mol⁻¹ above the reverse addition reaction of molecular oxygen which makes the formation of CHHP_RAD1 unfavorable and unlikely in the low-temperature oxidation region. However, when the CHHP_RAD1 species would be formed, unimolecular decomposition is considered by carbon-carbon β -scission, hydrogen-carbon β -scission, several hydrogen shifts, cyclic ether formation and by dissociation of the hydroperoxyl radical for formation of cyclohexene. These transition states are always treated for the axial positioned and equatorial positioned geometry, but these are only discussed if the reaction rate are remarkably different.

Cyclic ether formation with formation of an extra three-membered ring can proceed with the formation of the cis and trans isomer. The transition state of the cis form amounts to 51 kJ mol⁻¹ and results in a decrease of enthalpy of 70 kJ mol⁻¹ compared to the CHHP_RAD1 species. Compared with the dissociation reaction to oxygen and the cyclohexyl radical, this pathway is favorable and can be expected to be dominant for decomposition of the CHHP_RAD1. The formation of a three-membered ring for the trans form however requires a significant higher activation energy, *i.e.* 156 kJ mol⁻¹ instead of 51 kJ mol⁻¹. In addition, the formed trans geometry is enthalpically less favorable compared to the starting products, which makes this reaction pathway unlikely to proceed.

A second reaction pathway for decomposition is the scission of the carbon-oxygen bond to form the cyclohexene and hydroperoxyl radical with an activation energy of 65 kJ mol⁻¹ which is limited. The products formed are more stable compared to the cyclohexyl and molecular oxygen, for which an enthalpy of reaction at 0 K equal to 66 kJ mol⁻¹ is calculated.

Hydrogen shifts were expected to be of lesser importance which is partly confirmed. Hydrogen shift from the carbon atom labelled with number 2 leads to the formation of the CHHP_RAD2 species, which is enthalpically slightly favorable, *i.e.* 4 kJ mol⁻¹, but the activation energy is too high to be considered as an important reaction path. On the other hand, hydrogen shift with the 0 labelled carbon atom is immediately followed by the rearrangement to cyclohexanone and the hydroxyl radical due to scission of the oxygen-oxygen bond. The low enthalpy of formation of cyclohexanone makes this reaction path the most favorable on the PES what concerns the enthalpy of reaction. However, the reaction barrier is quite large with 122 kJ mol⁻¹ making this reaction less likely to proceed.

Moreover, two extra pathways are considered for decomposition which are carbon-carbon and hydrogen-carbon β -scission with the formation of a double carbon-carbon bond. Different products are obtained depending on the bond which is broken, but none of these is substantially favored by a lower activation energy or a higher enthalpy difference between products and reactants. Due to the minimal activation energy of 140 kJ mol⁻¹ for hydrogen-carbon β -scission these reaction pathways are unlikely to proceed, especially since the formed products are only slightly more favorable than the corresponding

transition states which are positioned higher than the starting products. The dissociation reaction is thus expected to proceed faster than the β -scissions. In case of carbon-carbon β -scission, the ring is broken and acyclic oxygenated hydrocarbons are formed which increases the entropy. Although they are enthalpically unfavorable these could still proceed due to entropic reasons. Because of the high activation barrier, *i.e.* at least 120 kJ mol^{-1} , and the high energy of the formed products these reactions are considered to be of minor importance in the low-temperature oxidation zone for the CHHP_RAD1 species.

Comparing the rate of the corresponding equations for all considered reactions indicates that the positioning of the hydroperoxide group does have an influence on the stability of the transition state and therefore influences the reaction rate. For example, the hydrogen shift for formation of the CHHP_RAD2 for which the axial positioned hydroperoxide group is 15 kJ mol^{-1} more stable compared to the equatorial positioned group. In the low temperature region these differences make that the coefficients can differ to a factor ten.

Similarly, potential energy surfaces have been constructed for the CHHP_RAD2 and CHHP_RAD3 which consider the same unimolecular decomposition reactions. However, several differences are observed when the different surfaces are compared. In Figure 4-10 the relevant part of the potential energy surface is constructed for unimolecular decomposition of the cyclohexyl peroxide radical which merges all reaction pathways with an activation energy below 180 kJ mol^{-1} relative to the cyclohexyl peroxide radical. In addition, several other reactions are included by which the formed cyclohexyl peroxide radical can decompose.

The cyclohexyl peroxide radical can decompose with formation of cyclohexene and the hydroperoxyl radical with an activation energy of 134 kJ mol^{-1} . In the transition state simultaneously the carbon-oxygen and carbon-hydrogen bond is broken. This reaction is energetically more favorable compared to the dissociation reaction forming molecular oxygen and cyclohexane.

A 1,3-hydrogen shift leads to the formation of stable cyclohexanone and the hydroxyl radical accompanied with a heat of reaction of -113 kJ mol^{-1} starting from the cyclohexyl peroxide radical. The high energy barrier of 165 kJ mol^{-1} makes this reaction path less favorable.

It is clear that the peroxide radical can be converted to a hydroperoxide by a hydrogen shift reaction. The difference in activation energy between these different transition states is remarkably. The energy of the formed peroxides, *i.e.* of CHHP_RAD1, CHHP_RAD2 and CHHP_RAD3, is similar but the barrier for the 1,4-shift, 1,5-shift and 1-6 shift amount to 156 kJ mol^{-1} , 108 kJ mol^{-1} and 129 kJ mol^{-1} respectively at 0 K. For the 1,4-shift and 1,5-shift this is below the reverse dissociation to molecular oxygen and the cyclohexyl radical for which an enthalpy difference between reactants and products of 156 kJ mol^{-1} is calculated. It can be expected that the 1,5-shift will also occur because it is entropically

favorable. It is thus the CHHP_RAD2 which will be preferentially formed and react further by β -scission or chain branching, discussed next. The favored β -scission of CHHP_RAD2 leads to the formation of 5-hexenal and the hydroxyl radical with an activation energy of 162 kJ mol⁻¹. Another remarkable difference is the formation of the cyclic ether from the CHHP_RAD3 radical which has an activation energy of only 71 kJ mol⁻¹ starting from CHHP_RAD3 and an accompanied reaction heat of 123 kJ mol⁻¹, this makes it the most favorable reaction path for the CHHP_RAD3 radical.

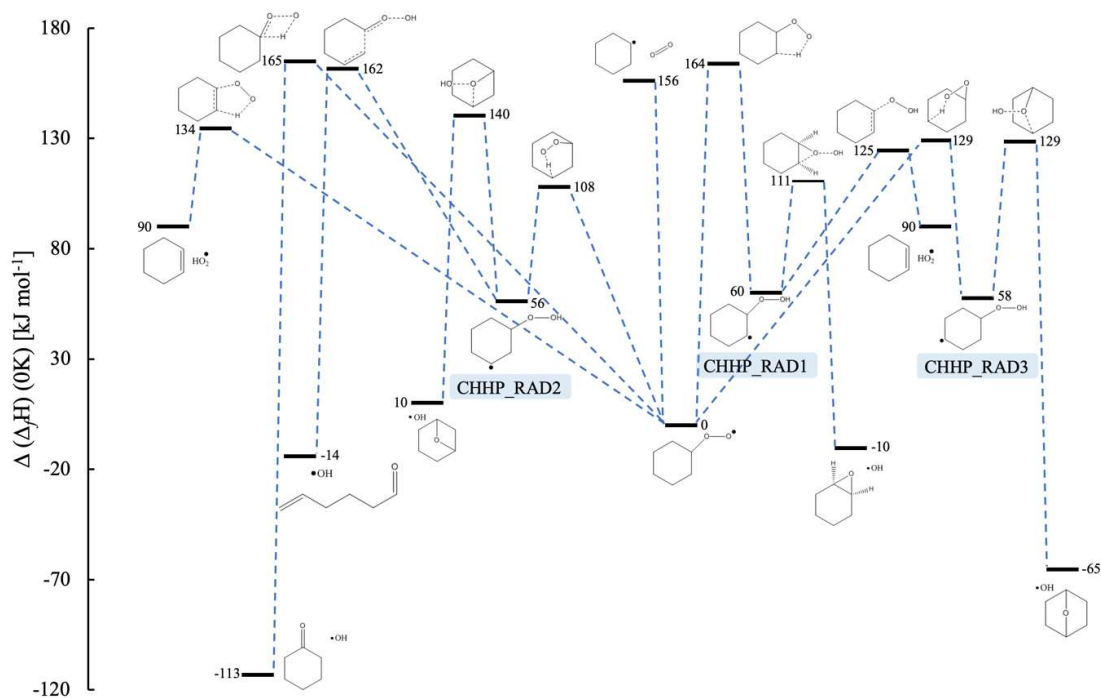


Figure 4-10: Potential energy surface for unimolecular decomposition of cyclohexyl peroxide radical.

Once the hydrogen-shift has proceeded to form CHHP_RAD1, CHHP_RAD2 or CHHP_RAD3 there is the possibility of molecular oxygen to add to these molecules which have an enthalpy of reaction of respectively -148 kJ mol⁻¹, -146 kJ mol⁻¹ and -156 kJ mol⁻¹. The resulting molecules can chain branch, this forms an important reaction pathway to have an increase of radicals in the low-temperature oxidation zone which results in significant conversion at these temperatures. Due to the importance of these reactions, these were also considered for ab initio calculations. The results are shown in the reaction scheme in Figure 4-11 for the CHHP_RAD1 species with indication of the most stable position of each substituent (Ax. or Eq.). Since it concerns a sequence of non-unimolecular reactions, these cannot be depicted on a PES.

After the addition of molecular oxygen, the formed species has a peroxide and hydroperoxide functionality. The peroxide group will proceed via a hydrogen shift to the associated ketone hydroperoxide with release of a hydroxyl radical. An activation energy of 118 kJ mol⁻¹ and a heat of reaction of -186 kJ mol⁻¹ is obtained from ab initio calculations. Via an oxygen-oxygen scission reaction

the ketone oxy species is formed, again with release of a hydroxyl radical. This final species is expected to decompose via a β -scission. Both β -scission have a similar activation energy of 24 kJ mol⁻¹, but the product with a radical on the carbonyl group is preferred due to stabilization. The latter can react further by an α -scission with formation of carbon monoxide and hydroxypentyl.

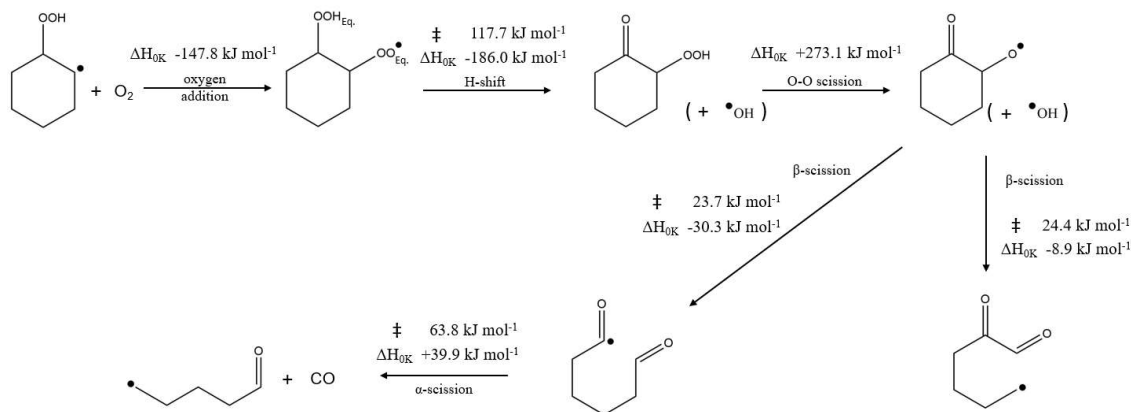


Figure 4-11: Reaction scheme for chain branching of the CHHP_RAD1 species with activation barrier and associated enthalpy of reaction, data obtained from ab initio calculations at the CBS-QB3 level of theory.

When these chain branching reactions proceed, three radicals are formed for every hydroperoxide radical. This indicates the acceleration effect which can be induced by these reactions. Moreover, in literature it is indicated that these reaction are important for low-temperature oxidation, which is confirmed from the model validation in the next section.

4.5. Performance of the developed kinetic model

Validation of the microkinetic model is performed with experiments published in literature which are originally used to assess the microkinetic model developed by Serinyel et al. [13]. The experimental conditions are a pressure of 1.07 atm, a reactor volume of 85 cm³, an equivalence ratio of 1.0 and a molar inlet flow of cyclohexane equal to 0.667 diluted with helium. Reactor simulations are performed with ANSYS chemkin V17.0. When the initial microkinetic model is compared with the experimental trends, these are not well predicted with even no reaction in the low-temperature oxidation zone. However, it was observed that Genesys had difficulties with the regression of modified Arrhenius parameters and the performance of the model was significantly enhanced by using the modified Arrhenius parameters regressed in a temperature region around 700 K. The latter model showed reactivity in the low temperature region but the significant drop in conversion around 650 K is not predicted.

The formation of the CHHP_RAD2 species is most dominant starting from cyclohexyl peroxide due to the lower activation energy and the influence of the entropy. This is also observed from the corresponding reaction rates, but the formed hydroperoxide species did not react further. The most favorable reaction pathway which leads to decomposition of the CHHP_RAD2 species is the β -scission to 5-hexenal and a hydroxyl radical. The initial modified Arrhenius parameters from ab initio calculations amounted to $1.38 \cdot 10^9 \text{ s}^{-1}$, 1.98 and 24.2 kJ mol⁻¹ for respectively the pre-exponential factor, the temperature exponent and the alternative activation energy. Adaptation of the alternative activation energy to 12.2 kJ mol⁻¹ was necessary to obtain the expected drop in conversion as observed by experimental measurements. The experimental measurements and several models are plotted in Figure 4-12.

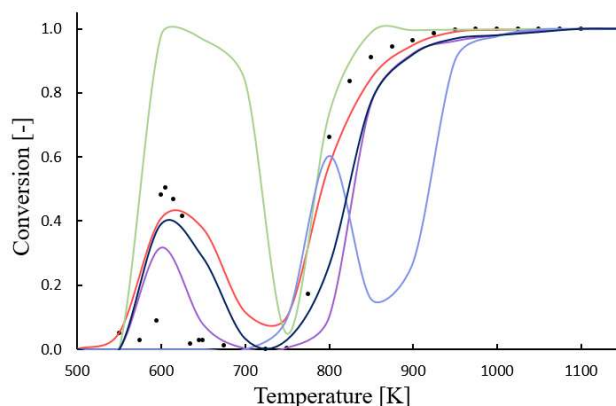


Figure 4-12: Comparison of several different models for the experimental conditions as used for model validation at Nancy: red line Serinyel et al. [13], purple line Ranzi et al. [23], green line JetSurf [21], light blue line Westbrook et al. [25] and dark blue the microkinetic model developed in this master dissertation. Black dots represent the experimental measurements.

From the plot comparing the different models with experimental measurements in Figure 4-12, it can be observed that the initial on-set temperature of the Genesys model developed in this master dissertation does predict the experimental measurements rather good, this is the region where no fitting has performed. The maximal conversion around 610 K and the steep drop in conversion between 650 and 750 K is the result of the adapted alternative activation energy. A conversion of 50 % should be obtained around 610 K which is currently calculated to be 40%. The model as it is thus able to predict the experimental measurements in the low-temperature oxidation zone with only one fitted parameter, which is rather good. Especially, since the other model performing good is the one from Serinyel et al. [13] which uses several lumped reactions. When the reactions important in the low-temperature oxidation region are investigated it turns out that the expected chain branching reactions are responsible for the major radical production and do create conversion in the region around 600 K. Around 610 K, other reactions start to become dominant, for example the β -scission with a fitted parameter, which reduces the generation of radicals and thus decreases the conversion. With higher temperature, these reactions becomes more dominant and the shift is made from chain branching to β -scissions, cyclic ether formations and other reactions which are favorable according to the potential energy surface.

The high-temperature region which was of lesser interest in this master dissertation since other chemistry proceeds in this region which is not observed during the liquid-phase oxidation of cyclohexane. However, also in this region the on-set temperature is well-predicted and in addition only a slight delay is observed in the steepness of the model curve.

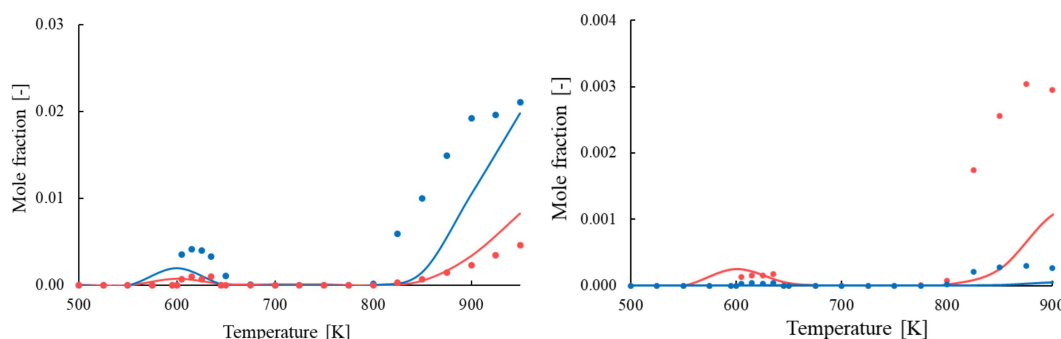


Figure 4-13: Left: carbon monoxide (blue) and carbon dioxide (red) production rate in function of the temperature. Right: ethylene (red) and propylene (blue) production rate in function of the temperature. Lines represent the model predictions and dots represent experimental measurements.

The production rate of several important products from cyclohexane oxidation are depicted in Figure 4-13. It is observed that the low-temperature oxidation zone does fit rather well which is in agreement with the conversion which is due to adaptation of a parameter also well. Especially the formation of carbon dioxide and ethylene is well predicted in the range between 550 and 650 K. However, for the high-temperature region the model predictions are not good which could be expected since the conversion is not good predicted. Future improvements should focus on the high-temperature regime

for which ab initio calculations should be carried out, in addition the other versions of the AramcoMech base mechanism should be investigated if this leads to better performance. Especially since the production of ethylene, propylene, carbon monoxide and carbon dioxide is mainly influenced from the reaction rates of this base mechanism. An in-depth sensitivity analysis and product analysis is not yet performed since the model is not considered as finished but requires some extra work to get a good fit over the whole temperature region. Also the low-temperature oxidation region should be investigated in closer detail to look if several reaction pathways are not overlooked. A better performing model would be one in which no fitting of kinetic parameters is required.

4.6. Conclusions

Before development of a microkinetic model for the liquid-phase oxidation of cyclohexane, a gas-phase model is constructed with Genesys to prove the performance of the developed automatic kinetic model generation framework. The focus is on the low-temperature oxidation region since literature indicates that the chemical reactions important in this temperature range are also important during liquid-phase oxidation of cyclohexane.

The chapter is started with covering the experimental and modelling research which has already been performed on the cyclohexane oxidation. Cyclic compounds, so-called naphthenes, can make up a significant fraction of energy fuels and for combustion optimization it is necessary to have a reliable model describing the oxidation chemistry of naphthenes. Cyclohexane is therefore frequently used as surrogate, due to its high symmetry number, which explains the numerous amounts of published works.

