
**Integrated Computer-Aided Design
of Molecules and Processes
using COSMO-RS**

**Integriertes Prozess- und Moleküldesign
mithilfe von COSMO-RS**

Von der Fakultät für Maschinenwesen der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Jan David Scheffczyk

Berichter: Univ.-Prof. Dr.-Ing. André Bardow
Univ.-Prof. Dr.-Ing. Kai Sundmacher
Apl.Prof. Dr. rer. nat. Kai Olaf Leonhard

Tag der mündlichen Prüfung: 05. Juni 2018

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Aachener Beiträge zur Technischen Thermodynamik Band 14

Jan David Scheffczyk

Integrated Computer-Aided Design of Molecules and Processes using COSMO-RS

Integriertes Prozess- und Moleküldesign mithilfe von COSMO-RS

ISBN: 978-3-95886-236-4

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwendung ist ohne die Zustimmung des Herausgebers außerhalb der engen Grenzen des Urhebergesetzes unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Herstellung & Vertrieb:

1. Auflage 2018

© Wissenschaftsverlag Mainz GmbH - Aachen

Süsterfeldstr. 83, 52072 Aachen

Tel. 0241/87 34 34

Fax 0241/87 55 77

www.Verlag-Mainz.de

ISSN: 2198-4832

Satz: nach Druckvorlage des Autors

Umschlaggestaltung: Druckerei Mainz

printed in Germany

D82 (Diss. RWTH Aachen University, 2018)

Danksagung

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Technische Thermodynamik der RWTH Aachen. Besonderer Dank gilt meinem Doktorvater, Herrn Prof. Dr.-Ing. André Bardow, der mich stets im Vorhaben der Promotion unterstützt hat. Das mir entgegengebrachte Vertrauen sowie die übertragene Verantwortung haben mich fachlich und persönlich entscheidend geprägt. Weiterhin danke ich Herrn Prof. Dr. rer. nat. Kai Leonhard für die wertvolle Unterstützung sowohl beim Verfassen meiner Arbeit als auch bei den abenteuerlichsten Flug-, Kletter- und Skimanövern. Besonderer Dank gilt außerdem Herrn Prof. Dr.-Ing. Kai Sundmacher für die Übernahme des Ko-Referats sowie Herrn Prof. Dr.-Ing. Andreas Jupke für die Übernahme des Prüfungsvorsitzes.

Während meiner Tätigkeit am Lehrstuhl hatte ich das Glück einer Vielzahl von inspirierenden und bemerkenswerten Menschen zu begegnen, die meine Zeit in Aachen zu etwas Großartigem gemacht haben. Allen voran möchte ich meinem Projektpartner Christian Redepenning danken. Unsere Kooperation war der entscheidende Baustein für das Entstehen der vorliegenden Arbeit. Ebenfalls danke ich Christian Jens, dessen wertvolle Unterstützung mir durch die anfänglichen Durststrecken bei der Arbeit mit COSMO-RS geholfen hat. Ich danke den Mitarbeitern des gesamten Lehrstuhls, insbesondere allen Mitarbeitern der Energiesystemtechnik und den Kollegen aus dem Exzellenzcluster *Tailor-Made Fuels from Biomass*. Besonders danke ich meiner langjährigen Büropartnerin Maike Hennen, die unermüdlich das Überleben unserer Büropflanzen sichergestellt hat. Großer Dank geht an die Arbeitsgruppe Prozessdesign, deren Leitung mir sehr viel Spaß gemacht hat. Außerdem danke ich meinen Mentoren Matthias Lampe, Philip Voll und Niklas von der Aßen, die alle inspirierende Vorbilder für mich waren. Besonderer Dank gilt darüber hinaus allen Studenten, die zum Gelingen dieser Arbeit beigetragen haben, insbesondere Benedikt Winter, Christiane Reinert, Lorenz Fleitmann und Pascal Schäfer.

Der BASF möchte ich für die ausgezeichnete Förderung im Rahmen des *European Talent Pool* und des *Discover Together* Mentoringprogramms danken. Insbesondere danke ich meinem Mentor Carsten Knösche, der mir spannende Einblicke in die Arbeits- und Denkweise in verschiedenen Bereichen der chemischen Industrie ermöglicht hat.

Ein herzliches Dankeschön geht außerdem an die beste E\$T-Sailing Crew für Präzisionssegelmanöver und die stets volle Fahrt voraus. Ganz besonders möchte ich mich schließlich bei meinem langjährigen Kollegen und Freund Johannes Schilling für die weltbesten Konferenzreisen und die tolle gemeinsame Zeit am Lehrstuhl bedanken.

Der wichtigste Dank gilt abschließend denjenigen, die mich neben meiner Arbeit am Lehrstuhl begleitet haben. Danke an meinen engsten Freund David, der mit mir Höhen und Tiefen durchgestanden hat. Danke an meine wundervolle Freundin Kristina, die mir liebevoll und selbstlos den Rücken freigehalten hat - ohne Dich wäre das alles nicht möglich gewesen! Danke an meine Familie. Eure Unterstützung und Eure Liebe hat mich angetrieben, beflügelt und getragen. Ihr seid die Besten - Danke für alles!

Aachen, im Juni 2018

Jan David Scheffczyk

*'Design is a funny word. Some people think design means how it looks.
But of course, if you dig deeper, it's really how it works.'*

Steve Jobs,
"Wired" (1996)

Contents

Contents	V
List of Figures	IX
List of Tables	XI
Notation	XII
Kurzfassung	XIX
Abstract	XXI
1 Introduction	1
1.1 Structure of this Thesis	2
2 State-of-the-Art in Computer-Aided Molecular and Process Design	5
2.1 General Problem Formulation	6
2.1.1 Molecular Design Problem	7
2.1.2 Integrated Molecular and Process Design Problem	8
2.2 Molecular Structure Exploration Methods	10
2.2.1 Selection Methods	10
2.2.2 Design Methods	11
2.3 Thermodynamic Models	15
2.3.1 Data-driven Models	16
2.3.2 Quantum Mechanics-based Models	18
2.4 Process Performance Models	21
2.4.1 Property-based Models	21
2.4.2 Process Models	22
2.5 Contribution of this Thesis	23
3 Massive, Automated Solvent Screening for Minimum Energy Demand in Hybrid Extraction-Distillation using COSMO-RS	27
3.1 Introduction	27