A microkinetic model is constructed with Genesys by using reaction families which are the same for oxidation of acyclic compounds, *e.g.* hydrogen abstraction by molecular oxygen, molecular oxygen addition, hydrogen shifts, cyclic ether formation, chain branching and β -scission reactions. In addition new GAVs have been calculated for the hydroperoxide and peroxide radical group which were missing in Genesys.

To obtain accurate thermodynamic and kinetic parameters, *ab initio* calculations have been performed at the CBS-QB3 level of theory. It was started with optimization of several species important in the low-temperature oxidation region for thermodynamic parameter calculation with Genesys. For single substituted functionalities on the six-membered ring structures all possible spatial arrangements are treated, *i.e.* an axial position and an equatorial position is considered, in order to find the most stable geometry. It turns out that for single substituted six-membered ring structures, without radicals and double bonds in the six ring, the equatorial position is always favored both for the entropy and the enthalpy of formation. The difference in enthalpy of formation remains in the latter case limited within 1 to 2 kJ mol⁻¹. For double substituted six-membered ring structures, no trend is observed in the stability of the different geometries. For these molecules, differences in enthalpy of formation can reach beyond 15 kJ mol⁻¹ since repulsive and attractive interactions can be present for one specific geometry which are not present for the other geometry, *e.g.* a hydrogen bond between the substituents. The latter makes that each geometry should be optimized to obtain accurate thermodynamic data of the lowest energy geometry which is not required in case of single substituted species because the error remains much more limited.

The potential energy surface (PES) for the addition of molecular oxygen to the cyclohexyl radical has been constructed by high-level *ab initio* calculations at the CBS-QB3 level of theory to study the low-

temperature oxidation reactivity. From the PES, it is clear that the peroxide radical can be converted to a hydroperoxide by several hydrogen shift reactions, *i.e.* a 1,4-shift, 1,5-shift or 1,6-shift. The associated activation barrier is below the reverse dissociation to molecular oxygen and the cyclohexyl radical except for the 1,4-shift which is therefore less likely. It can be expected that the 1,5-shift will occur faster because it is entropically favored. These formed hydroperoxide radicals can react further mainly through cyclic ether formation and β -scission reactions. Alternatively, oxygen addition on the hydroperoxide radicals leads to chain branching which accelerates the reaction since three radicals are formed starting from only one radical.

Besides the hydrogen shifts, also hydroperoxide elimination can take place with formation of cyclohexene which stays under the reaction enthalpy of dissociation to molecular oxygen and the cyclohexyl radical. The unimolecular decomposition reaction with the lowest reaction enthalpy is a 1,3-shift for which 113 kJ mol^{-1} is released due to the formation of the energetically favorable cyclohexanone and a hydroxyl radical. However, the activation energy of 165 kJ mol^{-1} makes proceeding of this reaction unlikely. Other hydrogen shifts, carbon-carbon and carbon-hydrogen β -scissions do not play an important role except for one β -scission possible after the 1,5-shift which forms the hex-5-enal and hydroxyl radical. This reaction has an energy barrier equal to 162 kJ mol^{-1} with an enthalpy change of -14 kJ mol^{-1} which is in addition more favorable due to ring-opening.

Before reactor simulations, the developed microkinetic model of Genesys is merged with AramcoMech 2.0 to include the chemistry of small (oxygenated) hydrocarbons. The final model is fitted with experiments performed in a jet-stirred reactor at l'Université de Lorraine by the group of Prof. Battin-Leclerc. The model predicts well the experimental trends of the conversion of cyclohexane in the low-temperature oxidation region when the alternative activation energy of the modified Arrhenius equation for the mentioned β -scission is lowered from 24.2 kJ mol^{-1} to 12.2 kJ mol^{-1} . In addition, the formation rate of major products such as ethylene and carbon dioxide in the low-temperature region does also match with experimental measurements. In the high-temperature region the on-set temperature is correct, but the increase in conversion is not fast enough. Extra *ab initio* calculations for the high-temperature chemistry should be performed in order to match with experimental observations in this region. The rather good match between experiments and the microkinetic model with only one fitted parameter confirms the power of the first principles based automatic kinetic model generation framework encoded in Genesys for the gas phase, which can now be extended to allow liquid-phase process modelling.

4.7. References

1. F. Buda, B. Heyberger, R. Fournet, P.-A. Glaude, V. Warth, and F. Battin-Leclerc, *Modeling of the Gas-Phase Oxidation of Cyclohexane*. Energy & Fuels, 2006. **20**(4): p. 1450-1459.
2. E.J. Silke, W.J. Pitz, C.K. Westbrook, and M. Ribaucour, *Detailed Chemical Kinetic Modeling of Cyclohexane Oxidation†*. The Journal of Physical Chemistry A, 2007. **111**(19): p. 3761-3775.
3. L. Cai and H. Pitsch, *Optimized chemical mechanism for combustion of gasoline surrogate fuels*. Vol. 162. 2015.
4. F. Battin-Leclerc, *Detailed chemical kinetic models for the low-temperature combustion of hydrocarbons with application to gasoline and diesel fuel surrogates*. Progress in Energy and Combustion Science, 2008. **34**(4): p. 440-498.
5. S.E. Klai and F. Baronnet, *Study of the homogeneous gas-phase oxidation of cyclohexane .1. Experimental-study*. Journal De Chimie Physique Et De Physico-Chimie Biologique, 1993. **90**(10): p. 1929-1950.
6. S.E. Klai and F. Baronnet, *Study of the homogeneous gas-phase oxidation of cyclohexane .1. Experimental-study*. Journal De Chimie Physique Et De Physico-Chimie Biologique, 1993. **90**(10): p. 1951-1998.
7. D. Voisin, A. Marchal, M. Reuillon, J.C. Boettner, and M. Cathonnet, *Experimental and Kinetic Modeling Study of Cyclohexane Oxidation in a JSR at High Pressure*. Combustion Science and Technology, 1998. **138**(1-6): p. 137-158.
8. A. El Bakali, M. Braun-Unkhoff, P. Dagaut, P. Frank, and M. Cathonnet, *Detailed kinetic reaction mechanism for cyclohexane oxidation at pressure up to ten atmospheres*. Proceedings of the Combustion Institute, 2000. **28**(2): p. 1631-1638.
9. C.J. Hayes, D.R. Burgess, and J.A. Manion, *Chapter Three - Combustion Pathways of Biofuel Model Compounds: A Review of Recent Research and Current Challenges Pertaining to First-, Second-, and Third-Generation Biofuels*, in *Advances in Physical Organic Chemistry*, I.H. Williams and N.H. Williams, Editors. 2015, Academic Press. p. 103-187
10. S. Granata, T. Faravelli, and E. Ranzi, *A wide range kinetic modeling study of the pyrolysis and combustion of naphthenes*. Combustion and Flame, 2003. **132**(3): p. 533-544.
11. B. Sirjean, P.A. Glaude, M.F. Ruiz-López, and R. Fournet, *Theoretical Kinetic Study of the Reactions of Cycloalkylperoxy Radicals*. The Journal of Physical Chemistry A, 2009. **113**(25): p. 6924-6935.
12. O. Lemaire, M. Ribaucour, M. Carlier, and R. Minetti, *The production of benzene in the low-temperature oxidation of cyclohexane, cyclohexene, and cyclohexa-1,3-diene*. Combustion and Flame, 2001. **127**(1): p. 1971-1980.
13. Z. Serinyel, O. Herbinet, O. Frottier, P. Dirrenberger, V. Warth, P.A. Glaude, and F. Battin-Leclerc, *An experimental and modeling study of the low- and high-temperature oxidation of cyclohexane*. Combustion and Flame, 2013. **160**(11): p. 2319-2332.
14. A.P. Zeelenberg and H.W. de Bruijn, *Kinetics, mechanism and products of the gaseous oxidation of cyclohexane*. Combustion and Flame, 1965. **9**(3): p. 281-295.
15. B.H. Bonner and C.F.H. Tipper, *The cool flame combustion of hydrocarbons I—Cyclohexane*. Combustion and Flame, 1965. **9**(3): p. 317-327.
16. S.M. Handford-Styring and R.W. Walker, *Arrhenius parameters for the reaction HO₂ + cyclohexane between 673 and 773 K, and for H atom transfer in cyclohexylperoxy radicals*. Physical Chemistry Chemical Physics, 2001. **3**(11): p. 2043-2052.

17. Z. Hong, K.-Y. Lam, D.F. Davidson, and R.K. Hanson, *A comparative study of the oxidation characteristics of cyclohexane, methylcyclohexane, and n-butylcyclohexane at high temperatures*. Combustion and Flame, 2011. **158**(8): p. 1456-1468.
18. S.M. Daley, A.M. Berkowitz, and M.A. Oehlschlaeger, *A shock tube study of cyclopentane and cyclohexane ignition at elevated pressures*. International Journal of Chemical Kinetics, 2008. **40**(10): p. 624-634.
19. B. Sirjean, F. Buda, H. Hakka, P.A. Glaude, R. Fournet, V. Warth, F. Battin-Leclerc, and M. Ruiz-Lopez, *The autoignition of cyclopentane and cyclohexane in a shock tube*. Proceedings of the Combustion Institute, 2007. **31**(1): p. 277-284.
20. S. Vranckx, C. Lee, H.K. Chakravarty, and R.X. Fernandes, *A rapid compression machine study of the low temperature combustion of cyclohexane at elevated pressures*. Proceedings of the Combustion Institute, 2013. **34**(1): p. 377-384.
21. F. Wu, A.P. Kelley, and C.K. Law, *Laminar flame speeds of cyclohexane and mono-alkylated cyclohexanes at elevated pressures*. Combustion and Flame, 2012. **159**(4): p. 1417-1425.
22. E.D. H. Wang, B. Sirjean, D. A. Sheen, R. Tango, A. Violi, J. Y. W. Lai, F. N. Egolfopoulos, D. F. Davidson, R. K. Hanson, C. T. Bowman, C. K. Law, W. Tsang, N. P. Cernansky, D. L. Miller, R. P. Lindstedt, *A high-temperature chemical kinetic model of n-alkane (up to n-dodecane), cyclohexane, and methyl-, ethyl-, n-propyl and n-butyl-cyclohexane oxidation at high temperatures*. September 19, 2010
23. E. Ranzi, A. Frassoldati, R. Grana, A. Cuoci, T. Faravelli, A.P. Kelley, and C.K. Law, *Hierarchical and comparative kinetic modeling of laminar flame speeds of hydrocarbon and oxygenated fuels*. Progress in Energy and Combustion Science, 2012. **38**(4): p. 468-501.
24. W.J. Pitz, C.V. Naik, T.N. Mhaolduin, C.K. Westbrook, H.J. Curran, J.P. Orme, and J.M. Simmie, *Modeling and experimental investigation of methylcyclohexane ignition in a rapid compression machine*. Proceedings of the Combustion Institute, 2007. **31**(1): p. 267-275.
25. L.M. Cai, H. Pitsch, S.Y. Mohamed, V. Raman, J. Bugler, H. Curran, and S.M. Sarathy, *Optimized reaction mechanism rate rules for ignition of normal alkanes*. Combustion and Flame, 2016. **173**: p. 468-482.
26. B. Heyberger, N. Belmekki, V. Conraud, P.-A. Glaude, R. Fournet, and F. Battin-Leclerc, *Oxidation of small alkenes at high temperature*. International Journal of Chemical Kinetics, 2002. **34**(12): p. 666-677.
27. P.D. Paraskevas, M.K. Sabbe, M.-F. Reyniers, G.B. Marin, and N.G. Papayannakos, *Group additive kinetic modeling for carbon-centered radical addition to oxygenates and β -scission of oxygenates*. AIChE Journal, 2016. **62**(3): p. 802-814.
28. P.D. Paraskevas, M.K. Sabbe, M.-F. Reyniers, N. Papayannakos, and G.B. Marin, *Kinetic Modeling of α -Hydrogen Abstractions from Unsaturated and Saturated Oxygenate Compounds by Carbon-Centered Radicals*. ChemPhysChem, 2014. **15**(9): p. 1849-1866.
29. P.D. Paraskevas, M.K. Sabbe, M.-F. Reyniers, N.G. Papayannakos, and G.B. Marin, *Kinetic Modeling of α -Hydrogen Abstractions from Unsaturated and Saturated Oxygenate Compounds by Hydrogen Atoms*. The Journal of Physical Chemistry A, 2014. **118**(40): p. 9296-9309.
30. J.T. Moss, A.M. Berkowitz, M.A. Oehlschlaeger, J. Biet, V. Warth, P.-A. Glaude, and F. Battin-Leclerc, *An Experimental and Kinetic Modeling Study of the Oxidation of the Four Isomers of Butanol*. The Journal of Physical Chemistry A, 2008. **112**(43): p. 10843-10855.
31. A.G. Vandeputte, M.K. Sabbe, M.F. Reyniers, V. Van Speybroeck, M. Waroquier, and G.B. Marin, *Theoretical study of the thermodynamics and kinetics of hydrogen abstractions from hydrocarbons*. Journal of Physical Chemistry A, 2007. **111**(46): p. 11771-11786.

- 32. M.K. Sabbe, M.F. Reyniers, V. Van Speybroeck, M. Waroquier, and G.B. Marin, *Carbon-centered radical addition and beta-scission reactions: Modeling of activation energies and pre-exponential factors*. Chemphyschem, 2008. **9**(1): p. 124-140.
- 33. M.K. Sabbe, M.-F. Reyniers, V. Van Speybroeck, M. Waroquier, and G.B. Marin, *Carbon-Centered Radical Addition and β -Scission Reactions: Modeling of Activation Energies and Pre-exponential Factors*. ChemPhysChem, 2009. **10**(12): p. 1951-1951.
- 34. J. Troe, *The thermal dissociation/recombination reaction of hydrogen peroxide $H_2O_2(+M) \rightleftharpoons 2OH(+M)$ III.. Analysis and representation of the temperature and pressure dependence over wide ranges*. Vol. 158. 2011.
- 35. B. Sirjean, P.A. Glaude, M.F. Ruiz-Lopez, and R. Fournet, *Detailed Kinetic Study of the Ring Opening of Cycloalkanes by CBS-QB3 Calculations*. The Journal of Physical Chemistry A, 2006. **110**(46): p. 12693-12704.
- 36. J. Bugler, K.P. Somers, E.J. Silke, and H.J. Curran, *Revisiting the Kinetics and Thermodynamics of the Low-Temperature Oxidation Pathways of Alkanes: A Case Study of the Three Pentane Isomers*. The Journal of Physical Chemistry A, 2015. **119**(28): p. 7510-7527.

5.

Fast assignment of liquid-phase thermodynamics with Genesys

5 . FAST ASSIGNMENT OF LIQUID-PHASE THERMODYNAMICS WITH GENESYS.....137

5.1.	INTRODUCTION	138
5.2.	STATE OF THE ART FOR THE LIQUID-PHASE OXIDATION OF CYCLOHEXANE	139
5.2.1.	Literature modelling and experimental work	139
5.2.2.	Experimental GC×GC analysis	141
5.3.	EXTENDING GENESYS TO A LIQUID-PHASE ENVIRONMENT	143
5.3.1.	Molecules.....	143
5.3.2.	Reaction families	144
5.3.3.	Reaction rules	145
5.4.	FAST ASSIGNMENT OF THERMODYNAMIC DATA: ALGORITHM TO ASSIGN THE GIBBS FREE ENERGY OF SOLVATION	146
5.5.	VALIDATION AND RESULTS	149
5.5.1.	Available validation data	149
5.5.1.1	COSMOtherm.....	149
5.5.1.1	FreeSolv database	150
5.5.2.	Abraham solute descriptors validation.....	150
5.5.3.	Platts fragments validation.....	151
5.5.3.1	Complete data set.....	152
5.5.3.2	Validation for cyclohexane as solvent	155
5.5.4.	Improving the performance for carbon dioxide	156
5.6.	CONCLUSIONS	158
5.7.	REFERENCES	160

5.1. Introduction

An extension of Genesys is required before the automatic generation of a microkinetic model for the liquid-phase oxidation of cyclohexane becomes possible. Currently the computer-aided framework is tailored to the gas phase, in which it is necessary to implement liquid-phase effects. The latter would allow model generation for liquid-phase processes as is envisioned in the longer term of this project. The presence of a liquid has an effect on all three main pillars of automatic kinetic model generation, *i.e.* molecules, the reaction family definition and reaction rules, which form the fundamentals for reaction network generation. By extension, the fast assignment of thermodynamic properties to all molecules and kinetic parameters to all reactions in the kinetic model needs to be extended with fundamental knowledge of the liquid phase. In this master dissertation, the first step is taken by introducing a calculation algorithm for the fast assignment of the Gibbs free energy of solvation to species via the empirical Abraham equation, see section 2.4. The latter will be assessed for its performance in general and more specifically for the application to the liquid-phase oxidation of cyclohexane.

First, an overview is given on the kinetic modelling, *ab initio* and experimental studies that have already been performed concerning the liquid-phase oxidation of cyclohexane. This is followed by an overview of the three pillars of automatic kinetic model generation and how these should be redefined for the implementation of liquid-phase chemistry. Next, the newly implemented algorithm in Genesys for the calculation of the Gibbs free energy of solvation is discussed in more detail. Similar to the gas-phase algorithm, a roadmap is followed for the assignment of the Gibbs free energy of solvation depending on the availability of data in several databases.

Finally, the procedure to calculate the Gibbs free energy of solvation is validated with *ab initio* determined values obtained from COSMOtherm [1, 2] and the FreeSolv database [3] found in literature which contains experimental determined values for the Gibbs free energy of hydration. A more detailed discussion is made on the validation of species important for the oxidation of cyclohexane. Some improvements to the scheme are made and suggested for a better performance.

5.2. State of the art for the liquid-phase oxidation of cyclohexane

A lot of research has already been performed in the past concerning the liquid-phase oxidation of cyclohexane since it is an important industrial process for the production of cyclohexanone and cyclohexanol, building blocks for the production of adipic acid and ϵ -caprolactam. Several groups (Steeman et al. [4], Lipes and Furman [5], Balcerzak et al. [6]) indicated the importance of the residence time on the selectivity, since the desired products are intermediates in a sequence of reactions, and over-oxidation results easily in a number of useless byproducts. Still, the industrial conversion of the process is kept low, around 5 %, to avoid the formation of ring-opened products. However, the goal should be to develop an accurate kinetic model with elementary reactions which allows an increase in selectivity and/or conversion with as outcome a higher production yield.

5.2.1. Literature modelling and experimental work

In the literature several simplified models are proposed to describe the reaction kinetics of the liquid-phase oxidation of cyclohexane. Spielman et al. [7] presented one of the earliest models with only a limited amount of lumped reactions as presented in the reaction scheme in Figure 5-1.

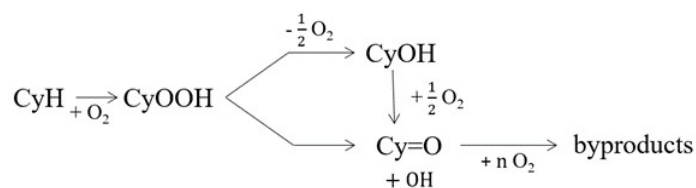


Figure 5-1: Liquid-phase oxidation scheme of the reactions as presented by Spielman et al. [7]; CyH, CyOOH, CyOH and Cy=O stand respectively for cyclohexane, cyclohexyl hydroperoxide, cyclohexanol and cyclohexanone.

Alagy et al. [8], proposed a more extensive kinetic model, but this neglected the formation of the hydroperoxide which is both an important intermediate and end-product of the liquid-phase oxidation. A more detailed model for the non-catalytic oxidation, *i.e.* the autoxidation, was presented by Khar'kova et al. [9] which also performed experiments for validation. However, this model still only contained 19 reactions which makes it impossible to accurately predict byproduct formation.

Pohorecki et al. [10] proposed a complete process model for the cyclohexane oxidation in the liquid phase, combining both a kinetic reaction model and a mass transfer model. The latter is necessary since the process is affected both by chemical kinetics and by mass transfer limitations especially for the presence of molecular oxygen in liquid cyclohexane. Earlier experimental work by Balcerzak et al. [6] at a temperature of 160 °C and a pressure of 9.0 bara were used to regress the kinetics of the model.

Page Removed due to confidentiality

Page Removed due to confidentiality

Page Removed due to confidentiality

5.3. Extending Genesys to a liquid-phase environment

The generation of a microkinetic model by an automatic kinetic model generator is based upon three pillars. These are the molecules, the reaction families and the reaction rules, as indicated in Figure 5-4. Extending the gas-phase framework will require adaptations for all three pillars to include the presence of the liquid phase.

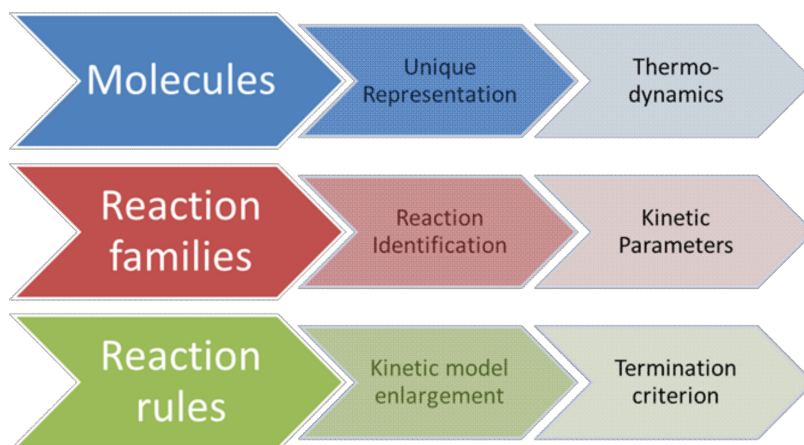


Figure 5-4: Schematic overview of the three pillars required for automatic kinetic model generators.

5.3.1. Molecules

The first pillar concerns the molecules which are basically a collection of atoms each bonded to one or more atoms. To construct a detailed predictive microkinetic model, it is necessary to represent these species unambiguously, *i.e.* represent each species in a unique manner. Most often, the molecule identifier, either via a SMILES or a unique InChI identifier, of the species is converted into a weighted graph structure. A molecule identifier is a textual identifier for chemical substances, designed to provide a standard way to encode molecular structural information and to facilitate the search for such information in databases and in the open literature. This representation comprises nodes, representing the atoms, and edges, representing the chemical bonds. Additional information such as chemical elements, charges, lone electron pairs, unpaired electrons, bonds orders, is given to these nodes and edges.

Currently in Genesys no information about the liquid phase or a solvent is available. The species representation through graph will need to be extended to entail solvent information. Therefore, a new concept will have to be introduced in Genesys for interactions between molecules: solute molecules are surrounded with solvent molecules, this is not equal to a chemical bond, which is currently the only possibility to link atoms/molecules. This information is needed during the model generation to allow the calculation of thermochemistry and kinetics, and also to decide whether certain reactions are probable

for a species in a solvent. The presence of a solvent will influence which reactions are kinetically significant depending on the given operating conditions.

Besides the molecular representation, it is also important to obtain thermodynamic properties for each species. For the gas phase a large database is available with ab initio calculated values at the CBS-QB3 level of theory. A similar database can be constructed with ab initio results and experimental values for the thermodynamic properties of species in different solvents. When database information lacks for a species, the GAV scheme of Benson is applied which allows an accurate thermodynamic property calculation for the gas phase, as extensively discussed in section 3.1.4.1. It will be investigated if a similar approach can be used for the prediction of thermodynamic properties of solutes and solvents. Empirical correlations are available in literature, *e.g.* the Abrahams equation, which allows calculation of the Gibbs free energy of solvation at standard conditions that can be used to determine thermodynamic properties in the liquid phase. The Abrahams equation is extended with the Platts fragment scheme that can be used for fast assignment of thermodynamic properties for the liquid phase, see literature study section 2.4. This method is implemented in Genesys and validated against experimental and ab initio obtained data. In case this method does not yield results for the thermodynamic properties of a solute, on-the-fly ab initio calculations can be carried out for which best practices have to be determined.

5.3.2. Reaction families

A kinetic model can contain from tens up to tens of thousands of reactions, *c.f.* Figure 1-1, but most of these reactions can be categorized in just a few reaction families. It is thus sufficient for automatic kinetic model generation software to contain information about a limited number of reaction families. Each reaction family is defined by the structural features (*i.e.* functional groups) of the reactant(s), a recipe on how to convert the reactant(s) in the product(s) and information on how to calculate the rate coefficients for each reaction. Similar to gas-phase reactions, reactions that occur in the liquid phase, especially for the radical oxidation of cyclohexane, the thousands of reactions can be classified in reaction families. Moreover, literature indicates that these reactions are similar to the reactions that occur in the low temperature gas-phase oxidation region. This makes that the knowledge obtained from the gas-phase oxidation model, developed in chapter 4, can be reused for the construction of a model for the liquid-phase oxidation of cyclohexane.

The reaction network requires assignment of kinetic parameters to the chemical reactions to calculate the reaction rate. In the liquid phase, a distinction is made between the apparent and intrinsic reaction rate due to mass transfer or diffusion limitations. The intrinsic rate coefficients can be either originating from direct measurements under well-chosen conditions or calculation of the liquid-phase reaction rate coefficient considering the diffusional effect, another option that is worth exploring is to start from the gas-phase reaction rate coefficients for which an additional correction is necessary to describe the effect

of the solvent. Databases will have to be constructed with available intrinsic kinetic data in liquids found in the open literature or from new *ab initio* calculations. Similar to the thermodynamic properties the current group-additivity scheme for reaction rates is tailored to the gas phase which has to be extended. Typically, for the calculation of rate coefficients in a liquid or solvent, a contribution of the diffusion is added to the intrinsic reaction rate coefficient which requires knowledge of the radius, viscosity and diffusivity of species. The intrinsic reaction rate and the diffusion rate are calculated separately and processed to obtain an outcome for the true or apparent rate coefficient in the liquid phase.

5.3.3. Reaction rules

The third pillar involves the applicability of so-called reaction rules. To automatically generate all the possible reactions that a species can undergo, elementary transformations are defined to execute a reaction specifying the molecular structure rearrangements. After the execution of these transformations, product species are formed which can subsequently be subjected to reactions themselves resulting in new products. The network generation iteratively continues until the reaction network is considered complete by user-defined criteria. The latter is included by so-called reaction rules which puts criteria on the reactions which are allowed to be executed and the products allowed to be formed. Limitations can be imposed for each reaction family, *e.g.* only react atoms with less than one oxygen atom, as to limit the size of the generated reaction network to the essential chemistry.

Experience in termination criteria has been gained for the construction of gas-phase models, but for liquid-phase models this experience will have to be gained via literature reviewing, *e.g.* simulation of the results of Pohorecki et al. [10, 11] and Hermans et al. [12, 13] and the study of experimental data.

5.4. Fast assignment of thermodynamic data: Algorithm to assign the Gibbs free energy of solvation

The first implementation of a liquid phase in Genesys concerns the extension of the scheme for fast assignment of thermodynamic properties. An algorithm is introduced to calculate the Gibbs free energy of solvation at standard conditions of 1 bara pressure of solute in the gas phase and a solvent concentration of 1 mol l⁻¹ at 298 K, *i.e.* $\Delta G_{solv}^{\circ}(298K)$, via the empirical Abraham equation. The latter requires the availability of solute and solvent descriptors. The popularity of the Abraham equation is indicated by the amount of recent publications which determine solute descriptors for a wide range of drug-like species [14-18]. When solute descriptors are missing in the database, it can be tried to calculate these with the Platts fragments method. In Figure 5-5 the algorithm implemented in Genesys to calculate $\Delta G_{solv}^{\circ}(298K)$ is visualized. The algorithm starts with the definition of a solute and solvent, which have to be specified as input via either SMILES or InChI identifiers, and subsequently generates as output the Gibbs free energy of solvation if enough data is available.

The algorithm requires reading and storing of all liquid-phase databases including ab initio data, but also Abraham solute and solvent descriptors and the Platts fragments with associated contributions. First, the databases with experimental and ab initio data are read which contain values for $\Delta G_{solv}^{\circ}(298K)$ for a specific combination of solute and solvent. The latter databases are discussed in more detail in section 5.5.1, as they are used in this master dissertation for validation of the Abraham equation and the Platts fragment method. Second, the databases with solute and solvent Abraham descriptors are read. Third, the database with Platts fragments and accompanying values for the solute descriptors is read.

When all data is read, a first query is performed to check if the combination of solvent and solute appears in the database with experimental or ab initio data. If so, the value of $\Delta G_{solv}^{\circ}(298K)$ of the database is assigned. If not, the next query is performed to check if the combination of solvent and solute molecules appears in the database with respectively Abraham solvent descriptors and solute descriptors. When the solvent descriptors are missing, it is not possible to determine $\Delta G_{solv}^{\circ}(298K)$. If the solute descriptors are missing, these can be predicted via the Platts fragments method. The latter method is similar to Benson's group additivity theory and requires that all non-hydrogen atoms can be matched with Platts fragments. In the latter case it is also impossible to calculate a value for $\Delta G_{solv}^{\circ}(298K)$.

The calculated value for each solute descriptor by Platts fragments is obtained by a very similar procedure as the GAVs algorithm implemented in Genesys which was discussed in the methodology chapter, section 3.1.4.2. Due to the similarity of the algorithm for GAV assignment, the latter is reused for the assignment of Platts fragments to a molecule. Differences between the GAV and Platts scheme is that there is an interception value, *i.e.* a fixed contribution which has to be added to the different solute

descriptors independent of the molecular structure, and that the groups in both schemes are sometimes differently defined. For example, a carbonyl group is one group in the GAV scheme, while it is composed out of two groups, *i.e.* a double bonded oxygen atom and a carbon atom with a double bond, for the Platts fragments method. When both the solute and solvent descriptors are available from a database or by calculation, the Abraham equation is utilized to calculate a value for $\Delta G_{solv}^{\circ}(298K)$, which is eventually assigned.

In appendix C section C.1 and section C.2, the descriptor values stored in respectively the solvent and solute database can be found which are obtained from the RMG database [19]. This includes solvent descriptors for 25 frequently used solvents including water, cyclohexane, benzene and 1-octanol, and solute descriptors for 152 solutes. Moreover, the list with 28 Platts fragments and corresponding solute descriptor contributions are listed in appendix C section 3 as used in this master dissertation, these are obtained from the original paper published by Platts et al. [20].

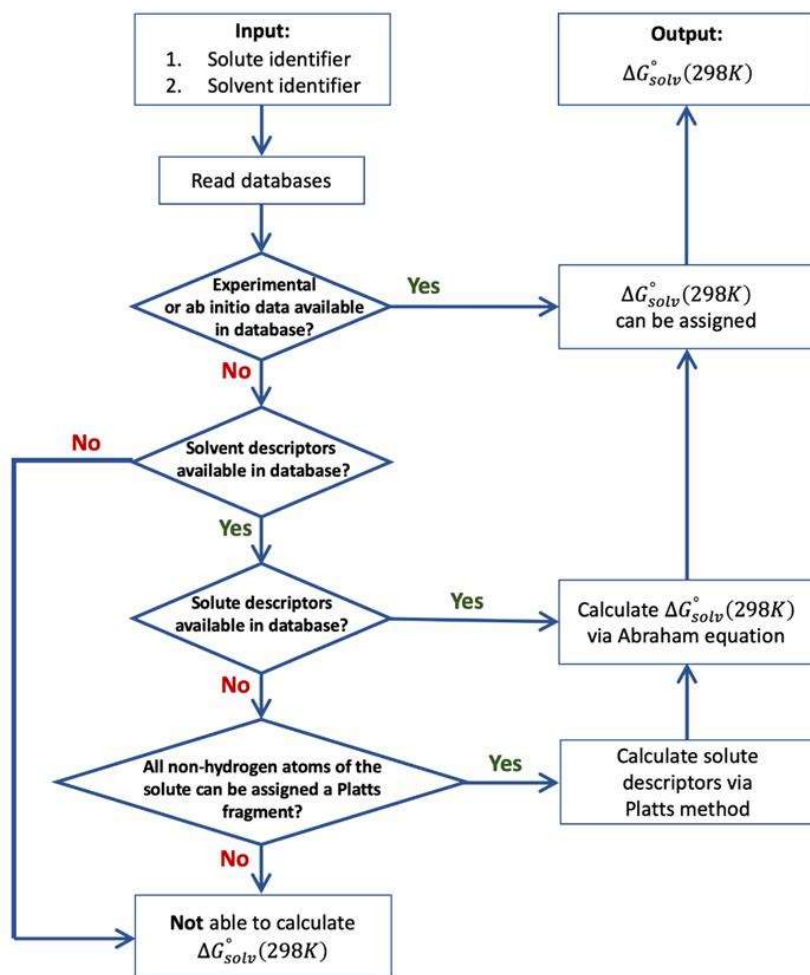


Figure 5-5: Schematic representation of the algorithm implemented in Genesys to assign the Gibbs free energy of solvation at 298 K to a species.

The necessary framework to implement the Mintz equation in Genesys is similar to the previous described algorithm for the Abraham equation and is already partially encoded. The Mintz equation is used for calculation of the enthalpy of solvation at 298 K. Note that for the Mintz equation the same solute descriptors are used as for the Abraham equation, but the used solvent descriptors are different. The Mintz equation is however not covered in this chapter. With both the enthalpy and Gibbs free energy of solvation calculated, the entropy of solvation can be obtained. The latter allows to account for the temperature dependency of the Gibbs free energy of solvation by assuming that the entropy and enthalpy are temperature independent. It should also be noted that the application of the Mintz equation is often limited to a small temperature region around 298 K in literature, see section 2.4.6.

The enthalpy of species in the liquid phase can be calculated from the enthalpy in the gas phase and the enthalpy of solvation, *cf.* equation (5-1), and similar for the entropy, *cf.* equation (5-2).

$$H_{liq}(T, P) = H_{Gas}(T, P) + \Delta H_{solv}(T, P) \quad (5-1)$$

$$S_{liq}(T, P) = S_{Gas}(T, P) + \Delta S_{solv}(T, P) \quad (5-2)$$

The latter is valid since both thermodynamic properties are state functions. The gas-phase enthalpy and entropy can be obtained from the Benson group additivity scheme that is already implemented in Genesys for gas-phase species, as discussed in section 3.1.4. Note that the current gas-phase framework in Genesys and the group additive values are currently valid for a temperature range around 1000 K. Similar developed GAVs for the gas-phase are available in literature, valid around 300 K which are more interesting since liquid-phase processes are typically performed at low temperatures. The enthalpy of solvation and the entropy of solvation are obtained after application of the Abraham and Mintz equation.

5.5. Validation and results

Validation of the above-mentioned algorithm is performed by comparing the Gibbs free energy of solvation calculated by Genesys with ab initio calculated and experimental tabulated values. Compared to the discussed algorithm, *cf.* Figure 5-5, all values from the experimental and ab initio database are compared to calculated values with the Abraham method and to calculated values using the Platts fragments method for validation of both methods.

5.5.1. Available validation data

In the literature not a lot of databases are reported with experimental determined values for the Gibbs free energy of solvation. Though, a well-known database is the Minnesota solvation database [21] from the group of Truhlar and Cramer with 3037 reported values for 92 solvents and 790 unique solutes. However, access was not granted in time. For the validation two other sets of data are used, *i.e.* one set obtained from a predictive property calculation tool (COSMOtherm) based on ab initio calculations and the set of Mobley et al. [3] found in literature with reported experimental values for the Gibbs free energy of hydration.

The complete validation set contains 4565 entries of which several duplicates are present. Typical values for the Gibbs free energy of solvation are situated in the range between -30 and 0 kJ mol⁻¹ with a minority of the values being positive. Results of the validation will be evaluated using the mean absolute error (MAE) as defined in equation (5-3) and the maximal error.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5-3)$$

Note that radicals are not considered in the validation set as these are not available in any of the validation data sets. The latter is crucial to obtain thermodynamic parameters for the intermediate products formed during liquid-phase oxidation reactions.

5.5.1.1 COSMOtherm

Because of lack of experimental validation data available from literature, ab initio values for ΔG_{solv}° are determined using the commercial software package COSMOtherm and used as validation data. The aim of this software package is to post-process ab initio calculations carried out by other software packages, *e.g.* Gaussian or TurboMole, and determine thermodynamic properties, such as ΔG_{solv}° at standard conditions. In COSMOtherm, besides the ability to include new ab initio calculations for the determination of thermodynamic properties, a large database is present with ab initio calculations for the most commonly used solvents and solutes. This database includes ab initio calculations for 272

species at the BP-TZVPD-FINE level of theory, *i.e.* the latest developed level of theory included in COSMOtherm with the current highest accuracy. Out of the data set of 272 species, the solvents that are present in the Abraham solvent database are selected and combined with all C/H/N/S/O solutes present, excluding ions and polymers, to determine ΔG_{solv}° . Although a lot of halogens are available in COSMOtherm, it was decided to not extract them and to not put them in the validation set, since halogens are not of interest for studying the liquid-phase oxidation of cyclohexane. A method is added to Genesys to automatically create input files, run these simulations and post-process the results to the required database format. This results in a final data set of 3922 entries for ΔG_{solv}° that can be used for validation of the Abraham and Platts method. Values are taken at a solvent concentration of 1 mol l⁻¹ and a partial pressure of the solute of 1 bara and 298 K, *i.e.* standard conditions.

5.5.1.1 FreeSolv database

Mobley et al. [3] has constructed a database, the so-called Free Solvation database or FreeSolv database, which contains experimental determined values of the Gibbs free energy of hydration, *i.e.* for solvation in water, for 643 solutes. The validation set thus only contains experimental data with water as solvent. A wide range of hydrocarbons are covered in the database, including hydrocarbons with hetero-atoms, *i.e.* nitrogen, sulfur and halogens, and several functional groups such as the carbonyl, carboxyl, nitro, thiol, amine and amide groups. Data was obtained from prior literature publications. The database is constructed in an open format including SMILES for the molecules. Several common values present in both data sets were compared and matched within 1 kJ mol⁻¹.

5.5.2. Abraham solute descriptors validation

For all combinations of solutes and solvents present in the validation set, *i.e.* the ab initio and experimental database, ΔG_{solv}° is calculated via the Abraham equation and the respective solute and solvent descriptors (see Appendix C Table C-1 and C-2). For 1438 of the 4565 combinations, entries for both solute and solvent descriptors are available. Note that data is available for 25 solvents in the solvent descriptor database, but only 20 of them are validated since several solvents are not represented in the validation set, the solvents not validated are carbon tetrachloride, chloroform, dichloroethane, undecane and iso-octane. The difference between the tabulated and calculated value of ΔG_{solv}° for each entry is depicted in Figure 5-6. The blue shaded area covers the data points with an absolute maximal error of 4 kJ mol⁻¹. Only 21 data points out of the 1438, *i.e.* 1.5%, are estimated with an absolute error above 4 kJ mol⁻¹. The MAE amounts to 1.1 kJ mol⁻¹ and the maximal error amounts to 7.67 kJ mol⁻¹ for the Gibbs free energy of solvation of water in 1-octanol.