3.2	Automated Screening Approach	29
3.2.1	Thermodynamic Model	29
3.2.2	Process Performance Model	30
3.2.3	Molecular Structure Exploration Method	32
3.3	Case Study: Purification of the Platform Chemical γ -Valerolactone . .	35
3.4	Comparison to Conventional Process Performance Indicators	41
3.5	Conclusions	44
4	COSMO-CAMD: A Framework for Optimization-Based Computer-Aided Molecular Design using COSMO-RS	45
4.1	Introduction	45
4.2	COSMO-CAMD Framework	47
4.2.1	Accuracy Levels in COSMO-RS	47
4.2.2	Genetic Algorithm LEA3D	49
4.2.3	COSMO-CAMD Procedure	51
4.3	Case Study 1: Solvent Design for Phenol Extraction	53
4.3.1	Problem Specification	53
4.3.2	Results	54
4.4	Case Study 2: Solvent Design for HMF Extraction	60
4.4.1	Problem Specification	60
4.4.2	Results	61
4.5	Conclusions	64
5	COSMO-CAMPD: A Framework for Integrated Design of Molecules and Separation Processes based on COSMO-RS	65
5.1	Introduction	65
5.2	COSMO-CAMPD Framework	67
5.2.1	Integrated Molecular and Process Design Approach	67
5.3	Case study: Integrated Solvent and Process Design for Hybrid Extraction-Distillation of γ -Valerolactone	70
5.3.1	Process Formulation	70
5.3.2	Application of COSMO-CAMPD	72
5.3.3	Results of the Design Phase	74
5.3.4	Results of the Refinement Phase	76
5.4	Experimental Validation of COSMO-RS Prediction	80
5.5	Conclusions	84

6	Integrated Process and Molecule Design using COSMO-RS for Reaction-Separation Processes	85
6.1	Introduction	85
6.2	Integrated Design of Molecules, Processes and Flowsheets	87
6.2.1	Problem Specification	87
6.2.2	Integrated Process and Molecular Design for Reaction-Separation Processes	87
6.3	Case Study: Production of CO from CO ₂ and H ₂	89
6.3.1	Process Specification	90
6.4	Reaction-Separation Process Models	92
6.4.1	Reactor Models	93
6.4.2	Separation Models	93
6.4.3	Application of COSMO-CAMPD	95
6.5	Results for Integrated Process and Solvent Design	97
6.6	Conclusions	101
7	Summary, Conclusions and Future Perspective	103
7.1	Summary and Conclusions	103
7.2	Future Perspective	105
A	Process-Level Solvent Evaluation	111
B	Computer-Aided Molecular Design	121
C	Computer-Aided Molecular and Process Design	125
D	Integrated Design of Reaction-Separation Processes	135
	Bibliography	145

List of Figures

2.1	General CAMD/CAMPD problem	6
2.2	Key elements of CAMD/CAMPD approaches	9
2.3	Content of this thesis in the broader CAMD/CAMPD context	24
3.1	Content of Chapter 3 in the broader CAMD/CAMPD context	28
3.2	Process flowsheet for hybrid extraction-distillation	31
3.3	Schematic procedure of the proposed solvent screening approach	33
3.4	Minimum energy demand $Q_{\min}$ and minimum amount of solvent $S_{\min}$ for valid solvent candidates	37
3.5	Ternary diagram of literature benchmark n-butyl acetate and best solvent identified by solvent screening.	39
3.6	Comparison of phase distribution coefficient P to minimum amount of solvent $S_{\min}$ and minimum energy demand $Q_{\min}$	42
4.1	Content of Chapter 4 in the broader CAMD/CAMPD context	46
4.2	Root mean square error (RMSE) for calculating the distribution coefficient $\log P_i$ for a representative set of 31 molecules	50
4.3	Schematic procedure of the COSMO-CAMD framework	51
4.4	σ -profiles of best COSMO-CAMD molecule, best molecule from database search and solute phenol	56
4.5	Evolution of the mean and maximum objective function values	57
4.6	Molecules generated in COSMO-CAMD using COSMOfrag and BP-TZVP-MF in comparison to refined molecules on BP-TZVPD-FINE level	59
4.7	Objective function values $F(y)$ for molecules generated in COSMO-CAMD	63
5.1	Content of Chapter 5 in the broader CAMD/CAMPD context	66
5.2	Integrated process and molecular design approach COSMO-CAMPD.	69
5.3	Hybrid extraction-distillation process for the purification of GVL	71
5.4	Minimum energy demand $Q_{\min}$ vs. minimum solvent demand $S_{\min}$	75

5.5	Comparison of the minimum energy demand $Q_{\min}$ calculated with the pinch-based process models to minimum energy demand Q_{rig} from rigorous simulations	78
5.6	Experimental validation for designed solvent 2-methylfuran	81
5.7	Minimum energy demand for COSMO-CAMPD prediction and experimental validation	83
6.1	Content of Chapter 6 in the broader CAMD/CAMPD context	86
6.2	Integrated process and molecule design approach COSMO-CAMPD for reaction-separation processes	88
6.3	Process variants for the conversion of CO_2 to CO	90
6.4	Exergy demand in process vs. storage molecule (SMO) concentration after synthesis reactor	98
6.5	Breakdown of contributions to process exergy demand for three solvents using DMF as storage molecule	100
7.1	Contribution of this thesis in the broader CAMD/CAMPD context . .	104
7.2	Panorama of further research topics classified by level of integration and scale	106
B.1	Molecular fragment library for extraction of phenol from water	122
B.2	Molecular fragment library for extraction of HMF from water	123
C.1	Molecular fragment library for GVL purification	127
C.2	Experimental validation for solvent n-butyl acetate	130
C.3	Experimental validation for solvent toluene	131
D.1	Schematic representation of the VLLE reactor	136
D.2	Comparison of the VLLE reactor to commercial process simulations . .	138
D.3	Process flowsheet structure for the conversion of CO_2 to CO	141
D.4	Molecular fragment library for reaction-separation process	143

List of Tables

3.1	Selected results for solvent screening	38
4.1	Ranked list of solvents designed by COSMO-CAMD	55
4.2	Ranked list of solvents designed by COSMO-CAMD	62
5.1	Ranked list of solvents designed by COSMO-CAMPD	79
5.2	Minimum energy demand of experimentally validated solvents	82
A.1	Extended list of the screening results	113
A.2	Extended list of the screening results	116
B.1	Parameters for optimization with LEA3D	121
C.1	Block specifications for the hybrid extraction-distillation process	125
C.2	Stream compositions for the hybrid extraction-distillation process	126
C.3	Statistics for the COSMO-CAMPD optimization for GVL	126
C.4	Assumptions for the rigorous simulation of the hybrid extraction-distillation process	128
C.5	List of chemicals	129
C.6	Gas chromatography parameters	129
C.7	LLE measurements for the system 2-methylfuran-GVL-water	132
C.8	LLE measurements for the system toluene-GVL-water	133
C.9	LLE measurements for the system n-butyl acetate-GVL-water	134
D.1	Statistics for the COSMO-CAMPD optimization for DMF	142
D.2	Statistics for the COSMO-CAMPD optimization for MeF	142