Large systematic deviations between calculated and tabulated values are not observed in Figure 5-6. Note that the largest part of the used validation set is based on ab initio calculated values which are

prone to errors. Similarly, experimental errors can be expected in the FreeSolv database. The points indicated by the blue circle in Figure 5-6 stand out. They have in common that the solvent is the same namely 1-octanol. The MAE within this reduced validation set containing only the data points with 1-octanol as solvent amounts to 2.2 kJ mol^{-1} , which is significantly higher than the MAE of the complete validation set. Except for the solvation of water in 1-octanol, the other points with outstanding errors are for solvation of n-nonane and n-decane in 1-octanol with an error of respectively 5.21 and 5.00 kJ mol^{-1} , and n-propyl formate and propylcyclopentane in water, respectively 5.27 and 4.79 kJ mol^{-1} .

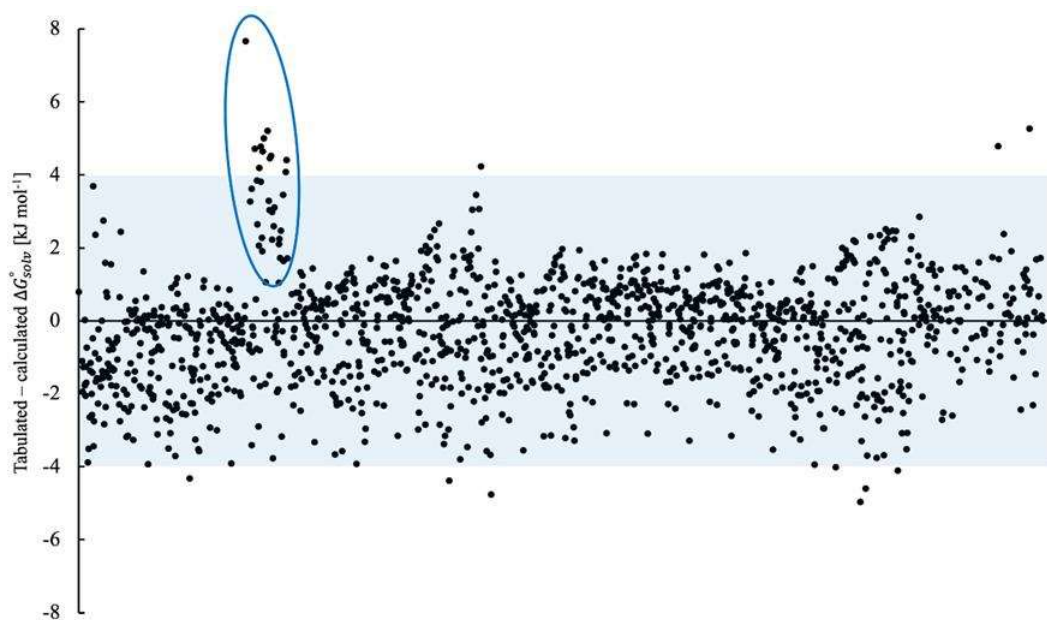


Figure 5-6: Plot of tabulated minus calculated values for ΔG_{solv}° at 298 K using tabulated Abraham solvent and solute descriptors.

5.5.3. Platts fragments validation

When solute descriptors are lacking these can be calculated via the Platts fragments method [20]. For the solvent descriptors, the need for tabulated values remains since a group additive scheme is not yet available. Predictive schemes are available in the open literature, but the performance is not optimal [22]. In addition, most frequently used solvents are present in the database, which decreases the need for urgent implementation of such a scheme. The Platts fragments method is compared with the available validation set to gain maximal information concerning the performance of the method. The implemented Platts fragments for carbon, oxygen and some nitrogen groups can be found in Appendix C section C.3 as these are published in the original paper of Platts [20]. The different hydrogen bond corrections which were also included in the original paper of Platts are not yet implemented in Genesys due to the rather rare molecular structure requirements which are seldom present during oxidation processes. For sulfur

and halogen functionalities, the Platts fragments are regressed but they are not yet implemented in the Genesys database since they are regressed on significantly less compounds and therefore considered less reliable. In addition, they are of minor importance for oxidation of hydrocarbons. The comparison between the Platts method and the validation data set is done, only in the case that (1) the solvent descriptors are available from the Abraham data set and (2) all non-hydrogen atoms in the solute must have been assigned a Platts fragment. Note that the presence of water in the validation set is more dominant and therefore influences the MAE more than any other solvent.

5.5.3.1 Complete data set

From the complete data set with 4565 entries, 3549 could be calculated with the Platts fragments method, *i.e.* 77.7 %. The sulfur and halogen containing species in the validation set cannot be calculated and several nitrogen functionalities are also not available as Platts fragment, *e.g.* the nitro-functionality which is frequently represented in the validation set. Of the 643 entries of the FreeSolv database only 354, *i.e.* only 55.1 %, are calculated due to the high presence of missing nitrogen and sulfur functionalities and a lot of halogen containing components.

The difference between tabulated and calculated values for ΔG_{solv}° are plotted in Figure 5-7 with a different color for each solvent. The acyclic hydrocarbon solvents are indicated in red shades, cyclic hydrocarbon solvents in orange shades, oxygenated solvents in blue shades and the nitrogen-containing solvents in green shades. The dark grey and light grey shaded area cover respectively the data points with an absolute maximal error of 4 kJ mol⁻¹ and 10 kJ mol⁻¹. The Platts method on the complete data set has a MAE of 2.5 kJ mol⁻¹. As expected, the performance is less compared to the tabulated Abraham solute descriptors. On the other hand, the number of calculated values has been increased from 1438 to 3549. Hence, the algorithm as implemented in Genesys is well structured, *i.e.* the Abraham method with tabulated solute descriptors is used for calculation of ΔG_{solv}° before the Platts fragments method is applied for the determination of solute descriptors. A maximal error of 34.5 kJ mol⁻¹ is noticed for solvation of 1,4-diamino-9,10-anthracenedione in water and 82.6 % of the ΔG_{solv}° for solute-solvent combinations are calculated within 4.0 kJ mol⁻¹. Even for compounds with exceptional low tabulated values for ΔG_{solv}° the calculated value with Platts fragments methodology matches very well, *e.g.* the experimental determined value for hydration of mannose, a sugar monomer, amounts to -106.6 kJ mol⁻¹ and the calculated value amounts to -107.7 kJ mol⁻¹.

From Figure 5-7, it is clear that the MAE is not the same for all solvents and that a different performance is observed for apolar and polar solvents. The acyclic and cyclic hydrocarbons perform best and only seldom data points are observed with an error above 10.0 kJ mol⁻¹. For the oxygenated and nitrogen-containing solvents, *e.g.* N,N-dimethyl formamide, dimethyl sulfoxide, ethyl acetate and acetonitrile, a

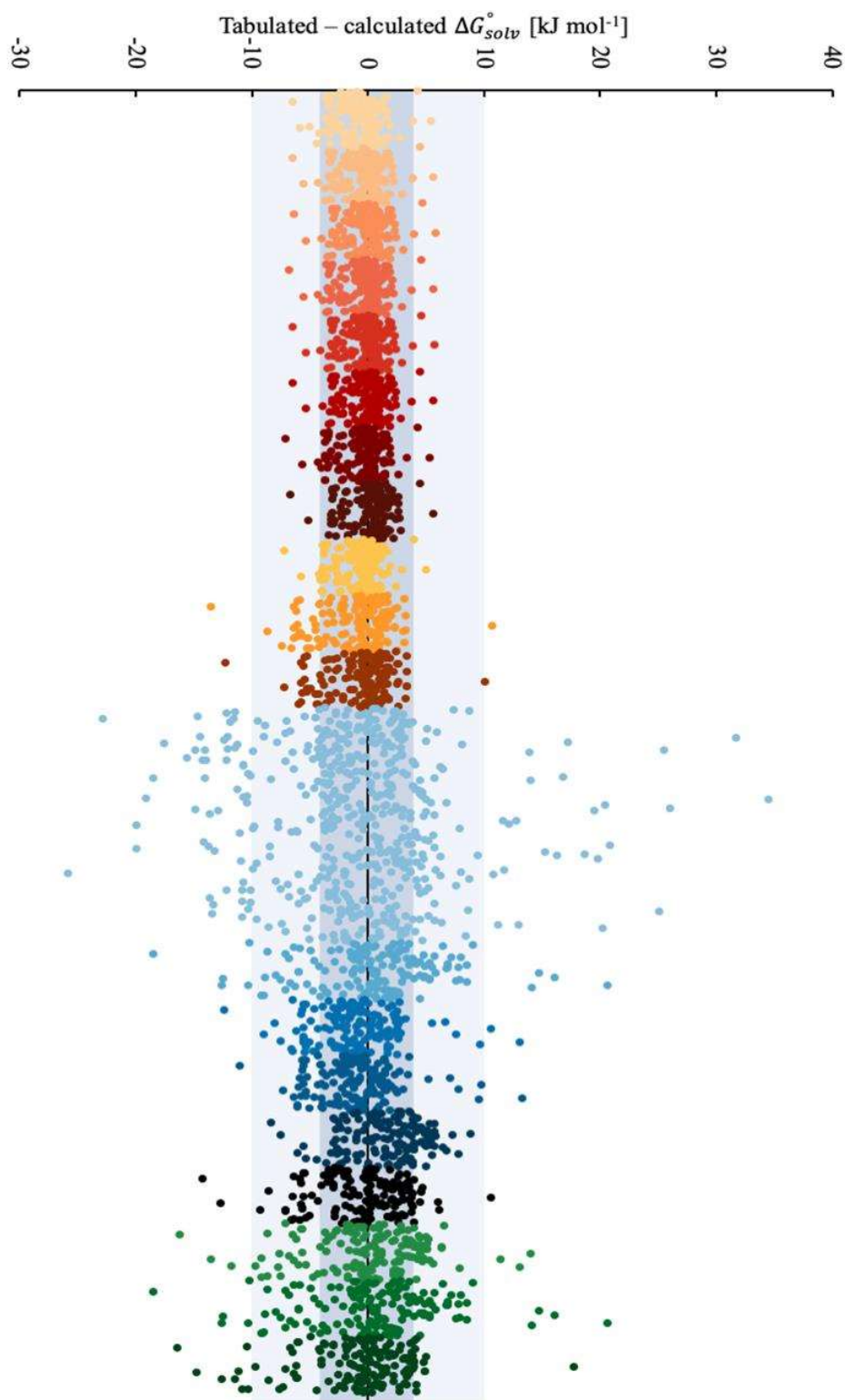


Figure 5-7: Plot of tabulated minus calculated values for ΔG_{solv}° using Platts fragments for solutes descriptors for 20 different solvents. The different solvents are indicated in different colors, from left to right, red shades: n-pentane, n-hexane, n-heptane, n-octane, n-nonane, n-decane, dodecane, hexadecane; orange shades: cyclohexane, benzene, toluene; blue shades: water, ethanol, 1-butanol, 1-octanol, di-n-butylether, ethyl acetate; green shades: N,N-dimethyl formamide, dimethyl sulfoxide, acetonitrile.

much broader spreading of the data points is observed, especially water stands out. All of these solvents have in common that they are polar solvents. Di-n-butylether which is not polar and 1-octanol which is less polar than water do perform better. Clearly for polar solvents it is more difficult to calculate the ΔG_{solv}° . 1-octanol has several outliers when calculated with tabulated Abraham solute descriptors, discussed in section 5.5.2, but this is not visible when calculating ΔG_{solv}° with Platts fragments.

Several outliers are detected in Figure 5-7. When the combination of solute and solvent molecules for the outlying data points are checked, it turns out that this are not always the same solutes which have a higher error in different solvents and thus they cannot be considered as systematic deviations. The outliers comprise all kinds of solutes, *i.e.* normal acyclic hydrocarbons, (cyclic) oxygenated hydrocarbons, nitrous compounds, monofunctional as well as bifunctional, *etc.* One exception is carbon dioxide as a solute which has a tabulated value for ΔG_{solv}° in all solvents except water between -2.0 and -5.5 kJ mol⁻¹ while the calculations range from -9.0 to -19.9 kJ mol⁻¹. In addition, the Gibbs free energy of solvation in water amounts to 2.3 kJ mol⁻¹ which is calculated via the Abraham equation with Platts fragments descriptors to be -29.9 kJ mol⁻¹, an error of more than 30 kJ mol⁻¹. Note that in the validation set only ab initio data is available for the solvation of carbon dioxide. Ab initio calculations are also prone to errors, but experimental data from literature matches well, *e.g.* the experimental versus ab initio calculated Gibbs free energy of hydration of carbon dioxide amounts to 0.8 kJ mol⁻¹ [23] and 2.3 kJ mol⁻¹ respectively. As an improvement, a regression will be performed in section 5.5.4. to obtain solute descriptors for the carbon dioxide molecule that can replace the calculation by Platts fragments and improve the accuracy.

Finally, when the solvents n-pentane, n-hexane, n-heptane, n-octane, n-nonane, n-decane, dodecane and hexadecane are considered, *i.e.* points indicated with red shades in Figure 5-7, it is clear that the same trend appears for each of these solvents, including the outliers. The solvents have in common that they are all normal acyclic hydrocarbons with which it is not possible to have polar interactions nor hydrogen bonds, characterized by a value of zero for the solvent descriptors a, b and s. Also cyclohexane, a naphthene, behaves similarly. The observed outliers are identified as carbon dioxide, 1,3-diazine, thymine and salicylic acid, indicated in Figure 5-8. Except for carbon dioxide, these are rather rare substances which are not immediately expected during oxidation processes. Hence, for now, the focus for improvement will be on regression of solely the Abraham solute descriptors for CO₂.

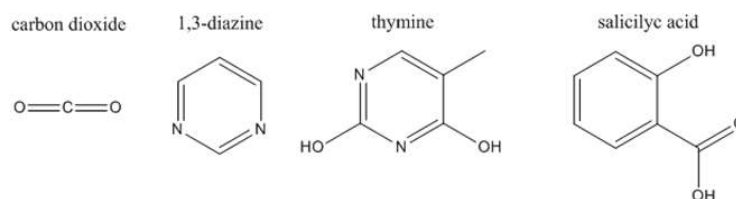


Figure 5-8: Molecular structure of carbon dioxide, 1,3-diazine, thymine and salicylic acid.

5.5.3.2 Validation for cyclohexane as solvent

Cyclohexane is also considered separately as solvent, since it is the solvent of interest when performing research on the liquid-phase oxidation of cyclohexane. A plot of the tabulated minus calculated values for ΔG_{solv}° is depicted in Figure 5-9 for this reduced set of data. When calculating ΔG_{solv}° with tabulated Abraham solute descriptors on a reduced data set with only cyclohexane as solvent, the MAE amounts to 0.9 kJ mol⁻¹ with 64 of the 152 species calculated. The MAE for the same reduced data set with 152 entries amounts to 1.3 kJ mol⁻¹ when calculating the solute descriptors via the Platts fragments methodology, which is significantly lower than the MAE of the overall validation set. Moreover, only 4% of the data points are not calculated within the boundary of 4 kJ mol⁻¹ with as outliers water, carbon dioxide, acetic anhydride and glycerin as observed in Figure 5-9. Also, the solvation of cyclohexanone, important for the liquid-phase oxidation of cyclohexane, is comprised in the validation set and has an error of only 0.5 kJ mol⁻¹. The Gibbs free energy of solvation for cyclohexanol is not comprised in the validation set.

Note that the peroxide group is missing as Platts fragment which is a crucial fragment for calculation of thermodynamic parameters of oxidation intermediates and products. In literature it is common to approximate the peroxide functionality by considering it as two ether functionalities. Platts fragments approach for the peroxide functional group is currently not validated since no peroxides are included in the validation set, but for the future it will be interesting to regress a value for this functional group in addition to new radical groups. Another important molecule for oxidation processes is molecular oxygen for which Abraham solute descriptors are available in the database but these cannot be validated at the moment since there is no experimental or ab initio data available in the validation set nor in the open literature.

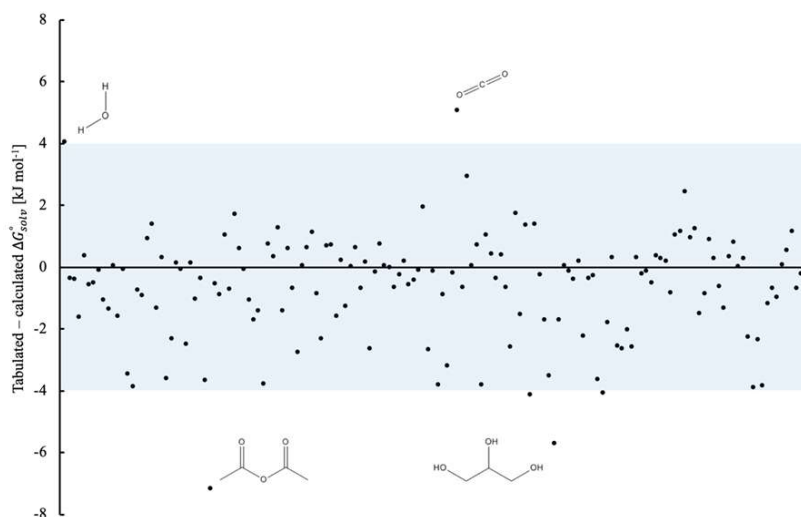


Figure 5-9: Plot of tabulated minus calculated values for ΔG_{solv}° using Platts fragments for solute descriptors and cyclohexane as solvent.

5.5.4. Improving the performance for carbon dioxide

Carbon dioxide was detected as an outlier in all solvents with a mean absolute error of 10.0 kJ mol⁻¹ and even a maximum error up to 31.8 kJ mol⁻¹. Clearly the Platts fragments method fails here, which is not a surprise since the carbon dioxide molecule is composed out of a carbon atom with two neighbors and two double bonded oxygen atoms. Considering that CO₂ is the only molecule for which two double bonded oxygen atoms are attached on one carbon atom, it can be expected that this molecular structure fails for Platts fragments. Solute descriptors for CO₂ are not available in the solute database of RMG and are neither found in the open literature. However, carbon dioxide is an important product formed during oxidations and it is therefore important to have an accurate estimation for its solvation Gibbs free energy. Since 16 data points are available from the COSMOtherm validation set it is possible to regress solute descriptors for the CO₂ molecule.

From the 16 data points, 10 data points are considered in the training set for regression of five solute descriptors and the 6 others are taken as test set. A multi-linear least squares regression, *cf.* equation (5-4), is performed in order to obtain the solute descriptors A, B, E, L and S.

$$RSSQ(\vec{b}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \xrightarrow{\vec{b}} Min \quad (5-4)$$

With $RSSQ(b)$ the residual sum of squares for a combination of solute descriptors, *i.e.* \vec{b} , y_i the observed value for data point i and \hat{y}_i the estimated value for data point i which depends on \vec{b} . The regressed values for the solute descriptors are given in Table 5-1 together with the descriptors calculated by the Platts fragments method, which differ significantly. The R² value amounts to 0.85 and from the corresponding correlation matrix it is detected that only the E and S descriptor are slightly correlated. The latter is acceptable due to the physical interpretation of the E and S descriptor.

Table 5-1: Solute descriptors for carbon dioxide: calculated with the Platts fragments method and regressed values via least squares regression.

	A	B	E	L	S
Platts fragments method	0.003	0.767	0.206	1.452	1.051
Regression	0.017	0.040	-1.133	0.480	1.101

A comparison of both obtained solute descriptors by means of a parity graph as depicted in Figure 5-10, which plots the calculated values in function of the tabulated values for both the data points of the training set (dots) and the test set (triangles). For optimal results, all points should be located on the first bisector. The bad estimation of the Platts fragments method in this case is clearly visible. However, the

regressed set of solute descriptors performs rather well, even describing the positive Gibbs free energy of solvation of carbon dioxide in water. Both the training set and the test set are predicted within an accuracy of 4 kJ mol⁻¹. The maximal error is observed for acetonitrile as solvent (part of the test set) with an error of 2.5 kJ mol⁻¹. The MAE of all data points and of the test set amounts to 0.4 kJ mol⁻¹ and 0.5 kJ mol⁻¹ respectively. The calculated descriptors are now included in the solute descriptor database and will be used for calculation of ΔG_{solv}° when experimental values are missing rather than the Platts fragments method.

When the physical meaning of the solute descriptors is linked back to the physical meaning, these are also well performing. Note that the A and B descriptor were respectively for the hydrogen bond donor and acceptor capacity. For carbon dioxide, no hydrogen atom is present which makes it impossible to have hydrogen bond donor capacities resulting in an A descriptor close to zero. The value of the B descriptor is slightly higher, since carbon dioxide has the possibility to accept hydrogen bonds. Moreover, the E and S descriptors are related to polarizability which is possible for carbon dioxide whereby these are significantly different from zero.

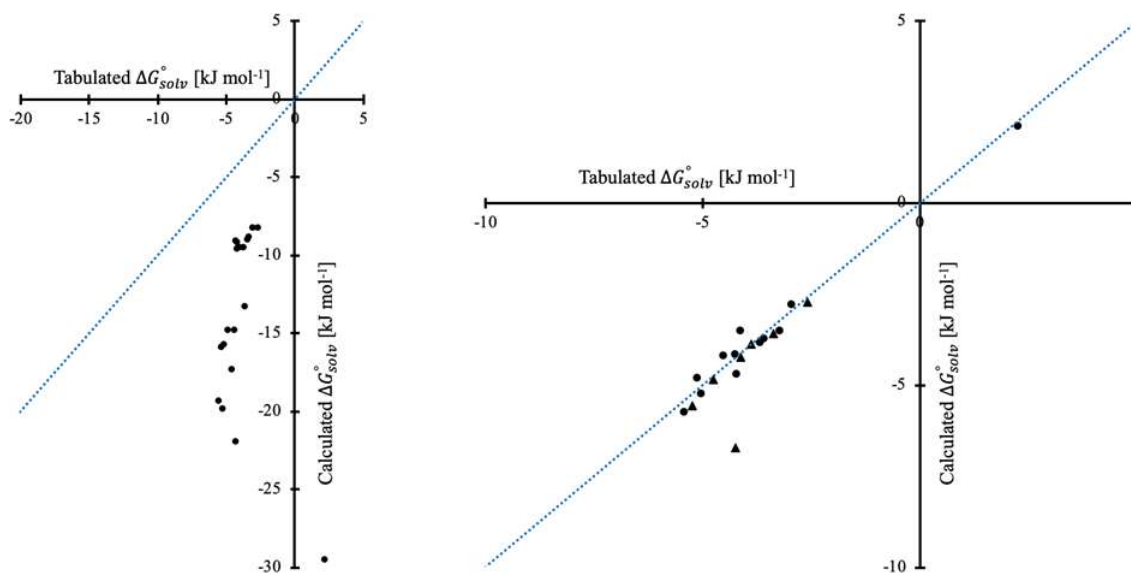


Figure 5-10: Parity plot of the tabulated ΔG_{solv}° versus calculated ΔG_{solv}° with the solute descriptors using Platts fragments method (Left). Parity plot of the tabulated ΔG_{solv}° versus calculated ΔG_{solv}° with self-regressed solute descriptors, dots (●) represent the training set, triangles (▲) the testing set (Right).

5.6. Conclusions

Extension of the gas-phase tailored framework for automatic kinetic model generation is necessary to allow model generation for liquid-phase processes, such as the liquid-phase oxidation of cyclohexane, with Genesys. All three pillars for kinetic model generation, *i.e.* molecules, reaction families and reaction rules, have to be redefined for liquid-phase chemistry. A lot of experimental and modelling research has already been performed concerning the industrial important liquid-phase oxidation of cyclohexane but modelling of the byproduct formation remains a challenging task.

In this master dissertation the first step is taken to introduce the liquid phase in Genesys, by focusing on the first pillar, *i.e.* molecules. An algorithm to calculate the Gibbs free energy of solvation for species in a liquid phase is implemented which will allow assignment of thermodynamic parameters to species in a solvent. The algorithm first searches the database with experimental and ab initio data if the combination of solute and solvent is tabulated. If not, an empirical correlation, *i.e.* the Abraham equation, is used to calculate the Gibbs free energy of solvation at standard conditions and 298 K. The empirical correlation requires solute and solvent descriptors. Solvent descriptors are tabulated for most common solvents. If solvent descriptors are not available in the database, the correlation cannot be used. Solute descriptors are tabulated for solutes, but in addition these can be calculated from a group additive scheme, *i.e.* Platts fragments method. The latter calculates the solute descriptors by summing up the contribution of each fragment the molecule is built from. It is impossible to calculate ΔG_{solv}° if solute descriptors are not tabulated and if not all non-hydrogen atoms can be assigned a contribution via Platts fragments.

After implementation, validation of both the Abraham equation with tabulated solute descriptors and calculated descriptors via Platts fragments method is performed. A validation set was used, composed out of COSMO*therm* data, based on ab initio calculations, and the FreeSolv database which contains a set of experimental determined values for the Gibbs free energy of hydration.

The mean absolute error (MAE) of the Gibbs free energy of solvation with tabulated solute descriptors amounts to 1.1 kJ mol⁻¹ with 1438 of the 4565 data points calculated spread over the 20 solvents represented in the data set. With Platts fragments much more solvation energies, 3549 of the 4565 data points, are calculated with a MAE of 2.5 kJ mol⁻¹ which is slightly higher than for the tabulated solute descriptors. Calculation of the Gibbs free energy of solvation with tabulated solute descriptors is thus preferred over calculated solute descriptors with the Platts fragment method. In addition, the polar solvents perform systematically worse than the apolar solvents. Except for carbon dioxide, the outlying data points, *i.e.* the solutes, over the different solvents are not the same.

Since cyclohexane is the solvent of interest in this master dissertation, the validation set was also limited to this solvent, which performs slightly better. A MAE of 0.9 and 1.3 kJ mol⁻¹ is reported respectively for calculation of ΔG_{solv}° with tabulated solute descriptors and solute descriptors calculated via Platts fragments. Of the 786 entries with cyclohexane as solvent in the validation set, 591 entries are calculated via the Platts fragments and only 195 via tabulated solute descriptors, *i.e.* 75.2 % and 24.8 % respectively. For carbon dioxide, the main outlier, molecular solute descriptors were regressed which decrease the MEA of the entries with carbon dioxide as solute from 10.0 kJ mol⁻¹ to 0.5 kJ mol⁻¹.

To enhance the performance, it is possible to regress values again for several solute and solvent descriptors and implement new Platts fragments, *e.g.* the missing peroxide functionality important for oxidation processes. Moreover, the Platts method has to be extended to allow calculation of the Gibbs free energy of solvation for radical species. Therefore, it will first be necessary to obtain accurate data for the solvation of radicals via *ab initio* calculations. Future work will then encompass regression to implement these new functionalities. Finally, it is of interest to enlarge the database with experimental data from literature, *ab initio* calculations, solute and solvent descriptors.

5.7. References

1. F. Eckert and A. Klamt, *Fast solvent screening via quantum chemistry: COSMO-RS approach*. *Aiche Journal*, 2002. **48**(2): p. 369-385.
2. A.P. Zeelenberg and H.W. de Bruijn, *Kinetics, mechanism and products of the gaseous oxidation of cyclohexane*. *Combustion and Flame*, 1965. **9**(3): p. 281-295.
3. D.L. Mobley and J.P. Guthrie, *FreeSolv: a database of experimental and calculated hydration free energies, with input files*. *Journal of Computer-Aided Molecular Design*, 2014. **28**(7): p. 711-720.
4. J.W.M. Steeman, S. Kaarsemaker, and P.J. Hoftyzer, *EI. A pilot plant study of the oxidation of cyclohexane with air under pressure*. *Chemical Engineering Science*, 1961. **14**(1): p. 139-149.
5. V.V. Lipes, Furman, M.S., *Influence of the feeding rate on the kinetics of liquid phase cyclohexane oxidation*. *Khimicheskaya Promyshlennost' Chemical industry*, 1964(8): p. 569-574 (in Russian).
6. K. Balcerzak, Janitz, A., Kowalska, W., Vieweger, T., Wielezynski, W., Zimowski, A., Zylbersztejn, M., Krzysztoforski, A., *Effect of residence time on the selectivity of oxidation of cyclohexane with air in an industrial reactor*. *Przemysk Chemiczny (Chemical Industry)*, 1982. **61**(8): p. 266-268 (in Polish).
7. M. Spielman, *Selectivity in hydrocarbon oxidation*. *AIChE Journal*, 1964. **10**(4): p. 496-501.
8. J. Alagy, P. Trambouze, and H. Van Landeghem, *Designing a Cyclohexane Oxidation Reactor*. *Industrial & Engineering Chemistry Process Design and Development*, 1974. **13**(4): p. 317-323.
9. T.V. Khar'kova, Arest-Yakubovich, I.L. and Lipes, V.V., *Kinetic model of the liquid-phase oxidation of cyclohexane. I. Homogeneous proceeding of the process*. *Kinetika i Kataliz (Kinetics and Catalysis)*, 1989. **30**(4): p. 954-958 (in Russian).
10. R. Pohorecki, J. Bałdyga, W. Moniuk, A. Krzysztoforski, and Z. Wójcik, *Liquid-phase oxidation of cyclohexane — modeling and industrial scale process simulation*. *Chemical Engineering Science*, 1992. **47**(9): p. 2559-2564.
11. R. Pohorecki, J. Bałdyga, W. Moniuk, W. Podgórska, A. Zdrójkowski, and P. T. Wierzychowski, *Kinetic model of cyclohexane oxidation*. *Chemical Engineering Science*, 2001. **56**(4): p. 1285-1291.
12. I. Hermans, P. Jacobs, and J. Peeters, *The Formation of Byproducts in the Autoxidation of Cyclohexane*. *Chemistry – A European Journal*, 2007. **13**(3): p. 754-761.
13. I. Hermans, P.A. Jacobs, and J. Peeters, *To the Core of Autocatalysis in Cyclohexane Autoxidation*. *Chemistry – A European Journal*, 2006. **12**(16): p. 4229-4240.
14. J.-C. Bradley, M.H. Abraham, W.E. Acree, Jr., A.S. Lang, S.N. Beck, D.A. Bulger, E.A. Clark, L.N. Condron, S.T. Costa, E.M. Curtin, S.B. Kurtu, M.I. Mangir, and M.J. McBride, *Determination of Abraham model solute descriptors for the monomeric and dimeric forms of trans-cinnamic acid using measured solubilities from the Open Notebook Science Challenge*. *Chemistry Central journal*, 2015. **9**: p. 11-11.
15. E. Qian, G. Lee, M. Che, L. Wang, D. Yue, R. Fischer, M. Jodray, A. Gupta, R. Neal, Y. Liu, W.E. Acree, and M.H. Abraham, *Determination of Abraham model solute descriptors for xanthone based on experimental solubility measurements at 298.2 K*. *Physics and Chemistry of Liquids*, 2019: p. 1-8.
16. E. Hart, A.M. Ramirez, S. Cheeran, M. Barrera, M.Y. Horton, A. Wadawadigi, W.E. Acree, and M.H. Abraham, *Determination of Abraham model solute descriptors for 2-methyl-3-nitrobenzoic acid from measured solubility data in alcohol, alkyl ether, alkyl acetate and 2-alkoxyalcohol mono-solvents*. *Physics and Chemistry of Liquids*, 2017. **55**(6): p. 796-804.

17. G. Lee, M. Che, E. Qian, L. Wang, A. Gupta, R. Neal, D. Yue, S. Downs, T. Mayes, O. Rose, W. Acree, and M. H. Abraham, *Determination of Abraham model solute descriptors for o-acetoacetanilide based on experimental solubility data in organic mono-solvents*. 2018. 1-8.
18. Brumfield, Wadawadigi, Kuprasertkul, Mehta, T. Stephens, Barrera, D. La Rosa, Kennemer, Meza, W. Acree, and Abraham, *Determination of Abraham model solute descriptors for three dichloronitrobenzenes from measured solubilities in organic solvents*. Vol. 53. 2015.
19. W. H. Green and R.H. West, *RMG Solvation database*, MIT, Editor. 2019.
20. J. Platts, D. Butina, M. H. Abraham, and A. Hersey, *Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach*. Vol. 39. 1999. 835-845.
21. A.V.K. Marenich, C. P.; Thompson, J. D.; Hawkins, G. D.; and C.C.G. Chambers, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., *Minnesota Solvation Database – version 2012*, U.o. Minnesota, Editor.: Minneapolis, 2012.
22. J.C. Bradley, M.H. Abraham, W.E. Acree, and A. Lang, *Predicting Abraham model solvent coefficients*. Chemistry Central Journal, 2015. **9**: p. 10.
23. D. Jiao and S.B. Rempe, *CO₂ solvation free energy using quasi-chemical theory*. Journal of Chemical Physics, 2011. **134**(22): p. 9.

6.

Conclusions and future outlook

6 . CONCLUSIONS AND FUTURE OUTLOOK.....	163
6.1. CONCLUSIONS	164
6.1.1. Cyclohexane gas-phase oxidation model	164
6.1.2. Fast assignment of liquid-phase thermodynamic parameters with Genesys	165
6.2. FUTURE OUTLOOK AND PERSPECTIVES	167
6.2.1. Cyclohexane gas-phase oxidation model	167
6.2.2. Extension of Genesys to the liquid phase	167

6.1. Conclusions

Nowadays, there is a strong drive to design new and optimize existing large-scale chemical processes in the pursuit of sustainability and increased profitability. Currently these topics are on the mindset of many people, considering that process optimization can result in significant energy savings and associated reduction of greenhouse gas emissions, more specifically CO₂ emissions, as well as reduction of waste streams. An accurate microkinetic model of the chemical process is therefore essential.

Currently, the limited fundamental understanding of reactions important in many liquid-phase processes does not allow automatic kinetic model generation for liquid-phase processes. This is in sharp contrast to gas-phase models which can be developed from first principles due to theoretical and computational advancements the last decade. In this master dissertation, the liquid-phase oxidation of cyclohexane to cyclohexanone and cyclohexanol was investigated, since it is an important industrial process for the synthesis of nylon-6 and nylon-6,6 which is only partially understood. A gas-phase model was constructed to prove the power of the automatic kinetic model generation gas-phase framework and to gain insight in the molecular reactivity between cyclohexane and molecular oxygen. Elements from the gas-phase model can later be reused for the development of the liquid-phase model, especially since literature indicates that the low-temperature oxidation zone shows the same type of reactions as during liquid-phase oxidation. Afterwards, Genesys was extended with an algorithm to calculate the Gibbs free energy of solvation as first step to include the liquid-phase effect. The latter is validated with experimental and ab initio data, and will allow fast assignment of thermodynamic properties to solute species.

6.1.1. Cyclohexane gas-phase oxidation model

A microkinetic model for the gas-phase oxidation of cyclohexane is developed which is focused on the low-temperature oxidation region. The reaction families are defined based on similarities to acyclic hydrocarbon oxidation. Thermodynamic and kinetic parameters are determined with ab initio calculations for the most important species and reaction pathways. For accurate thermodynamic parameter determination, it was observed that it is important to treat both axial and equatorial positioned substituents on the six-membered ring since the energy difference between the geometries can reach beyond 15 kJ mol⁻¹.

The potential energy surface (PES) for the addition of molecular oxygen to the cyclohexyl radical has been constructed by high-level ab initio calculations at the CBS-QB3 level of theory. From the PES, it is clear that the peroxide radical can be converted to a hydroperoxide by a hydrogen shift reaction with a barrier for the 1,5-shift and 1-6 shift of 108 kJ mol⁻¹ and 129 kJ mol⁻¹ respectively at 0 K. This barrier is below the reverse dissociation to molecular oxygen and the cyclohexyl radical for which an enthalpy

difference between reactants and products of 156 kJ mol^{-1} is calculated. The barrier for the 1,4-shift amounts to 164 kJ mol^{-1} which is slightly higher than the enthalpy of reaction of the dissociation reaction. It can be expected that the 1,5-shift will also occur because it is entropically favored. These formed hydroperoxide radicals can react further mainly through cyclic ether formation and β -scission reactions. Alternatively, oxygen addition on the hydroperoxide radicals leads to chain branching, which is not shown on the PES.

Besides the hydrogen shifts, also hydroperoxide elimination can take place with formation of cyclohexene for which the activation barrier amounts to 134 kJ mol^{-1} , which stays under the reaction enthalpy of dissociation to molecular oxygen and the cyclohexyl radical. The unimolecular decomposition reaction considered with the lowest enthalpy of formation is a 1,3-shift for which 113 kJ mol^{-1} is released due to the formation of the energetically favorable cyclohexanone and a hydroxyl radical. However, the activation energy of 165 kJ mol^{-1} makes proceeding of this reaction unlikely. Other hydrogen shifts, carbon-carbon and carbon-hydrogen β -scissions do not play an important role except for one β -scission possible after the 1,5-shift which forms hex-5-enal and a hydroxyl radical. This reaction has an energy barrier equal to 162 kJ mol^{-1} with an enthalpy change of -14 kJ mol^{-1} which is even more favorable due to ring-opening.

The developed microkinetic model of Genesys is merged with AramcoMech 2.0 and validated with experiments performed in a jet-stirred reactor at l'Université de Lorraine by the group of Prof. Battin-Leclerc. The model predicts well the experimental trends in the low-temperature oxidation region when the alternative activation energy of the modified Arrhenius equation for the mentioned β -scission is lowered from 24.2 kJ mol^{-1} to 12.2 kJ mol^{-1} . The match between experiments and the microkinetic model confirms the power of the first principles based automatic kinetic model generation framework encoded in Genesys for the gas phase, which can now be extended to allow liquid-phase process modelling.

6.1.2. Fast assignment of liquid-phase thermodynamic parameters with Genesys

Currently, the applications of the microkinetic model generation framework are limited to gas-phase processes. With this master dissertation the first step has been taken to introduce liquid-phase effects in Genesys by considering the fast assignment of thermodynamic properties to solutes, in order to extend the application domain to liquid-phase processes.

A new algorithm is encoded which assigns the Gibbs free energy of solvation based on the available data in the database. The algorithm first searches the database with experimental and *ab initio* data if the combination of solute and solvent is tabulated. If not, an empirical correlation, *i.e.* the Abraham equation, has been implemented in Genesys to calculate the Gibbs free energy of solvation at 298 K and

standard conditions. The empirical correlation requires solute and solvent descriptors. Solvent descriptors are tabulated for most used solvents. Solute descriptors are tabulated for a lot of solutes, but in addition these can be calculated from a group additive scheme, *i.e.* Platts fragments method. It is impossible to calculate the Gibbs free energy of solvation if solute descriptors are not tabulated and if not all non-hydrogen atoms can be assigned a contribution via Platts fragments.