Abbreviations

AR	Additional Reactant
ASC	Apparent Surface Charge
BM	benchmark
BP-TZVP	COSMO-RS parametrization
BP-TZVPD-FINE	COSMO-RS parametrization
BP-TZVP-MF	COSMO-RS parametrization
CAMD	Computer-Aided Molecular Design
CAMPD	Computer-Aided Molecular and Process Design
COSMO-RS	Conductor-like Screening Model for Realistic Solvation
CPA	Cubic-Plus-Association
CSM	Continuum Solvation Model
DDB	Dortmund Data Bank
DFT	Density Functional Theory
EOS	Equation of State
GA	Genetic Algorithm
GC	Group Contribution
GGA	Generalized Gradient Approximation
HB	Hydrogen bond
IL	ionic liquid
LEA	Ligand by Evolutionary Algorithm
LLE	liquid-liquid equilibrium
MAPE	mean average percentage error
MINLP	Mixed Integer Nonlinear Program
MOO	Multi-Objective Optimization
NLP	Nonlinear Program
NRTL	Non-Random-Two-Liquid
ORC	Organic Rankine Cycle
PA	Purification After Reforming
PB	Purification Before Reforming
PCM	Polarizable Continuum Model
PR	Peng-Robinson
QM	Quantum Mechanic
RBM	Rectification Body Method
RMSE	root mean square error
SA	Solvent After Synthesis

SAFT	Statistical Associating Fluid Theory
SB	Solvent before Synthesis
SD	Structural Descriptors
SMILES	Simplified Molecular Input Line Entry Specification
SMO	Storage Molecule
SRK	Soave-Redlich-Kwong
TI	Topological Indices
UNIFAC	Universal Quasichemical Functional Group Activity Coefficients
UNIQUAC	Universal Quasichemical
VLE	vapor-liquid equilibrium
VLLE	vapor-liquid-liquid equilibrium

Chemical Species

CO	carbon monoxide
CO ₂	carbon dioxide
DEF	diethylformamide
DIPF	diisopropyl-formamide
DMA	dimethylamine
DMF	dimethylformamide
DNBE	di-n-butylether
EtF	ethylformate
FA	formic acid
GVL	γ -valerolactone
H ₂	hydrogen
H ₂ O	water
hfOCT	1h,8h-hexadecafluorooctane
HMF	hydroxymethylfurfural
MeF	methylformate
MeOH	methanol
MIBK	methylisobutylketone
nBAC	n-butyl acetate
TOL	toluene

Latin Symbols

a	solute
c	carrier
c	molecular constrains
E	extract stream/phase
F	objective function
F	feed stream
g	Gibbs free energy
g	thermodynamic properties
h	enthalpy
h	inequality constraints
K	mass-based phase distribution coefficient
M	molar weight
n	number of (heavy) atoms
N	number of molecules
P	mole-based phase distribution coefficient
p	pressure
Q	energy demand
R	raffinate stream
R^2	coefficient of determination
S	solvent stream/solvent demand
T	temperature
$\mathbf{w}$	weight fraction
x	process variables
y	(structure of) molecule
Y	design space of all possible molecules
$\mathbf{z}$	molar composition
z	process structure
Z	set of possible flowsheet variants

Greek symbols

α	relative volatility
γ	(isothermal) activity coefficient
ρ	correlation coefficient
σ	screening charge density
ξ	reaction conversion

Subscripts

B	bottom
boil	boiling
comp	compression
cond	condensation
D	distillate
E	extract
F	feed
i	component
max	maximum
melt	melting
min	minimum
R	raffinate
rig	rigorous
S	solvent
sat	saturation
vap	vaporization

Superscripts

*	optimal
~	feasible/final
DB	database
extr	extraction
TD	thermodynamic properties

Kurzfassung

Der Erfolg chemischer Prozesse hängt entscheidend von zwei Faktoren ab: Zum einen von der optimalen Betriebsweise des Prozesses und zum anderen von den darin eingesetzten Molekülen wie beispielsweise Lösungsmitteln. Die Auswahl optimaler Moleküle und Prozesse erfolgt jedoch häufig aufgrund voneinander getrennter Aspekte: Entweder werden die Prozesse selbst optimiert, wobei auf eine festgelegte Vorauswahl an Molekülen zurückgegriffen wird, oder Moleküle werden anhand vereinfachender Kriterien ausgewählt. Darüber hinaus werden bei der Suche nach optimalen Molekülen oft stark vereinfachte thermodynamische Modelle zur Stoffdatenvorhersage genutzt, die zahlreiche experimentell bestimmte Modellparameter benötigen und dadurch den molekularen Designraum einschränken. Derzeitige Auswahlmethoden für Prozesse und Moleküle erfassen somit häufig nicht die komplexen Zusammenhänge molekularer Eigenschaften im Prozess und betrachten nur eine begrenzte Menge an möglichen Molekülen. Diese Vorgehensweise birgt das Risiko, nicht immer optimale Moleküle und Prozesse auszuwählen.

Um gleichzeitig optimale Moleküle und Prozesse zu identifizieren, wird in dieser Arbeit ein Ansatz für integriertes *Computer-Aided Molecular and Process Design* (CAMPD) präsentiert. Der Ansatz nutzt COSMO-RS zur Stoffdatenvorhersage, welches auf Quantenmechanik basiert und daher nur gering von experimentell bestimmten Modellparametern abhängt. Um eine verlässliche Bewertung komplexer Prozesse zu ermöglichen, werden sogenannte pinch-basierte Prozessmodelle genutzt. Diese pinch-basierten Prozessmodelle berücksichtigen den inhärenten Zusammenhang molekularer Eigenschaften im Prozess und sind im Vergleich zu rigorosen Prozessmodellen sehr effizient. Der integrierte Designansatz wird in dieser Arbeit schrittweise von molekularen Screenings auf Prozessebene bis hin zum molekularen Design für Trenn- und Reaktionsprozesse erweitert. Die Anwendung des vorgestellten Designansatzes wird dabei für verschiedene Beispiele der Lösungsmittelauswahl und Prozessoptimierung veranschaulicht. Dabei werden Prozesse und Lösungsmittel für die Aufbereitung biobasierter Plattformchemikalien sowie die Herstellung von CO aus CO₂ betrachtet. In dieser Arbeit wird somit COSMO-RS erfolgreich in CAMPD integriert und der Lösungsraum und die Anwendbarkeit aktueller CAMPD Ansätze deutlich erweitert.

Abstract

Optimal performance of chemical processes requires both optimized operating conditions and carefully selected molecules such as solvents. However, the search for optimal molecules and process concepts often has a limited focus: Either processes are optimized using a pre-defined set of molecules or molecules are selected for novel applications based on simplified process indicators. At the same time, the search for optimal molecules often relies on strongly simplified thermodynamic models that require experimentally determined group interaction parameters and confine the molecular design space. Overall, current design approaches often do not capture complex process trade-offs and are limited to prescriptive sets of molecules which likely results in suboptimal choices.