Besides the implementation of an algorithm, both the Abraham equation and the Platts fragments method are validated with a data set containing Gibbs free energies of solvation from *ab initio* calculations, exported from COSMOtherm, and experiments from the FreeSolv database. The mean absolute error (MAE) of the Gibbs free energy of solvation calculated via the Abraham equation with tabulated solute descriptors amounts to 1.1 kJ mol⁻¹ with 1438 of the 4565 data points calculated. With Platts fragments much more solvation energies, *i.e.* 3549 of the 4565 data points, are calculated with a MAE of 2.5 kJ mol⁻¹ which is slightly higher than for the tabulated solute descriptors. Calculation of the Gibbs free energy of solvation with tabulated solute descriptors is thus preferred over calculated solute descriptors with the Platts fragments method. A total of 20 solvents are considered in the validation set, whereby the polar solvents perform systematically worse than the apolar solvents for the Platts fragments method. Except for carbon dioxide, the discrepancies over the different solvents are not the same. Cyclohexane as solvent was also assessed with a MAE of 0.9 kJ mol⁻¹ and 1.3 kJ mol⁻¹ respectively for calculation of the Gibbs free energy of solvation with tabulated solute descriptors and solute descriptors calculated via Platts fragments. This outcome is a good basis for modelling of processes with cyclohexane as solvent. Due to the bad performance of Platts fragments for carbon dioxide, molecular solute descriptors were regressed which decrease the MAE of the entries with carbon dioxide as solute from 10.0 kJ mol⁻¹ to 0.5 kJ mol⁻¹.

6.2. Future outlook and perspectives

The future work encompasses two parts. Further improvement of the gas-phase oxidation model for cyclohexane and extension of the gas-phase tailored framework of Genesys to the liquid phase.

6.2.1. Cyclohexane gas-phase oxidation model

The developed gas-phase microkinetic model can predict well the formation of most products, *e.g.* carbon monoxide and ethylene, in the low-temperature oxidation region which is of interest. Though, some species should be considered in more detail and to make the model also valid in the high-temperature region extra *ab initio* calculations are required. For this improved model, a sensitivity and product formation analysis should be executed which will give a more quantitative overview of the importance of several reactions in the gas phase. However, it is confirmed that the gas-phase tailored framework of Genesys functions well after adaptation of only one single parameter, which can therefore be extended to the liquid phase. A lot of knowledge from the gas-phase model can be reused for construction of a liquid-phase model and in addition the focus will be on the formation of cyclohexanone and cyclohexanol which are currently not present in the gas-phase model.

6.2.2. Extension of Genesys to the liquid phase

The algorithm for calculation of the Gibbs free energy of solvation performs very well for the largest part of the validation set. To enhance the performance even more, it is possible to regress values again for several solute and solvent descriptors and implement new Platts fragments, *e.g.* the missing peroxide functionality important for oxidation processes. Currently, the Platts fragments method cannot calculate the solute descriptors for radicals due to missing fragments which should also be regressed since radicals are important reaction intermediates during oxidation. Therefore, it will first be necessary to obtain accurate data for the solvation of radicals via high-level *ab initio* calculations. Moreover, it is of interest to enlarge the Genesys liquid-phase database with experimental data from literature, *e.g.* the Minnesota database and SAMPL data set, in addition to *ab initio* calculated data, solute and solvent descriptors. Future work encompasses the integration of the Gibbs free energy of solvation in the overall thermodynamic calculation scheme, amongst others by implementation of the Mintz equation, to allow assignment of the enthalpy, entropy and heat capacities to solute and solvent species.

Further research is still ongoing to extend Genesys to a broader application domain and improve the performance of the automatic kinetic model generation framework. In the end, fully automatic model generation for gas-phase, liquid-phase and catalyzed processes is envisioned.

A.

Activity coefficient models

A . ACTIVITY COEFFICIENT MODELS	169
A.1. STANDARD AND REFERENCE STATES	171
A.2. UNIQUAC MODEL	172
A.3. NRTL MODEL	174
A.4. UNIFAC MODEL	175
A.5. REFERENCES	177

An activity coefficient is a factor used in thermodynamics to describe the deviation from ideal behavior in a mixture of chemical substances. In an ideal mixture, the intermolecular interactions between each pair of chemical species are the same, and as a result the thermodynamic properties of the mixture can be described directly in terms of concentrations or the partial pressure, e.g. Henry's or Raoult's law [1]. Deviations from this ideal behaviour are accommodated by modifying the concentration by an activity coefficient. The chemical potential of a species A in an ideal mixture of liquids is given by equation (A-1).

$$\mu_A = \mu_A^\circ + RT \ln(x_A) \quad (\text{A-1})$$

where μ_A° is the chemical potential of the pure substance A, x_A is the mole fraction of species A in the mixture, R is the gas constant and T the absolute temperature. Non-ideal behavior is described in a very similar way, cf. equation (A-2), by introduction of the activity of a species A, i.e. a_A .

$$\mu_A = \mu_A^\circ + RT \ln(a_A) \quad (\text{A-2})$$

with the activity, i.e. a_A , defined as the product of the mole fraction and the activity coefficient, i.e. γ_a , for a non-ideal mixture of liquids, cf. equation (A-3).

$$a_A = x_a \cdot \gamma_a \quad (\text{A-3})$$

The activity can be considered as a pseudo or effective mole fractions that carries the composition dependency of μ_A . Moreover, the activity is dimensionless and in general it also depends on the temperature and pressure. Similar to the activity, the fugacity can be used to calculate the chemical potential of a non-ideal liquid mixture according to equation (A-4) [2].

$$\mu_A = \mu_A^\circ + RT \ln\left(\frac{f_A}{f_A^\circ}\right) \quad (\text{A-4})$$

in which f_A is the fugacity of species A and f_A° the reference fugacity of species A in the standard state (related to the standard conditions of μ_A°). From which directly the relation between activity and fugacity follows, cf. equation (A-5).

$$a_A = \frac{f_A}{f_A^\circ} \quad (\text{A-5})$$

In solvation models, frequently the Gibbs free energy of solvation is estimated for the transition from an ideal gas to an ideal solvent. Application of an activity coefficient model can take into account the difference of the non-ideality and allows one to estimate more accurately the change in free energy. Several activity coefficient models have been published in literature and are frequently used in modelling software, some important ones are covered in this chapter.

A.1. Standard and reference states

Several different standard and reference states are commonly used in the literature, so care must be taken when comparing data from different sources. Moreover, standard state issues immediately appear when one attempts to compare computed dilute-limit pure-solvent solvation energies with experimental data or when attempting to use these solvation energies to make predictions for real solvent mixtures [3].

There are several different commonly used standard states for the solutions phase. For problems involving gas-liquid equilibria, all species are mostly referenced to by the gas phase standard state, i.e. a pressure of 1 atmosphere or thus $C_0 \approx 0.045 \text{ mol l}^{-1}$ for ideal gases. In aqueous solutions, a 1 mol l^{-1} is often assumed for the solute molecules. For solutes that can exist also as a pure liquid, the pure standard state, cf. equation (A-6) with ρ the density of the fluid and MM the molar mass, is often used.

$$C_0 = \frac{\rho}{MM} \quad (\text{A-6})$$

For non-ionic liquid systems, there are two frequently used reference states. The first one describes the limiting behavior as the mole fraction of the species approaches zero, the so-called infinite dilution state, which is referred to as Henry's Law state frequently denoted with the ' ∞ ' symbol. The other reference state assumes an ideal-solution behavior and is known as Raoult's Law. Henry's law and the ideal-solution reference fugacity are respectively given by equation (A-7) and equation (A-8).

$$f_{solute}^{\infty}(x_{solute}) = \left. \frac{\partial f_{solute}^{true}}{\partial x_{solute}} \right|_{x_{solute} \rightarrow 0} \cdot x_{solute} = f_{solute}^{\infty} \cdot x_{solute} \quad (\text{A-7})$$

$$f_{solute}(x_{solute}) = f_{solute}^{pure} \cdot x_{solute} \quad (\text{A-8})$$

The activity coefficients can be reported to either of these reference states, e.g. for the Henry's Law reference state it is indicated in equation (A-9) with x_{solute}^* the mole fraction in the mixture of interest and $\gamma^{\infty \rightarrow *}$ the activity coefficient.

$$\gamma^{\infty \rightarrow *} = \frac{f_{solute}^{real}(x_{solute}^* \text{ in real mixture})}{f_{solute}^{\infty}(x_{solute}^*)} = \frac{f_{solute}^{real}(x_{solute}^* \text{ in real mixture})}{\left. \frac{\partial f_{solute}^{true}}{\partial x_{solute}} \right|_{x_{solute} \rightarrow 0} \cdot x_{solute}^*} \quad (\text{A-9})$$

The activity coefficient can be obtained from activity coefficient models, which can then be used with the previous formulas to calculate the chemical potential for the species in the mixture of interest.

A.2. UNIQUAC model

UNIQUAC, originating from universal quasichemical, is an empirical-based model to estimate the activity coefficients of components related to their molar fraction for description of phase equilibria, i.e. liquid-solid, liquid-liquid or liquid-vapor equilibria [4]. The model is derived from a first order approximation of interacting molecules on a lattice with statistical thermodynamics. In the derivation, it is assumed that the Wilson hypothesis is true, i.e. the concentration around a molecule is independent of the local concentration around another type of molecule. Formulated in another manner: the local concentration around a certain molecule is different than the bulk concentration. This difference originates from the difference in interaction energy of the central molecule with molecules of its own kind and that with the molecules of the other kind. The corresponding energy difference introduces a non-randomness at the molecular level, which is also referred to by the local composition theory. Due to this assumption, these models are not thermodynamically consistent for a real one-component fluid [5]. Moreover, the UNIQUAC model is based on quasi-chemical theory which calculates the configurational entropy change taking into account the non-random distribution of molecules and the distribution of molecules with different sizes. The UNIQUAC model is considered as a second-generation activity coefficient model since the expression for the excess Gibbs free energy consists of an entropy and an enthalpy contribution, cf. equation (A-10).

$$\ln(\gamma_i) = \ln(\gamma_i^C) + \ln(\gamma_i^R) \quad (\text{A-10})$$

The entropic term, i.e. $\ln(\gamma_i^C)$, describes the deviation from ideal solubility of the two-component mixture. Differences in the molecular shape of the components affect the entropy and the accompanying contribution is calculated via combinatorial techniques based on the lattice theory. In equation (A-11), the formula is given to determine the entropic contribution for a component i .

$$\ln(\gamma_i^C) = (1 - V_i + \ln(V_i) - \frac{z}{2} q_i \left(1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right) \right)) \quad (\text{A-11})$$

In which V_i represents the volume fraction per mixture mole fraction for component i , which is calculated by equation (A-12).

$$V_i = \frac{x_i r_i}{\sum_j x_j r_j} \quad (\text{A-12})$$

And the surface area fraction per mixture mole fraction, i.e. F_i , for component i is given by equation (A-13).

$$F_i = \frac{x_i q_i}{\sum_j x_j q_j} \quad (\text{A-13})$$

With r_i the relative Van der Waals volumes, q_i the surface area of the pure chemical component i and z the coordination number and x_i the mole fraction of component i . The residual or enthalpic contribution, i.e. $\ln(\gamma_i^R)$, accounts for the change in interacting forces between different molecules upon mixing and is calculated by equation (A-14).

$$\ln(\gamma_i^R) = q_i \left(1 - \ln \left(\frac{\sum_j x_j q_j \tau_{ji}}{\sum_j x_j q_j} \right) - \sum_j \frac{x_j q_j \tau_{ij}}{\sum_k x_k q_k \tau_{kj}} \right) \quad (\text{A-14})$$

With τ_{ij} defined in equation (A-15) in which Δu_{ij} represents the binary interaction energy parameter [J mol⁻¹] given in equation (A-16).

$$\tau_{ij} = \exp \left(-\frac{\Delta u_{ij}}{RT} \right) \quad (\text{A-15})$$

$$\Delta u_{ij} = u_{ij} - u_{ii} \quad (\text{A-16})$$

Where u_{ij} represents the reaction energy between molecules i and j .

The derived equations in the model rely upon two basic underlying parameters, i.e. the relative external surface area, i.e. q_i , and a volume parameter, i.e. r_i , as chemical constants which are to be obtained from binary mixture experiments and fitting of the model to the experimental data. An advantage is that for multicomponent mixtures, it suffices to dispose of all binary mixture parameters such that ternary (or higher) parameters are not required. The UNIQUAC model gives a good representation of vapor-liquid and liquid-liquid equilibria for binary and multicomponent mixtures containing a variety of nonelectrolyte components including azeotropes, e.g. alcohols, ketones, aldehydes and water. More recently COSMOSPACE [6] has been developed which is based on the same theoretical background of UNIQUAC, i.e. a lattice model, but takes more interactions into account and thereby outperforms the original model in the description of liquid-vapor and liquid-liquid equilibria.

A.3. NRTL model

The non-random two-liquid model or short the NRTL, is another activity coefficient model that relates the activity coefficient to the molar fraction of a compound in the liquid phase of interest [7]. The NRTL method is the predecessor of the UNIQUAC model and in contrast to that model only consists of enthalpic contributions. Similar to UNIQUAC, the model parameters are obtained by fitting model parameters to experimentally obtained activity coefficients. In Figure 1, an example is given of the fit of the NRTL model to experimental data for an azeotropic mixture, as well as extrapolation to other pressures. Care has to be taken since for some binary mixtures, multiple sets of NRTL parameters have been published in literature. One should then have a look at the assumptions made for the parameter set, e.g. the saturated vapor pressure of pure compounds used and if the gas phase is treated as ideal or non-ideal. When description of phase equilibria over a large temperature regime is necessary, it is important to take into account that the interactions between solute molecules mutually and between solute - solvent molecules are temperature dependent. This observation is not considered in the derivation and therefore, two formats are frequently applied to correct for it, i.e. the extended Antoine equation [8] or a second order polynomial format as shown in equation (A-17) and equation (A-18) respectively.

$$f(T) = A + \frac{B}{T} + C \cdot \ln(T) + D \cdot T \quad (\text{A-17})$$

$$f(T) = A + B \cdot T + C \cdot T^2 \quad (\text{A-18})$$

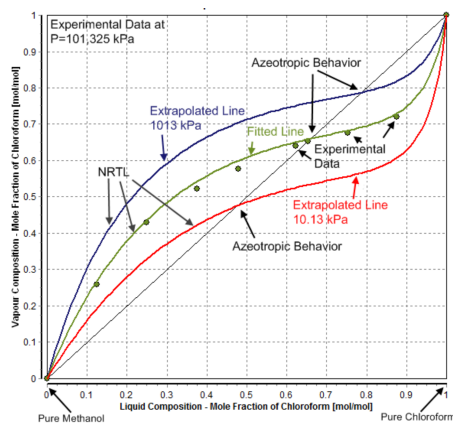


Figure 1: Vapor-Liquid equilibrium of the mixture of chloroform and methanol, NRTL fit and experimental data, and extrapolation to different pressures [9].

A.4. UNIFAC model

The UNIFAC model, UNIQUAC Functional-group Activity Coefficients, is an extension of the UNIQUAC model to predict the activity coefficient of non-electrolyte compounds in non-ideal mixtures [10]. UNIFAC uses the functional groups in the components that make up the liquid mixture to estimate the two necessary parameters, i.e. the external surface area and the volume parameter, to calculate the activity coefficient. Using contributions to account for each of the functional groups as well as the interaction between each pair of functional groups present in the molecules, the activity of each of the components can be estimated. The functional group contributions are obtained by experiments after which regression has been performed. For new compounds and new mixtures, this approach of working is more interesting since it is not necessary to perform experiments. Currently, this method is still updated with the introduction of new functional groups and the accompanying interactions with other groups. The method has become really popular by its implementation in lot of process simulation software in combination with its simple usage, reliable outcomes and large range of applicability. Though, simultaneous estimation of vapor-liquid equilibrium data and excess enthalpies is not possible with the required accuracy which means that the temperature dependency, described by the Gibbs-Helmholtz equation, is not described correctly. Moreover, poor results are obtained in the diluted regions as well for asymmetric systems, i.e. mixtures containing molecules of very different sizes and shapes. To solve these drawbacks, a new model was developed the so-called Modified UNIFAC (Dortmund) [11, 12]. By introduction of extra parameters to account for temperature dependency, adaptation of the underlying formulas and again perform regression with an extended set of data, the cons of the UNIFAC model form less of a problem. Despite the better results of the latter model, both models are next to each other further developed and each year new functional groups are introduced to both.

Activity is a central quantity in thermodynamics since it unifies the description of several relevant functions and properties such as concentrations in the liquid and gas phase as well as it is directly related to the chemical potential. For real mixtures, the activity coefficient embodies a lot of information concerning the solute-solvent and solute-solute interactions. The activity coefficient is thereby directly linked to the solvation of a species in a solvent. An accurate estimation of the activity coefficient is thus an important step for correct description of the solvation for real non-ideal mixtures. The importance of an activity coefficient model has already been indicated in the thermochemical cycle in the literature survey chapter, where it plays an important role in the last step where the solution evolves from the infinite dilution state to the actual concentration in a non-ideal solvent.

Models for activity coefficients are frequently used for prediction of simple phase equilibria or for estimation of related physical properties, e.g. the viscosity of mixtures. Process simulation programs, e.g. ASPEN PLUS [13], are commonly featured with activity coefficient models to calculate mass

balances in and around separation units like distillation columns and liquid-liquid extractions. For electrolyte systems, an adapted version of the NRTL mixture has been developed, i.e. Electrolyte NRTL [14].

A.5. References

1. Z. Hens, *Fysische Chemie: Chemische thermodynamica*. First ed. 2010: Academia Press. 224.
2. J.D. Ramshaw, *Fugacity and Activity in a Nutshell*. Journal of Chemical Education, 1995. **72**(7): p. 601.
3. A. Jalan, R.W. Ashcraft, R.H. West, and W.H. Green, *Predicting solvation energies for kinetic modeling*. Annual Reports Section "C" (Physical Chemistry), 2010. **106**(0): p. 211-258.
4. D.S. Abrams and J.M. Prausnitz, Statistical thermodynamics of liquid-mixtures – New expression for excess Gibbs energy of partly or completely miscible systems. Aiche Journal, 1975. **21**(1): p. 116-128.
5. G. Maurer and J.M. Prausnitz, *On the derivation and extension of the uniquac equation*. Fluid Phase Equilibria, 1978. **2**(2): p. 91-99.
6. A. Klamt, G.J.P. Krooshof, and R. Taylor, *COSMOSPACE: Alternative to conventional activity-coefficient models*. Aiche Journal, 2002. **48**(10): p. 2332-2349.
7. H. Renon and J.M. Prausnitz, Local compositions in thermodynamic excess functions for liquid mixtures. Aiche Journal, 1968. **14**(1): p. 135-+.
8. A. C., *Vapor Pressure: a new relationship between pressure and temperature*. Comptes Rendus des Séances de l'Académie des Sciences, 1888. **107**: p. 681–684, 778–780, 836–837.
9. Wikipedia. *Non-random two-liquid model*. 2019; Available from: https://en.wikipedia.org/wiki/Non-random_two-liquid_model.
10. A. Fredenslund, R.L. Jones, and J.M. Prausnitz, Group-contribution estimation of activity-coefficients in non-ideal liquid-mixtures. Aiche Journal, 1975. **21**(6): p. 1086-1099.
11. J. Lohmann, R. Joh, and J. Gmehling, *From UNIFAC to modified UNIFAC (Dortmund)*. Industrial & Engineering Chemistry Research, 2001. **40**(3): p. 957-964.
12. J. Gmehling, J. Lohmann, A. Jakob, J.D. Li, and R. Joh, *A modified UNIFAC (Dortmund) model. 3. Revision and extension*. Industrial & Engineering Chemistry Research, 1998. **37**(12): p. 4876-4882.
13. aspentech. *Technology That Loves Complexity*. Available from: <https://www.aspentech.com/>.
14. C.C. Chen and Y.H. Song, *Generalized electrolyte-NRTL model for mixed-solvent electrolyte systems*. Aiche Journal, 2004. **50**(8): p. 1928-1941.



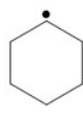

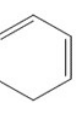

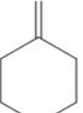
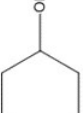
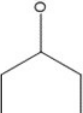
B.

Thermodynamic parameters from ab initio calculations

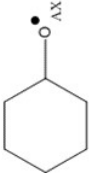
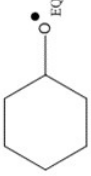
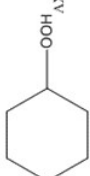
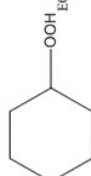
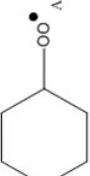
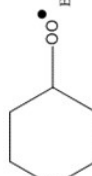
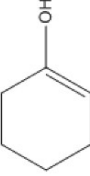
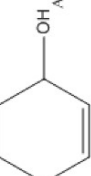
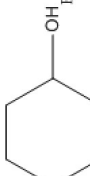
B . THERMODYNAMIC PARAMETERS FROM AB INITIO CALCULATIONS	179
---	------------

Appendix B: Thermodynamic parameters from ab initio calculations

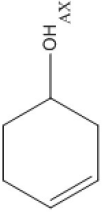
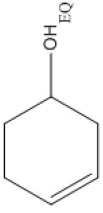
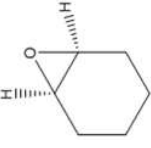
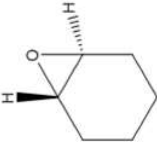

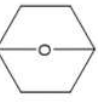
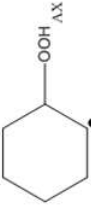
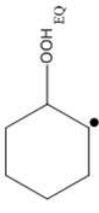
Table B-1: Thermodynamic parameters introduced in the Genesys species database, calculated by high-level ab initio calculations at the CBS-QB3 level of theory.

SMILES	MOLECULAR REPRESENTATION	$\Delta_f H_{298\text{ K}}$ [kJ mol ⁻¹]	$S_{298\text{ K}}$ [J mol ⁻¹ K ⁻¹]	C_p [J mol ⁻¹ K ⁻¹]						
				300 K	400 K	500 K	600 K	800 K	1000 K	1500 K
C1CCCCC1		-120.40	292.44	106.19	147.30	185.07	217.24	267.21	303.50	358.41
		-93.68	300.12	107.29	147.95	185.47	217.46	267.19	303.36	358.21
C1[CH]CCCC1		78.38	318.87	109.09	147.63	182.52	211.99	257.49	290.44	340.27
C1C=CCCC1		-1.04	304.11	100.19	137.11	170.40	198.43	241.54	272.61	319.44
C1C=CC=CC1		-120.54	326.32	107.50	142.10	172.56	197.90	236.29	263.51	303.91
		-107.96	331.15	108.67	143.30	173.64	198.81	236.92	263.96	304.14
C1C(=O)CCCC1		-227.66	328.14	114.47	153.26	188.09	217.41	262.39	294.65	343.96
C1C(O)CCCC1		-267.58	345.08	124.66	166.68	204.91	237.23	287.02	323.03	377.79
		-268.72	347.99	124.38	166.39	204.64	236.99	286.87	322.94	377.77

Appendix B: Thermodynamic parameters from ab initio calculations

<chem>C1C([O])CCCC1</chem>		-59.82	337.67	119.70	161.70	199.15	230.44	278.19	312.38	363.67
		-61.55	338.34	119.75	161.49	198.88	230.18	278.00	312.26	363.63
<chem>C1C(OO)CCCC1</chem>		-212.83	373.53	141.39	187.40	228.15	261.86	312.52	348.30	401.76
		-213.12	378.41	142.57	187.03	226.95	260.31	310.97	347.04	401.11
<chem>C1C(O[O])CCCC1</chem>		-72.84	369.58	134.49	177.64	215.96	247.92	296.58	331.34	383.41
		-73.51	374.38	133.17	175.78	214.12	246.31	295.45	330.56	383.06
<chem>C1C(O)=CCCC1</chem>		-183.43	334.76	124.04	163.46	197.13	224.65	266.09	295.72	340.76
<chem>C1C(O)C=CCC1</chem>		-163.30	346.91	122.70	159.86	192.90	220.52	262.69	293.04	339.22
		-162.37	346.88	124.93	160.74	192.99	220.26	262.69	293.04	339.22

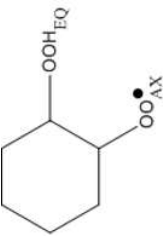
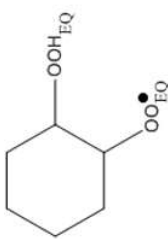
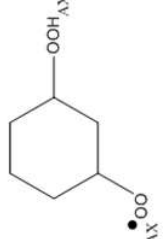
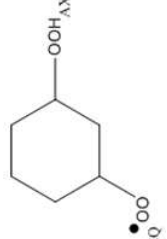
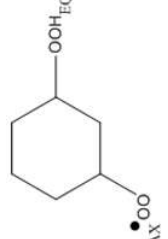
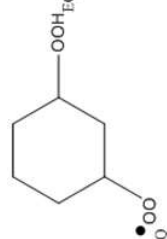
Appendix B: Thermodynamic parameters from ab initio calculations


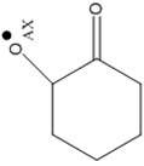
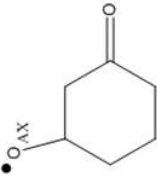
C1C(O)CC=CC1		-166.51	347.11	121.39	158.36	191.58	219.42	261.94	292.51	338.96
		-166.49	350.29	118.35	156.20	189.97	218.20	261.22	292.08	338.83
C1C(O2)C2CCCC1		-116.74	316.58	109.68	150.35	186.53	216.69	262.38	294.81	343.10
		60.78	318.41	112.93	153.42	189.24	218.94	263.81	295.70	343.38
C1C(O2)CC2CC1		-97.20	310.05	106.27	148.39	185.63	216.33	262.35	294.80	343.03
C1C(O2)CCC2C1		-179.32	313.10	104.03	147.24	185.13	216.12	262.25	294.64	342.79
C1C(OO)[CH]CCC1		-10.34	384.29	146.89	189.09	225.91	256.24	301.87	334.26	382.26
		-7.82	390.20	144.85	186.71	223.83	254.60	301.00	333.86	382.98

Appendix B: Thermodynamic parameters from ab initio calculations

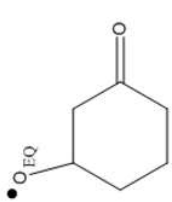
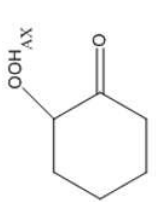
C1C(OO)C[CH]CC1		-12.90	386.24	144.78	188.15	225.91	256.81	302.86	335.26	383.64
		-14.88	390.33	144.94	187.00	224.13	254.87	301.16	333.94	382.97
		-13.63	383.99	144.14	187.69	225.57	256.55	302.68	335.11	383.51
		-13.10	389.70	145.61	187.50	224.51	255.15	301.31	334.04	383.01
C1C(OO)(O[O])CCCC1		-186.88	407.29	173.64	223.41	265.77	300.01	349.82	383.69	432.87
		-186.07	409.46	177.23	229.12	273.32	308.30	357.28	389.26	434.37
C1C(OO)C(O[O])CCC1		-157.74	436.79	178.83	223.58	262.64	294.79	342.87	376.67	426.99
		-155.36	422.64	177.48	225.47	265.90	298.46	346.37	379.70	428.90

Appendix B: Thermodynamic parameters from ab initio calculations

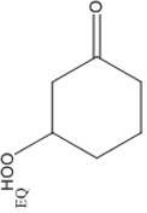

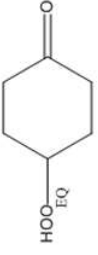
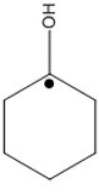
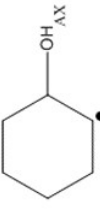
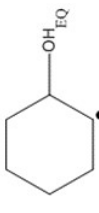
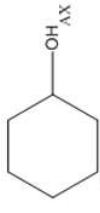
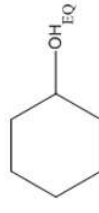
      C1C(OO)CC(O[O])CC1	-155.92	422.19	177.10	225.23	265.74	298.50	346.49	379.99	429.01
	-155.82	429.41	181.29	226.22	264.95	296.66	344.03	377.41	427.24
	-170.79	437.59	169.05	216.48	257.78	291.52	341.46	376.21	427.35
	-164.48	455.09	170.34	217.03	257.78	291.14	340.65	375.21	426.35
	-164.01	454.76	171.35	217.24	257.48	290.61	340.10	374.80	426.19
	-161.85	448.83	168.48	214.71	255.42	288.99	339.11	374.19	426.02

<chem>C1C(OO)CCC(O[O])C1</chem> 	-	-	-	-	-	-	-	-	-
	-165.01	434.95	168.79	216.22	257.50	291.20	341.03	375.65	426.70
	-163.29	439.30	171.51	217.69	257.98	291.07	340.42	375.01	426.28
	-160.81	443.30	169.11	215.37	255.76	289.38	339.48	374.65	426.13
<chem>C1C(=O)C([O])CCCC1</chem> 	-164.74	358.17	128.52	167.64	201.85	230.18	272.97	303.25	348.11
	-145.87	365.08	130.06	169.05	203.10	231.30	273.92	304.07	348.68
	-160.82	358.61	128.10	167.74	202.23	230.67	273.46	303.65	348.34
<chem>C1C(=O)CC([O])CC1</chem> 									

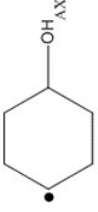
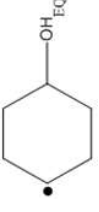
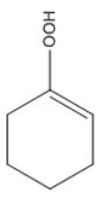
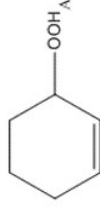
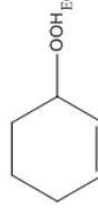
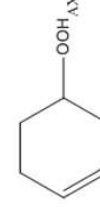
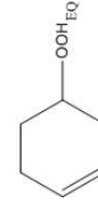
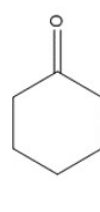
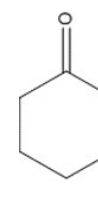
Appendix B: Thermodynamic parameters from ab initio calculations

 <chem>C1C(=O)CCC([O])C1</chem>	-162.37	360.73	128.82	168.45	202.96	231.40	274.14	304.26	348.76
	-163.97	358.32	127.78	167.46	202.01	230.49	273.32	303.52	348.24
	-161.82	359.43	128.20	167.52	201.92	230.37	273.24	303.50	348.27
 <chem>C1C(=O)C(OO)CCC1</chem>	-373.57	403.21	158.46	200.44	236.44	265.62	308.94	339.52	385.6
	-382.74	375.05	146.67	190.33	228.43	259.90	307.26	340.68	390.41
	-388.84	390.36	150.52	194.10	231.78	262.63	308.55	340.74	388.63

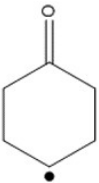
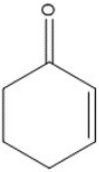
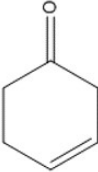
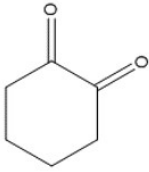
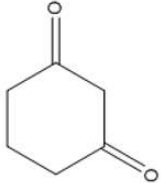

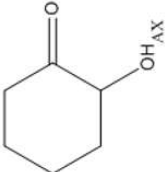
Appendix B: Thermodynamic parameters from ab initio calculations

	-381.40	402.32	150.20	192.43	229.55	260.23	306.2	338.44	386.04
<chem>C1C(=O)CCC(OO)C1</chem> 	-384.69	394.81	151.00	194.02	231.41	262.00	307.48	339.24	386.23
	-369.74	462.04	184.71	225.99	261.61	290.55	333.12	362.65	406.39
<chem>C1[C](O)CCCC1</chem> 	-109.38	346.03	125.25	166.02	202.34	232.60	278.54	311.33	360.74
<chem>C1C(O)[CH]CCC1</chem> 	-84.10	357.86	131.47	169.76	204.28	233.33	277.98	310.30	359.68
	-84.12	354.97	132.05	169.88	204.14	233.13	277.87	310.31	359.80
<chem>C1C(O)[CH]CC1</chem> 	-87.41	355.31	130.38	169.14	203.98	233.23	278.07	310.46	359.85
	-88.29	359.82	127.51	166.90	202.22	231.85	277.22	309.93	359.65

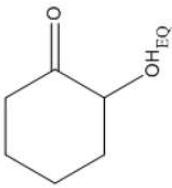
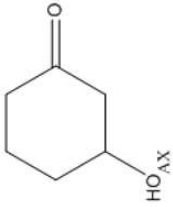
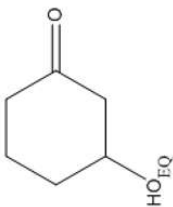



Appendix B: Thermodynamic parameters from ab initio calculations

<chem>C1C(O)CC[CH]C1</chem>		-86.87	356.33	127.63	167.08	202.40	232.01	277.31	309.97	359.64
		-87.04	359.23	127.17	166.63	202.01	231.68	277.12	309.87	359.63
<chem>C1C(OO)=CCCC1</chem>		-95.08	379.30	140.74	179.05	212.95	241.14	283.94	314.54	360.76
<chem>C1C(OO)C=CCC1</chem>		-89.51	372.44	138.33	178.67	213.86	242.56	286.25	316.89	362.57
		-85.02	378.79	136.02	176.45	212.00	241.32	285.20	316.07	362.03
<chem>C1C(OO)CC=CC1</chem>		-91.43	369.08	135.31	177.30	213.58	243.10	286.81	317.36	362.80
		-91.87	375.76	135.85	176.30	211.87	241.20	285.15	316.09	362.16
<chem>C1C(=O)[CH]CCC1</chem>		-62.78	338.24	113.72	150.64	183.27	210.43	251.63	280.91	324.51
<chem>C1C(=O)C[CH]CC1</chem>		-30.24	349.12	118.25	154.23	186.02	212.53	252.93	281.79	324.93

Appendix B: Thermodynamic parameters from ab initio calculations

<chem>C1C(=O)CC[CH]C1</chem>		-26.72	339.98	117.38	153.59	185.55	212.16	252.66	281.57	324.79
<chem>C1C(=O)C=CCC1</chem>		-120.54	326.32	107.50	142.10	172.56	197.90	236.29	263.51	303.91
<chem>C1C(=O)CC=CC1</chem>		-107.96	331.15	108.67	143.30	173.64	198.81	236.92	263.96	304.14
<chem>C1C(=O)C(=O)CCCC1</chem>		-295.52	349.73	123.22	159.15	190.82	217.25	257.40	285.81	327.66
<chem>C1C(=O)CC(=O)CC1</chem>		-330.33	347.69	123.06	159.65	191.57	218.01	257.95	286.14	327.74
<chem>C1C(=O)CCC(=O)C1</chem>		-324.48	344.36	122.89	159.28	191.15	217.61	257.63	285.89	327.60
<chem>C1C(=O)C(O)CCCC1</chem>		379.42	362.65	138.37	177.56	211.91	240.54	284.22	315.57	363.00

Appendix B: Thermodynamic parameters from ab initio calculations

	398.41	354.69	132.46	173.68	210.49	241.38	288.23	320.96	368.02
	-389.65	354.61	134.03	174.16	209.17	238.41	282.93	314.74	362.65
<chem>C1C(=O)CC(O)CC1</chem>									
	-388.20	367.89	132.99	172.59	207.87	237.37	282.25	314.29	362.65
	-389.68	363.18	133.97	174.29	209.56	238.82	283.22	314.92	362.70
<chem>C1C(=O)CCC(O)C1</chem>									
	-385.74	368.96	132.67	172.30	207.59	237.10	282.03	314.11	362.36
<chem>C=CCCC[CH2]</chem>	167.67	414.52	133.91	165.22	193.49	217.50	266.29	283.54	327.99
									
<chem>[CH2]C(OO)CCC=C</chem>	86.23	461.78	177.47	210.09	238.84	262.86	299.93	327.82	370.17

Appendix B: Thermodynamic parameters from ab initio calculations

<chem>[CH2]CC(OO)CC=C</chem>		79.25	450.51	176.01	213.47	244.06	268.57	305.28	331.86	373.20
<chem>[CH2]CCC(OO)C=C</chem>		79.88	472.37	183.63	218.55	247.23	270.24	305.09	330.85	371.96
<chem>[CH2]CCCC=COO</chem>		86.67	497.96	165.91	200.69	231.46	257.03	296.21	324.82	369.31
<chem>C=CCCCC=O</chem>		-114.76	417.33	150.51	178.71	204.27	226.67	262.87	290.13	332.23

C.

Solvent descriptors, solute descriptors & Platts fragments

C . SOLVENT DESCRIPTORS, SOLUTE DESCRIPTORS & PLATTS FRAGMENTS.....193

C.1	ABRAHAM SOLVENT DESCRIPTORS	194
C.2	SOLUTE DESCRIPTORS.....	195
C.3	PLATTS FRAGMENTS	200
C.4	REFERENCES	201

C.1 Abraham solvent descriptors

Table C-1: Abraham solvent descriptors from the RMG database [1].