To address the challenge of identifying optimal processes and molecules, this thesis presents an integrated computer-aided molecular and process design (CAMPD) approach. The design approach uses quantum mechanics (QM)-based property prediction by COSMO-RS and is thus independent of experimentally determined group interaction parameters while not relying on group additivity. For reliable and fast evaluation of complex processes, advanced pinch-based process models are employed. These pinch-based process models account for the inherent trade-off in molecular properties while being both computationally efficient and accurate in comparison to rigorous process models. The integrated design approach in this thesis is stepwise extended from process-level molecular screenings towards molecular design for separation and reaction-separation processes. The application of the presented integrated design approach is illustrated for various examples of solvent selection and process optimization. In particular, process concepts and solvents are investigated for the purification of bio-based platform chemicals as well as the production of CO from CO₂. Overall, this thesis successfully integrates COSMO-RS property prediction in CAMPD and thus significantly expands the range and applicability of current CAMPD approaches.

CHAPTER 1

Introduction

The key to realize efficient implementations of challenging chemical processes such as complex reaction or purification steps are suitable molecules. Molecules play this key role as adsorbents in gas adsorption (Joos et al., 2015; Bai et al., 2015; Huck et al., 2014; Braun et al., 2016), absorbents in washing processes (Pereira et al., 2011; Papadopoulos et al., 2016) and solvents in reaction processes (Zhou et al., 2016; Austin et al., 2016a; Jens et al., 2016) or liquid-liquid extraction (Redepenning and Marquardt, 2017; Song et al., 2017). For economically and ecologically efficient processes, molecules thus need to be carefully selected (Ng et al., 2015).

Commonly, molecules are selected based on expert knowledge and heuristics such as selection guidelines (Ashcroft et al., 2015; Prat et al., 2014; Tobiszewski et al., 2015) or prescriptive design rules (Barnicki and Fair, 1990, 1992; Jaksland et al., 1995). These design rules often rely on simple performance indicators, e.g., separation factors (Bonet et al., 2015) or phase distribution coefficients (Salleh et al., 2017; Fang et al., 2016) which provide a quick estimate of performance for large numbers of molecules with low computational costs. However, simple performance indicators do not capture complex trade-offs in molecular properties on the process level, e.g., molecules with a large phase distribution coefficient might stabilize a value chemical in an extraction process but can be unfavorable in a subsequent distillation. Neglecting these trade-offs likely results in suboptimal choices for molecules and processes (Papadopoulos and Linke, 2006a). Selecting suitable molecules is thus intrinsically linked to the process in which the molecule is used. However, a direct integration of molecule selection into process design is challenging since the molecular design space is merely endless (Fink et al., 2005). In addition, model equations relating molecular structure to thermodynamic properties and these thermodynamic properties to process performance are usually highly non-linear (Adjiman et al., 2014). The large molecular design space and the large set of highly non-linear equations lead to complex mixed integer non-linear programming (MINLP) problems (Gani, 2004a).

In recent years, the resulting problems have been tackled in so-called computer-aided molecular design (CAMD) or computer-aided molecular and process design (CAMPD) approaches (Ng et al., 2015; Austin et al., 2016b). Typically, CAMD/CAMPD approaches rely on reduced (process) performance models and data-driven thermodynamic property prediction methods such as group-contribution methods to be computationally tractable. These data-driven prediction methods have to be trained on initial experiments, which usually define and limit the molecular design space. In contrast, quantum mechanics (QM)-based methods can accurately predict properties independent from extensive experimental data, but a direct integration of QM into CAMD/CAMPD approaches is computationally challenging. A good trade-off in computational demand and accuracy has been achieved by COSMO-RS (Klamt et al., 2010). COSMO-RS combines efficient QM calculations with fast statistical thermodynamics allowing for comparably fast computations of thermodynamic properties. Currently, COSMO-RS is however mostly used as property prediction method and its predictive power is not fully exploited in CAMD/CAMPD approaches.

1.1 Structure of this Thesis

In this work, QM-based thermodynamic predictions are integrated into computer-aided molecular and process design by combining efficient property prediction by COSMO-RS with fast but accurate process-models. **Chapter 2** reviews common CAMD/CAMPD methods to provide a general overview of current approaches and challenges in the field. Here, theoretical foundations are summarized that are relevant to integrate COSMO-RS into CAMD/CAMPD: General CAMD/CAMPD problem formulations are presented and state-of-the-art solution approaches to CAMD/CAMPD are reviewed. This review reveals limitations of current CAMD/CAMPD approaches and the scientific contribution of this thesis is outlined (Chapter 2.5).

To establish sound but still efficient process-level evaluation of molecules, a method is introduced in **Chapter 3** that allows for large-scale, process-based solvent screening using COSMO-RS property prediction. Here, efficient pinch-based process models (Marquardt et al., 2008) are employed for fast but accurate process-level assessment of solvents. The resulting method is applied in a massive, automated database screening of several thousand solvents for a hybrid extraction-distillation process. To further extend the range of COSMO-RS from prediction to design, an optimization framework (COSMO-CAMD) is introduced in **Chapter 4** which allows for computer-aided molecular design with COSMO-RS using thermodynamic property design targets. COSMO-CAMD is applied in two case studies for solvent design.

In **Chapter 5**, COSMO-RS is fully integrated in CAMPD. For this purpose, the COSMO-CAMD framework presented in Chapter 4 is combined with pinch-based process models (Chapter 3) that allow for rapid process-level evaluation of designed solvents. The resulting COSMO-CAMPD framework is a hybrid stochastic-deterministic optimization method for integrated design of molecules in separation processes with COSMO-RS property prediction. The illustrative example of a hybrid extraction-distillation process from Chapter 4 is extended to include process variables as additional degrees of freedom. COSMO-CAMPD results are validated experimentally by automated liquid-liquid equilibrium experiments.

Besides separation processes, complex reactions are often relevant in chemical engineering. In **Chapter 6**, the optimization framework COSMO-CAMPD is thus extended to include multiphase reactions and automated process flowsheet selection. The extended framework for solvent design in reaction-separation processes is applied to a case study of CO production from H_2 and CO_2 . Finally, **Chapter 7** summarizes the work and gives a perspective on future work.

State-of-the-Art in Computer-Aided Molecular and Process Design

This chapter introduces general CAMD/CAMPD methods, discusses state-of-the-art solution approaches and highlights current challenges as well as limitations in CAMD/CAMPD. First, CAMD and CAMPD are defined via generic problem formulations (Section 2.1). In Section 2.2, current solution approaches to CAMD/CAMPD problems are presented. Thermodynamic process models employed in CAMD/CAMPD are discussed in Section 2.3 and methods for evaluating (process) performance are summarized in Section 2.4. Finally, open research tasks in the field of CAMD/CAMPD are highlighted and the contribution of this thesis is outlined (Section 2.5).

Parts of this chapter have been published in:

J. Scheffczyk, C. Redepenning, C. M. Jens, B. Winter, K. Leonhard, W. Marquardt and A. Bardow. Massive, Automated Solvent Screening for Minimum Energy Demand in Hybrid Extraction-Distillation using COSMO-RS, *Chemical Engineering Research and Design*, 2016, 115 Part B, 433-442.

J. Scheffczyk, L. Fleitmann, A. Schwarz, M. Lampe, A. Bardow and K. Leonhard. COSMO-CAMD: A Framework for Optimization-Based Computer-Aided Molecular Design using COSMO-RS, *Chemical Engineering Science*, 2017, 159, 84-92.

2.1 General Problem Formulation

Generally, CAMD/CAMPD approaches are systematic methods that aim at solving an inverse design problem (Fig. 2.1).

In this context, the solution of the analysis (direct) problem predicts the (process) performance of a given molecular structure based on thermodynamic property prediction models. In contrast, the solution of the inverse design problem finds the optimal (or at least better) molecular structure that meets a given process or property target. CAMD/CAMPD approaches differ in their respective design target: CAMD is the inversion of property prediction whereas CAMPD inverts the prediction of process performance (Ng et al., 2015). In order to solve the inverse design problem, CAMD/CAMPD approaches depend on three key elements (Fig. 2.1): A model of the (process) performance as design target, a thermodynamic model and an approach

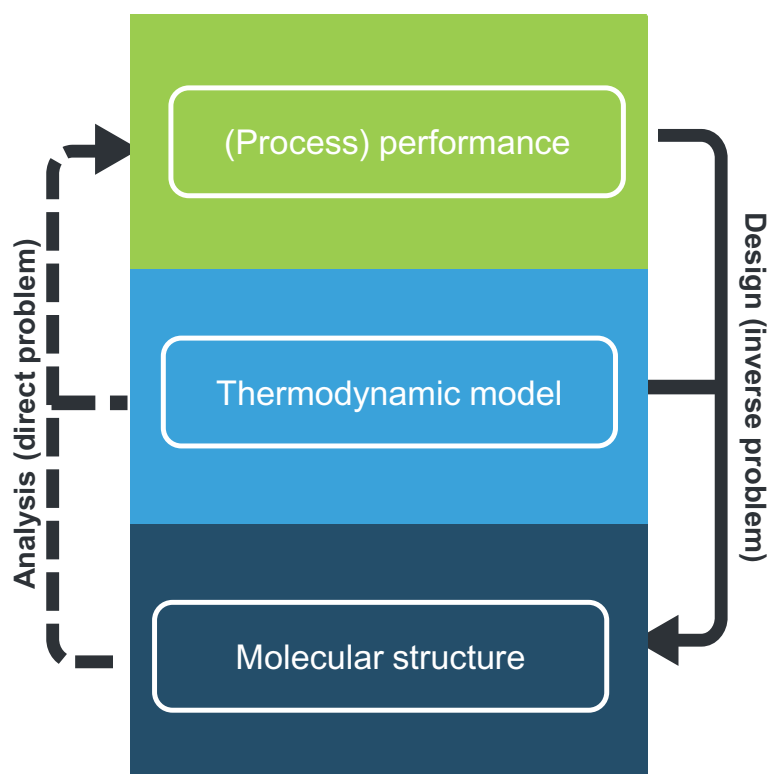


Figure 2.1: General CAMD/CAMPD problem as inverse design problem (solid line) to the direct analysis problem (dashed line) with three key elements of CAMD/CAMPD: (Process) performance, thermodynamic model and the molecular structure.

to explore molecular structures. Combining these three key elements in an efficient solution approach is the main challenge in CAMD/CAMPD. This challenge can be posed as a numerical optimization problem for CAMD/CAMPD which is presented in the following section.

2.1.1 Molecular Design Problem

The molecular design problem in CAMD can be stated as following: Given a process (e.g., absorption) that requires a molecule (e.g., solvent), find the optimal molecular structure that maximizes a selected performance criterion (e.g., solvent boiling point). This leads to a CAMD problem that can be formulated as:

$$\begin{aligned}
 & \underset{y}{\text{minimize}} && F(y) \\
 & \text{subject to} && g(y) = 0 \text{ (thermodynamic model),} \\
 & && h(y) \leq 0 \text{ (property constraints),} \\
 & && c(y) \leq 0 \text{ (molecular constraints),} \\
 & && y \in Y.
 \end{aligned} \tag{2.1}$$

Here, $F(y)$ is the objective function (e.g., distance to desired solvent property) and depends on the structure of the molecule y (e.g., solvent) as design variable. The molecular structures y are found within the design space of all possible molecules Y that can be generated from a defined set of building blocks (e.g., molecular fragments or groups). Typically, molecules are discrete structures expressed by integer variables. Equality constraints $g(y)$ encompass thermodynamic models which relate the molecular structure y to the thermodynamic properties g . $h(y)$ represents inequality constraints, e.g., due to property limits such as desired maximum boiling temperature. Molecular constraints are imposed by $c(y)$ such as structural feasibility and size of the designed solvents.

The choice of the objective function $F(y)$ depends on the respective design target and the application of the CAMD problem. Notably, in classical CAMD problem formulations, no process model is included in the objective function $F(y)$, e.g., the problem contains no mass- or energy balances for process unit operations. In contrast, the objective function quantifies the performance of the molecular structure y by specific design targets such as desired thermodynamic properties g that aim at assessing the process performance of the molecule (Sahinidis et al., 2003).

2.1.2 Integrated Molecular and Process Design Problem

Typically, process-level information is crucial for the success of molecular and process design (Adjiman et al., 2014). In order to integrate process-level information in the design problem, the CAMD problem (Eq. (2.1)) has to be extended to include explicit process models. The resulting CAMPD problem is a mixed-integer nonlinear programming (MINLP) problem (Gani, 2004b):

$$\begin{aligned}
 & \underset{x, y}{\text{minimize}} && F(x, y) \\
 & \text{subject to} && g_1(x, y) = 0 && (\text{thermodynamic model}), \\
 & && g_2(x, y) = 0 && (\text{process model}), \\
 & && h(x, y) \leq 0 && (\text{operating limits}), \\
 & && c(y) \leq 0 && (\text{molecular constraints}), \\
 & && x \in \mathbb{R}^n, y \in Y.
 \end{aligned} \tag{2.2}$$

Here, $F(x, y)$ is the objective function (e.g., process energy demand) and depends on two types of design variables: process variables x (e.g., process temperature) and the structure y of the molecule employed in the process (e.g., solvent). As for the CAMD problem (Eq. (2.1)), the molecular structures y are expressed as discrete integer variables and found within the design space of all possible molecules Y that can be generated from a defined set of building blocks (e.g., molecular fragments or groups). Equality constraints $g_1(x, y)$ in CAMPD encompass thermodynamic models as well as $g_2(x, y)$ the process model, e.g., mass- and energy balances for all unit operations. $h(x, y)$ represents inequality constraints due to process operating limits (e.g., maximum operating temperature) or limits on thermodynamic properties (e.g., existence of a liquid-liquid equilibrium). Molecular constraints are imposed by $c(y)$ such as structural feasibility and size of the designed solvents.