Solvent name	SMILES	c	a	b	e	s	l
water	O	-1.271	3.904	4.814	0.822	2.743	-0.213
1-octanol	CCCCCCCCO	-0.12	3.56	0.702	-0.203	0.56	0.939
benzene	C1=CC=CC=C1	0.107	0.457	0.169	-0.313	1.053	1.02
cyclohexane	C1CCCCC1	0.163	0	0	-0.11	0	1.013
dibutylether	CCCCOCCCC	0.369	2.626	-0.499	-0.216	0.026	1.124
octane	CCCCCCCC	0.215	0	0	-0.049	0	0.967
butanol	CCCCO	-0.039	3.781	0.995	-0.276	0.539	0.934
carbon tetrachloride	ClC(Cl)(Cl)Cl	0.282	0	0	-0.303	0.46	1.047
chloroform	ClC(Cl)Cl	0.168	0.28	1.37	-0.595	1.256	0.981
decane	CCCCCCCCC	0.156	0	0	-0.143	0	0.989
dichloroethane	CC(Cl)Cl	0.011	0.649	0.736	-0.15	1.436	0.936
dimethylformamide	N(C)(C)C=O	-0.174	4.112	0	-0.339	2.315	0.83
dimethylsulfoxide	CS(C)=O	-0.59	5.46	0	-0.2	2.89	0.732
dodecane	CCCCCCCCCCCC	0.053	0	0	0	0	0.968
ethanol	CCO	0.012	3.635	1.311	-0.206	0.789	0.853
heptane	CCCCCC	0.275	0	0	-0.162	0	0.983
hexadecane	CCCCCCCCCCCCCCCC	0	0	0	0	0	1
hexane	CCCCCC	0.292	0	0	-0.169	0	0.979
iso-octane	CC(C)CC(C)(C)C	0.275	0	0	-0.244	0	-6.708
nonane	CCCCCCCCC	0.2	0	0	-0.145	0	0.98
pentane	CCCCC	0.335	0	0	-0.276	0	0.968
toluene	CC1C=CC=CC=C1	0.121	0.467	0.099	-0.222	0.938	1.012
undecane	CCCCCCCCCCC	0.113	0	0	0	0	0.971
acetonitrile	CC#N	-0.007	2.085	0.418	-0.595	2.461	0.738
ethyl acetate	CCOC(C)=O	0.203	2.949	0	-0.335	1.251	0.917

C.2 Solute descriptors

Table C-2: Solute descriptors from the RMG database [1].

Solute name	SMILES	A	B	E	S	L
methane	C	0	0	0	0	-0.323
ethane	CC	0	0	0	0	0.492
propane	CCC	0	0	0	0	1.05
n-butane	CCCC	0	0	0	0	1.615
2-methylpropane	CC(C)C	0	0	0	0	1.409
n-pentane	CCCCC	0	0	0	0	2.162
2-methylbutane	CCC(C)C	0	0	0	0	2.013
2,2-dimethylpropane	CC(C)(C)C	0	0	0	0	1.82
n-hexane	CCCCCC	0	0	0	0	2.668
2-methylpentane	CCCC(C)C	0	0	0	0	2.503
3-methylpentane	CCC(C)CC	0	0	0	0	2.581
2,2-dimethylbutane	CCC(C)(C)C	0	0	0	0	2.352
2,3-dimethylbutane	CC(C)C(C)C	0	0	0	0	2.495
n-heptane	CCCCCCC	0	0	0	0	3.173
2-methylhexane	CCCCC(C)C	0	0	0	0	3.001
3-methylhexane	CCCC(C)CC	0	0	0	0	3.044
2,2-dimethylpentane	CCCC(C)(C)C	0	0	0	0	2.796
2,3-dimethylpentane	CCC(C)C(C)C	0	0	0	0	3.016
2,4-dimethylpentane	CC(C)CC(C)C	0	0	0	0	2.809
3,3-dimethylpentane	CCC(C)(C)CC	0	0	0	0	2.809
n-octane	CCCCCCCC	0	0	0	0	3.677
3-methylheptane	CCCCC(C)CC	0	0	0	0	3.51
2,2,4-trimethylpentane	CC(C)CC(C)(C)C	0	0	0	0	3.106
2,3,4-trimethylpentane	CC(C)C(C)C(C)C	0	0	0	0	3.481
n-nonane	CCCCCCCCC	0	0	0	0	4.182
2,2,5-trimethylhexane	CC(C)CCC(C)(C)C	0	0	0	0	3.567
n-decane	CCCCCCCCC	0	0	0	0	4.686
cyclopropane	C1CC1	0	0	0.18	0.15	1.314
cyclopentane	C1CCCC1	0	0	0.263	0.1	2.477
methylcyclopentane	CC1CCCC1	0	0	0.225	0.1	2.816

Solute name	SMILES	A	B	E	S	L
n-propylcyclopentane	CCCC1CCCC1	0	0.225	0	0.1	3.803
cyclohexane	C1CCCCC1	0	0	0.305	0.1	2.964
methylcyclohexane	CC1CCCCC1	0	0	0.244	0.1	3.323
1,2-dimethylcyclohexane	CC1CCCCC1C	0	0	0.236	0.1	3.6665
ethene	C=C	0	0.07	0.107	0.1	0.289
propene	C=CC	0	0.07	0.103	0.08	0.946
1-butene	C=CCC	0	0.07	0.1	0.08	1.491
1-pentene	C=CCCC	0	0.07	0.093	0.08	2.047
2-pentene	CC=CCC	0	0.07	0.141	0.08	2.211
3-methylbut-1-ene	C=CC(C)C	0	0.07	0.063	0.08	1.91
2-methylbut-2-ene	CC=C(C)C	0	0.07	0.159	0.08	2.226
1-hexene	C=CCCCC	0	0.07	0.078	0.08	2.572
2-methylpent-1-ene	C=C(C)CCC	0	0.07	0.09	0.08	2.588
1-heptene	C=CCCCCC	0	0.07	0.092	0.08	3.063
1-octene	C=CCCCCCC	0	0.07	0.094	0.08	3.568
1-nonene	C=CCCCCCCC	0	0.07	0.09	0.08	4.073
1,3-butadiene	C=CC=C	0	0.1	0.32	0.23	1.543
2-methylbuta-1,3-diene	C=CC(=C)C	0	0.1	0.313	0.23	2.101
2,3-dimethylbuta-1,3-diene	C=C(C)C(=C)C	0	0.14	0.352	0.23	2.69
cyclopentene	C1=CCCC1	0	0.1	0.335	0.2	2.402
cyclohexene	C1=CCCCC1	0	0.1	0.395	0.2	3.021
1-methylcyclohexene	CC1=CCCCC1	0	0.1	0.391	0.2	3.483
cyclohepta-1,3,5-triene	C1=CC=CCC=C1	0	0.18	0.764	0.46	3.442
propyne	C#CC	0.13	0.15	0.183	0.25	1.025
but-1-yne	C#CCC	0.13	0.15	0.178	0.23	1.52
pent-1-yne	C#CCCC	0.13	0.1	0.172	0.23	2.01
hex-1-yne	C#CCCCC	0.13	0.1	0.166	0.23	2.51
hept-1-yne	C#CCCCC	0.13	0.1	0.16	0.23	3
oct-1-yne	C#CCCCCCC	0.13	0.1	0.155	0.23	3.521
diethylether	CCOCC	0	0.45	0.041	0.25	2.015
di-n-propyl ether	CCCOCC	0	0.45	0.008	0.25	2.954
diisopropyl ether	CC(C)OC(C)C	0	0.45	0	0.19	2.482

Solute name	SMILES	A	B	E	S	L
di-n-butyl ether	CCCCOCCCC	0	0.45	0	0.25	3.924
tetrahydrofuran	C1CCOC1	0	0.48	0.289	0.52	2.636
2-methyltetrahydrofuran	CC1CCCO1	0	0.53	0.241	0.48	2.82
2.5-dimethyltetrahydrofuran	CC1CCC(C)O1	0	0.58	0.204	0.38	2.98
tetrahydropyran	C1CCOCC1	0	0.55	0.275	0.47	3.057
14-dioxane	CC1COCCO1	0	0.64	0.329	0.75	2.892
formaldehyde	C=O	0	0.33	0.22	0.7	0.73
acetaldehyde	CC=O	0	0.45	0.208	0.67	1.23
propionaldehyde	CCC=O	0	0.45	0.196	0.65	1.815
butyraldehyde	CCCC=O	0	0.45	0.187	0.65	2.27
isobutyraldehyde	CC(C)C=O	0	0.45	0.144	0.62	2.12
pentanal	CCCCC=O	0	0.45	0.163	0.65	2.851
hexanal	CCCCCC=O	0	0.45	0.146	0.65	3.357
heptanal	CCCCCCC=O	0	0.45	0.14	0.65	3.865
octanal	CCCCCCCC=O	0	0.45	0.16	0.65	4.361
nonanal	CCCCCCCCC=O	0	0.45	0.15	0.65	4.856
but-2-enal	CC=CC=O	0	0.49	0.387	0.8	2.57
hex-2-enal	CCCC=CC=O	0	0.45	0.404	0.8	3.4
propanone	CC(C)=O	0.04	0.51	0.179	0.7	1.696
butanone	CCC(C)=O	0	0.51	0.166	0.7	2.287
pentan-2-one	CCCC(C)=O	0	0.51	0.143	0.68	2.755
pentan-3-one	CCC(=O)CC	0	0.51	0.154	0.66	2.811
3-methylbutan-2-one	CC(=O)C(C)C	0	0.51	0.134	0.65	2.692
hexan-2-one	CCCCC(C)=O	0	0.51	0.136	0.68	3.262
4-methylpentan-2-one	CC(=O)CC(C)C	0	0.51	0.111	0.65	3.089
heptan-2-one	CCCCC(C)=O	0	0.51	0.123	0.68	3.76
heptan-4-one	CCCC(=O)CCC	0	0.51	0.113	0.66	3.705
octan-2-one	CCCCCCC(C)=O	0	0.51	0.108	0.68	4.257
nonan-2-one	CCCCCCCC(C)=O	0	0.51	0.119	0.68	4.735
nonan-5-one	CCCCC(=O)CCCC	0	0.51	0.103	0.66	4.698
decan-2-one	CCCCCCCCC(C)=O	0	0.51	0.108	0.68	5.245
undecan-2-one	CCCCCCCCCC(C)=O	0	0.51	0.101	0.68	5.732

Solute name	SMILES	A	B	E	S	L
cyclopentanone	<chem>O=C1CCCC1</chem>	0	0.52	0.373	0.86	3.221
cyclohexanone	<chem>O=C1CCCCC1</chem>	0	0.56	0.403	0.86	3.792
methyl formate	<chem>COC=O</chem>	0	0.38	0.192	0.68	1.285
ethyl formate	<chem>CCOC=O</chem>	0	0.38	0.146	0.66	1.845
n-propyl formate	<chem>CCCOC=O</chem>	0	0.56	0.132	0.63	2.433
isopropyl formate	<chem>CC(C)OC=O</chem>	0	0.4	0.091	0.6	2.23
isobutyl formate	<chem>CC(C)COC=O</chem>	0	0.4	0.095	0.6	2.789
isoamyl formate	<chem>CC(C)CCOC=O</chem>	0	0.4	0.092	0.6	3.306
methyl acetate	<chem>CCC(=O)OC</chem>	0	0.45	0.142	0.64	1.911
ethyl acetate	<chem>CCOC(C)=O</chem>	0	0.45	0.106	0.62	2.314
n-propyl acetate	<chem>CCCOC(C)=O</chem>	0	0.4	0.092	0.6	2.819
isopropyl acetate	<chem>CC(=O)OC(C)C</chem>	0	0.47	0.055	0.57	2.546
n-butyl acetate	<chem>CCCCOC(C)=O</chem>	0	0.45	0.071	0.6	3.353
isobutyl acetate	<chem>CC(=O)OCC(C)C</chem>	0	0.47	0.052	0.57	3.161
n-pentyl acetate	<chem>CCCCCOC(C)=O</chem>	0	0.45	0.067	0.6	3.844
isoamyl acetate	<chem>CC(=O)OCCC(C)C</chem>	0	0.47	0.051	0.57	3.74
n-hexyl acetate	<chem>CCCCCOC(C)=O</chem>	0	0.45	0.056	0.6	4.351
methyl propanoate	<chem>CCC(=O)OC</chem>	0	0.45	0.128	0.6	2.431
ethyl propanoate	<chem>CCOC(=O)CC</chem>	0	0.45	0.087	0.58	2.807
n-propyl propanoate	<chem>CCCOC(=O)CC</chem>	0	0.45	0.07	0.56	3.338
n-pentyl propanoate	<chem>CCCCCOC(=O)CC</chem>	0	0.45	0.05	0.56	4.331
methyl butanoate	<chem>CCCC(=O)OC</chem>	0	0.45	0.106	0.6	2.893
ethyl butanoate	<chem>CCCC(=O)OCC</chem>	0	0.45	0.068	0.58	3.271
n-propyl butanoate	<chem>CCCOC(=O)CCC</chem>	0	0.45	0.05	0.56	3.783
methyl pentanoate	<chem>CCCCC(=O)OC</chem>	0	0.45	0.108	0.6	3.392
ethyl pentanoate	<chem>CCCCC(=O)OCC</chem>	0	0.45	0.049	0.58	3.769
methyl hexanoate	<chem>CCCCCC(=O)OC</chem>	0	0.45	0.08	0.6	3.874
ethyl hexanoate	<chem>CCCCCC(=O)OCC</chem>	0	0.45	0.043	0.58	4.251
isobutyl isobutanoate	<chem>CC(C)COC(=O)C(C)C</chem>	0	0.47	0	0.5	3.885
acetic acid	<chem>CC(=O)O</chem>	0.61	0.45	0.265	0.65	1.75
propanoic acid	<chem>CCC(=O)O</chem>	0.6	0.45	0.233	0.65	2.29
butanoic acid	<chem>CCCC(=O)O</chem>	0.6	0.45	0.21	0.62	2.83

Solute name	SMILES	A	B	E	S	L
pentanoic acid	<chem>CCCCC(=O)O</chem>	0.6	0.45	0.205	0.6	3.38
3-methylbutanoic acid	<chem>CC(C)CC(=O)O</chem>	0.6	0.5	0.178	0.57	3.14
hexanoic acid	<chem>CCCCCC(=O)O</chem>	0.6	0.45	0.174	0.6	3.92
water	<chem>O</chem>	0.82	0.35	0	0.45	0.26
methanol	<chem>CO</chem>	0.43	0.47	0.278	0.44	0.97
ethanol	<chem>CCO</chem>	0.37	0.48	0.246	0.42	1.485
propan-1-ol	<chem>CCCO</chem>	0.37	0.48	0.236	0.42	2.031
propan-2-ol	<chem>CC(C)O</chem>	0.33	0.56	0.212	0.36	1.764
butan-1-ol	<chem>CCCCO</chem>	0.37	0.48	0.224	0.42	2.601
2-methylpropan-1-ol	<chem>CC(C)CO</chem>	0.37	0.48	0.217	0.39	2.413
butan-2-ol	<chem>CCC(C)O</chem>	0.33	0.56	0.217	0.36	2.338
2-methylpropan-2-ol	<chem>CC(C)(C)O</chem>	0.31	0.6	0.18	0.3	1.963
pentan-1-ol	<chem>CCCCCO</chem>	0.37	0.48	0.219	0.42	3.106
petan-2-ol	<chem>CCCC(C)O</chem>	0.33	0.56	0.195	0.36	3.106
pentan-3-ol	<chem>CCC(O)CC</chem>	0.33	0.56	0.218	0.36	2.84
2-methylbutan-1-ol	<chem>CCC(C)CO</chem>	0.37	0.48	0.219	0.39	3.011
3-methylbutan-1-ol	<chem>CC(C)CCO</chem>	0.37	0.48	0.192	0.39	2.86
2-methylbutan-2-ol	<chem>CCC(C)(C)O</chem>	0.31	0.6	0.194	0.3	2.63
hexan-1-ol	<chem>CCCCCCO</chem>	0.37	0.48	0.21	0.42	3.61
hexan-3-ol	<chem>CCCC(O)CC</chem>	0.33	0.56	0.2	0.36	3.343
2-methylpentan-2-ol	<chem>CCCC(C)(C)O</chem>	0.31	0.6	0.169	0.3	3.081
Ar	[Ar]	0	0	0	0	0
Ne	[Ne]	0	0	0	0	0
He	[He]	0	0	0	0	0
N2	<chem>N#N</chem>	0	0	0	0	-0.978
O2	<chem>[O][O]</chem>	0	0	0	0	-0.723

C.3 Platts fragments

Table C-3: Platts fragments in SMARTS with contributions to the solute descriptors [2].

Platts fragment (SMARTS)	A	B	E	S	L
<chem>[\$([C;X4H3])]</chem>	0.0	0.007	-0.104	-0.075	0.321
<chem>[\$([C;X4H2])]</chem>	0.0	0.0	0.0	0.0	0.499
<chem>[\$([C;X4H1])]</chem>	0.0	0.011	0.089	0.036	0.499
<chem>[\$([C;X4H0])]</chem>	0.037	0.037	0.187	0.071	0.443
<chem>[\$([C;X3v4H2])]</chem>	0.0	0.019	0.045	-0.085	0.244
<chem>[\$([C;X3v4H1])]</chem>	0.0	0.011	0.068	0.05	0.469
<chem>[\$([c;v4H1])]</chem>	0.0	0.011	0.068	0.05	0.469
<chem>[\$([c;v4]-[*;!c])]</chem>	0.0	0.0	0.180	0.101	0.624
<chem>[\$([c;v4](-[c])(-[c])(-[c]))]</chem>	0.018	0.019	0.300	0.121	0.744
<chem>[\$([C;X3v4H0])]</chem>	0.0	0.0	0.180	0.101	0.624
<chem>[\$([C;X2v4H0])]</chem>	0.0	0.028	0.040	0.034	0.332
<chem>[\$([C;X2v4H1])]</chem>	0.082	0.028	0.040	0.034	0.332
<chem>[\$([O;X2](-[c;v4])(-[#1]))]</chem>	0.543	0.307	0.061	0.247	0.672
<chem>[\$([O;X2])(-[#1])]</chem>	0.345	0.307	0.061	0.247	0.672
<chem>[\$([O;X2H0])]</chem>	0.0	0.211	0.014	0.185	0.360
<chem>[\$([O;X1H0])]</chem>	0.0	0.334	-0.041	0.370	0.495
<chem>[\$([N;X3H2]-[c;v4])]</chem>	0.247	0.275	0.163	0.383	0.949
<chem>[\$([N;X3H2]-[C;v4])]</chem>	0.177	0.481	0.085	0.175	0.781
<chem>[\$([N;X3H1](-[c;v4])(-[c;v4]))]</chem>	0.194	0.415	0.192	0.311	0.912
<chem>[\$([N;X3H1](-[C;v4])(-[C;v4]))]</chem>	0.087	0.541	0.138	0.265	0.568
<chem>[\$([N]#[c;v4])]</chem>	0.0	0.103	0.0	0.390	0.732
<chem>[\$([N]#[C;v4])]</chem>	0.0	0.242	0.0	0.694	0.757
<chem>[\$([n](:[c])(-[c]))]</chem>	0.300	0.321	0.460	0.223	0.574
<chem>[\$([O]=[C]-[O])]</chem>	0.0	0.134	-0.161	0.140	0.095
<chem>[\$([C;X3v4])(=[O;X1])(-[O;X2H1])]</chem>	0.243	-0.308	-0.012	-0.311	0.255
<chem>[\$([C;X4v4](-[O;v2])(-[#1])(-[C;X4v4](-[O;v2])(-[#1])))]</chem>	0.0	0.0	-0.043	0.052	0.100
<chem>[\$([C;X3v4H0])(=[O;X1])(-[O;X2])]</chem>	0.0	-0.206	0.067	-0.124	0.234
<chem>[\$([C;X3v4H0])(=[O;X1])(-[O;X2])(-[O;X2])]</chem>	0.0	-0.267	0.0	-0.19	0.0
Intercept	0.003	0.071	0.248	0.277	0.130

C.4 References

1. W. H. Green and R.H. West, *RMG Solvation database*, MIT, Editor. 2019.
2. J. Platts, D. Butina, M. H. Abraham, and A. Hersey, *Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach*. Vol. 39. 1999. 835-845.