Typically, CAMD/CAMPD problems (Eq. (2.1) and Eq. (2.2)) can be classified according to various aspects such as the intended application or the employed solution strategy. In this work, a classification is proposed according to the *scope* of each key element in the CAMD/CAMPD problem (Fig. 2.2): Here, the term *scope* refers to the information that is considered in the respective key element, e.g., a (process) performance model with limited information about the actual process is considered as limited-scope model. In contrast, a molecular structure exploration method that is able to explore a vast molecular space by design is considered as large-scope method. Accordingly, each key element in the CAMD/CAMPD problem, i.e., (process) performance, thermodynamic model and molecular structure, can be further subdivided

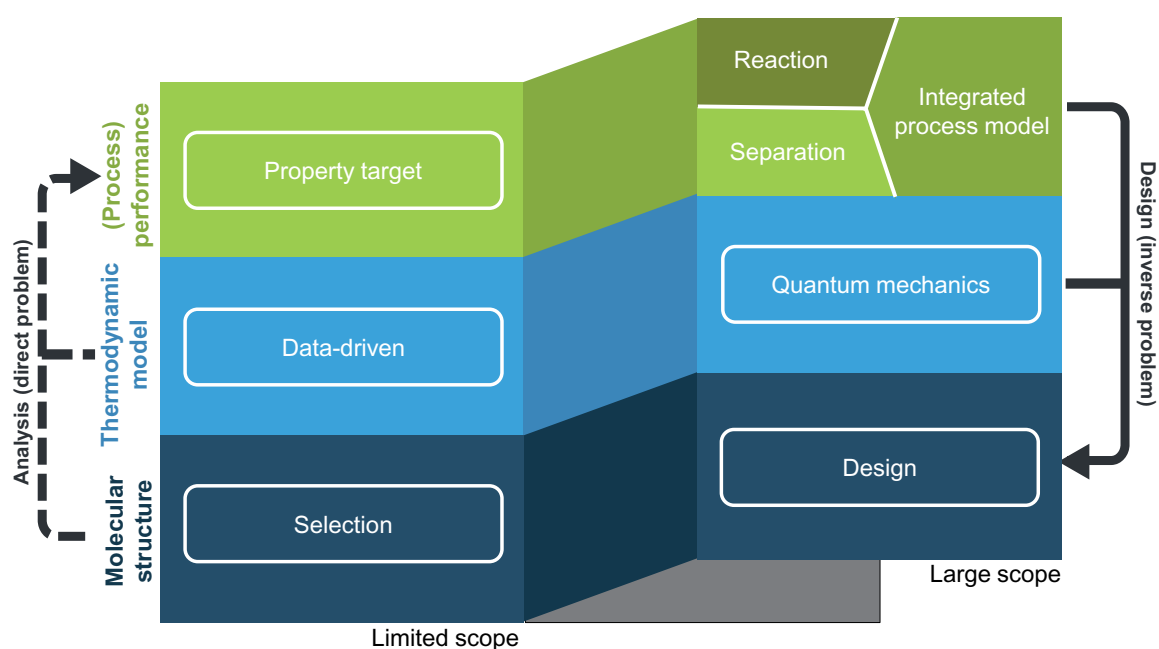


Figure 2.2: Key elements of CAMD/CAMPD approaches with limited/large scope classification for (process) performance (limited scope: property targets; large scope: integrated process models such as flowsheets with reaction and separation steps), thermodynamic models (limited scope: data-driven methods; large scope: quantum mechanics) and molecular structure exploration (limited scope: selection; large scope: design).

in limited- or large-scope elements. Here, large-scope process performance models (Fig. 2.2) comprise full process models (CAMPD), e.g., process flowsheets with rigorous process models for reaction and separation steps. These models take into account all process steps, e.g., by consideration of mass- and energy balances for unit operations. In contrast, limited-scope performance models (Fig. 2.2) are thermodynamic property targets (CAMD), e.g., a design target based on pure component vapor pressure. Similarly, data-driven thermodynamic property prediction methods are considered as limited-scope approaches (Fig. 2.2), e.g., first-order group-contribution methods with experimentally determined functional group interaction parameters. These experimentally determined parameters usually limit and confine the molecular design space in CAMD/CAMPD problem solution approaches and thus limit the scope of the CAMD/CAMPD problem. In contrast, quantum mechanics (QM)-based thermodynamic models employ first principle calculations (ab initio) and thus do not rely on experimentally determined functional group interaction parameters. Thus, QM-based property prediction methods are considered large-scope key elements in CAMD/CAMPD (Fig. 2.2). Lastly, molecular structure exploration approaches are distinguished, i.e., heuristic strategies that select molecules from a fixed set such as databases (limited scope in Fig. 2.2). In contrast to heuristic molecule selection, large-scope molecular structure exploration methods (Fig. 2.2) design molecular structures, e.g., by numerical optimization. For each of the three key elements of CAMD/CAMPD, state-of-the-art methods are reviewed for limited- and large-scope approaches in the following sections.

2.2 Molecular Structure Exploration Methods

The success of solution approaches to the CAMD/CAMPD problems (Eq. (2.1) and Eq. (2.2)) strongly depends on the method to identify optimal molecular structures. In this work, methods for exploring the design space of molecular structures are distinguished by their scope (Fig. 2.2), i.e., selection methods (limited scope) or design methods (large scope).

2.2.1 Selection Methods

Probably the most intuitive approach to solve the inverse design problem (solid line Fig. 2.2) is to repeatedly solve the analysis problem (dashed line in Fig. 2.2). In this work, such approaches are referred to as (systematic) selection methods (limited scope,

Fig. 2.2). Systematic selection can be performed, e.g., by enumerating a fixed set of molecules. Molecules in this set can either be selected based on expert knowledge or systematically enumerated from molecule databases. For each selected molecule, the objective function of the CAMD/CAMPD problem (Eq. (2.1) or Eq. (2.2)) is evaluated to rank molecules based on their predicted (process) performance. From this ranked list, top performing molecules are usually selected for detailed investigation, e.g., rigorous process model evaluation after expert molecule pre-selection based on heuristic indicators (Le Nhien et al., 2016; Kossack et al., 2008).

Systematic molecule selection is often performed for property-based CAMD tasks. Here, many applications use expert selection guidelines which contain molecules such as solvents with benign properties, e.g., safety (Prat et al., 2014; Henderson et al., 2011), toxicity (Voutchkova et al., 2011) or environmental indicators (Capello et al., 2007). To assess these molecular properties, clustering techniques are often employed that classify molecules based on statistical analysis of extensive datasets (Moity et al., 2012). Thereby, heuristic molecular descriptors can be derived for physicochemical properties such as solubility or density (Tobiszewski et al., 2015). These property-based descriptors are then employed, e.g., to substitute molecules with undesired properties by similar molecules with benign properties (Diorazio et al., 2016). Commonly, molecule selection methods largely rely on expert-knowledge to pre-select a few molecules for detailed investigation (Lapkin et al., 2010; Burghoff et al., 2008). In contrast, more systematic molecule selection approaches have been developed that automatically screen comprehensive databases for molecules with desired properties (Sendek et al., 2017; Blumenthal et al., 2016). Recently, systematic selection strategies have further addressed process design in large-scale molecule screenings based on integrated process optimization (Song et al., 2017; Preißinger et al., 2017; Schwöbel et al., 2017; Jens et al., 2016).

2.2.2 Design Methods

To extend the design space beyond pre-determined sets of molecules, design methods create novel molecular structures and are thus considered large-scope solution approaches for CAMD/CAMPD problems in this work (Fig. 2.2). Several design methods exist for the solution of CAMD/CAMPD problems depending on the complexity and size of the CAMD/CAMPD problem. Design methods in CAMD/CAMPD encompass generate-and-test methods, mathematical optimization, decomposition methods and heuristic optimization.

Generate-and-test methods generate large numbers of possible novel molecular

structures by enumerative combination of structural building blocks such as functional groups (Gani and Brignole, 1983). For each generated molecular structure, the (process) performance model is evaluated or tested against design specifications and promising molecules are selected for further consideration (Harper et al., 1999). Generate-and-test approaches commonly rely on simple property prediction methods and/or process models due to the combinatorial explosion of possible molecular structures (Harper and Gani, 2000). Here, the challenge lies in generating all possible molecular structures with little redundancy which is often tackled by multi-stage approaches: In successive design stages, the modeling detail of the CAMD/CAMPD problem is increased while the number of possible molecular structures is gradually reduced by applying constraints such as structural feasibility or limits on thermodynamic properties (Conte et al., 2011; Harper and Gani, 2000; Harper et al., 1999). In a final stage, the full MINLP problem (Eq. (2.2)) can be solved for a reduced set of molecules (Gani et al., 1991).

Mathematical optimization methods are used for CAMD problems that are amenable to direct optimization, e.g., single-molecule design problems. Most commonly, outer approximation (OA) (Duran and Grossmann, 1986) is employed to solve the MINLP problem (Eq. (2.2)) using relaxation of the MINLP problem and sequential solution of mixed-integer linear programming (MILP) as well as non-linear programming (NLP) subproblems. Direct mathematical optimization for CAMD problems has been applied, e.g., for refrigerant design (Duvedi and Achenie, 1996; Churi and Achenie, 1996), where global optimal solutions are found only if the NLP subproblems are convex. Global optimal solutions for CAMD/CAMPD problems can further be identified by interval analysis (Sinha et al., 2003) or branch-and-bound (BnB) algorithms (Sahinidis et al., 2003; Sinha et al., 1999). Currently, direct mathematical optimization methods are usually limited to CAMD formulations using thermodynamic property targets (Zhang et al., 2015; Zhou et al., 2015a). However, recent approaches have been proposed for integrated process- and molecular design tasks by directly solving the MINLP problem (Eq. (2.2)). Direct solution approaches rely on tractable problems such as pure-component design in Organic Rankine Cycles (ORC) (Schilling et al., 2017a,b). To address more complex process models, direct mathematical optimization methods employ shortcut process models (Zhou et al., 2015b) or impose problem-specific design constraints to facilitate the solution procedure (Gopinath et al., 2016).

In order to tackle more complex CAMD/CAMPD problems, **decomposition methods** are commonly used. Decomposition methods aim at reducing the problem size by breaking down originally complex problems in several sub-problems that can be

optimized more easily. Typical examples are multi-stage generate-and-test approaches which apply increasingly stringent constraints on generated molecular structures such as structural feasibility in successive design stages (Conte et al., 2011). Similarly, decomposition methods in CAMD are often employed in mathematical optimization methods where, e.g., optimization problems for structural feasibility targets and molecular properties are solved prior to the overall MINLP objective function (Karunanithi et al., 2006, 2005). Decomposition methods are *classic* CAMD solution strategies which have been employed for a large variety of applications such as gas absorption (Odele and Macchietto, 1993; Gani et al., 1991), polymer design (Zhang et al., 2015; Pavurala and Achenie, 2013; Eslick et al., 2009), reaction solvents (Zhou et al., 2016; Folić et al., 2008, 2007) or liquid-liquid extraction (Gebreslassie and Diwekar, 2015; Xu and Diwekar, 2005; Harper and Gani, 2000; Harper et al., 1999; Marcoulaki and Kokossis, 1998; Ourique and Silva Telles, 1998; Odele and Macchietto, 1993; Gani et al., 1991; Gani and Brignole, 1983).

Decomposition methods are also very popular in CAMPD where separation of the design of molecules and processes in the MINLP problem (Eq. (2.2)) leads to a natural problem decomposition in integer solutions (molecular structure) and non-linear programming problem (process performance). Several types of decomposition approaches in CAMPD can be distinguished:

- Early examples of decomposition methods in CAMPD are approaches by Papadopoulos and Linke (2005, 2006a,b). In these approaches, molecular and process design is decomposed based on multi-objective optimization (MOO) of property targets, which determine possible best molecular structures based on pre-defined heuristic criteria. MOO approaches leads to a Pareto set of molecular structures in so-called property clusters which are tested in successive process design. This *bottom-up* decomposition approaches start from the molecular level to find optimal processes. Thereby, these approaches rely on the assumption that optimal CAMPD solutions lie on a property-based Pareto front which may not be the case for complex CAMPD problems. Thus, property-based MOO decomposition approaches have been extended, e.g., by including shortcut process models in the Pareto set generation which leads to improved initialization points for process design (Burger et al., 2015).
- Other decomposition approaches decompose the process and molecule design problem by inverting the process design problem to identify optimal molecules (*top-down* decomposition approaches). This inversion of the process design problem can be performed by relaxation of molecular structure properties (continuous molecular targeting, CoMT). This relaxation leads to a lower bound by

process optimization which can be targeted in a subsequent molecule structure optimization. The overall approach is referred to as continuous-molecular targeting - computer-aided molecular design (CoMT-CAMD) (Bardow et al., 2010). CoMT-CAMD approaches have been applied using molecule selection from databases (Lampe et al., 2014; Stavrou et al., 2014) as well as integrated process and molecular design (Wang and Lakerveld, 2017; Lampe et al., 2015).

- Similar top-down decomposition approaches to CAMPD problems invert the process model by so-called property operators (Eden et al., 2004). These operators are derived from constitutive process model equations and transfer process information to the molecular design space by deriving property targets that optimize the process performance. These properties are then targeted in a separate CAMD step. CAMPD based on property operators has been adopted in various design approaches (Chemangattuvalappil and Eden, 2013; Bommareddy et al., 2010; Chemangattuvalappil et al., 2010a).

If problems are too complex for mathematical optimization or decomposition methods, **heuristic optimization methods** are usually employed in CAMD/CAMPD. Heuristic optimization methods are directed (random) search methods that iteratively generate sampling points in the design space (Rios and Sahinidis, 2013). For each sampling point, the objective function is evaluated and new sampling points can be determined, e.g., based on the objective function of previous sampling points. Usually, heuristic algorithms terminate by specified determination criteria such as a fixed number of iterations or a pre-defined optimality target (Fouskakis and Draper, 2002). In general, heuristic methods in CAMD/CAMPD approaches are performed in the domain of molecular structures (Austin et al., 2016b). Heuristic optimization methods are commonly employed to find near-optimal solutions and generate robust sets of promising molecular structures, e.g., to account for uncertainties in property prediction (Xu and Diwekar, 2005; Kim and Diwekar, 2002). Three popular methods of heuristic optimization in CAMD/CAMPD are simulated annealing, tabu search and genetic algorithms:

- Simulated annealing applies random alteration to molecular structure and trial solutions are evaluated based on the altered molecular structures (Marcoulaki and Kokossis, 1998). If the new solution is better within an error function, it is accepted as a best current solution and otherwise discarded. During the progress of the optimization, the error function becomes more stringent eventually forcing the convergence (Kim and Diwekar, 2002).
- In tabu search algorithms, an initially proposed molecular structure is altered until solutions based on the molecular structure appear in a tabu list (Eslick

et al., 2009). The tabu list contains forbidden solutions, e.g., frequently occurring molecules with inferior process performance. Thus, previous unsatisfying solutions are excluded from further optimization which creates a memory of the optimization. Tabu lists are often adjusted dynamically, i.e., novel solutions are allowed later in the progress that were previously forbidden (Lin et al., 2005). Tabu search algorithms can be advantageous over simulated annealing, e.g., when memory effects help to escape local optima (Chavali et al., 2004).

- The most prominent heuristic optimization methods in CAMD/CAMPD are genetic algorithms (GA). The generation of novel molecular structures with GA is based on the concept of natural selection (Holland, 1973). Here, a population of molecular structures evolves under selective pressure that favors better solutions in terms of a specified objective function. The optimization of molecular structures is based on two types of evolutionary operators: crossover (recombination of two molecular fragments) and mutation (structural modification of one single molecular fragment) that operate on the molecular structure with a specified probability. GA typically are applied in the molecular structure domain and features of high-performing molecules are passed on to the next generation until a convergence criterion is reached (Venkatasubramanian et al., 1995; Xu and Diwekar, 2005).

One of the main challenges in CAMD/CAMPD solution approaches is to identify suitable methods for the problem at hand. Often, different solution methods are combined to overcome limitations of individual solution approaches. E.g., challenges in using heuristic optimization algorithms are the risk of slow convergence speed due to random search and the lack of knowledge about global optimal solutions. Thus, heuristic approaches are often combined in hybrid stochastic-deterministic optimization procedures that combine advantages of heuristic and deterministic optimization methods (Zhou et al., 2017).

2.3 Thermodynamic Models

CAMD/CAMPD approaches usually require a detailed, reliable knowledge of thermodynamic behavior of components in the system of interest. This thermodynamic behavior can be predicted by thermodynamic models, which relate the structure of a molecule quantitatively to a thermodynamic property.

Properties of interest in CAMD/CAMPD largely coincide with properties relevant to chemical engineering problems, i.e., thermophysical properties (densities, heat ca-

capacities, enthalpies, viscosities) of pure components and mixtures (Gmehling, 2009). Problems involving multiphase equilibria or mixtures often require the calculation of non-ideal phase behavior by activity coefficients (Austin et al., 2016a; Jonuzaj et al., 2016) or fugacities (Linke et al., 2015; Papadopoulos et al., 2016). Depending on the application, equilibrium or non-equilibrium can be of importance, e.g., for CAMD/CAMPD problems involving multiphase equilibria (McBride et al., 2016; Burger et al., 2015), reaction equilibria (Zhou et al., 2016; Spieß et al., 2008), chemical kinetics (Zhou et al., 2015a; Struebing et al., 2013; Folić et al., 2008) or transport phenomena (Schilling et al., 2017b).

Commonly, thermodynamic equilibrium predictions are performed using either excess free Gibbs energy (g^E) models or equation of state (EOS) models. Classical examples for g^E models are the Wilson model (Wilson and Deal, 1962), NRTL ('Non-Random-Two-Liquid', Renon and Prausnitz (1968)) and UNIQUAC ('Universal Quasichemical', Fredenslund et al. (1975)). Most common EOS methods are SRK ('Soave-Redlich-Kwong', Soave (1972)) and PR ('Peng-Robinson', Peng and Robinson (1976)). Currently, much effort in research focuses on the development and improvement of these thermodynamic prediction models, e.g., the development of advanced EOS methods such as CPA ('Cubic-Plus-Association', Kontogeorgis et al. (2006)) and SAFT ('Statistical Associating Fluid Theory', Chapman et al. (1989)) type equations, which increase range and reliability of predictions.

Both g^E and EOS-models require mixture and/or component-specific model parameters to predict thermodynamic properties. Depending on how the model parameters are obtained, two types of methods can be distinguished for predicting thermodynamic properties in CAMD/CAMPD : Firstly, data-driven methods that determine model parameters from extensive experimental databases, e.g., by regression of interaction-parameters to phase equilibrium measurement data (limited scope, Fig. 2.2). Secondly, methods that obtain model parameters from first principle QM calculations requiring only very few empirical parameters (large scope, Fig. 2.2). QM-based methods are thus independent from extensive experimental data but instead require more elaborate QM calculations. The following section reviews these two types of current thermodynamic property prediction models and their application in CAMD/CAMPD.

2.3.1 Data-driven Models

The most common methods for predicting thermodynamic properties in CAMD/CAMPD are data-driven models such as group contribution (GC) methods. All GC methods assume the additivity of functional groups, i.e., thermodynamic properties

of a molecular structure are predicted based on the occurrence and type of defined molecular substructures. E.g., the molecular structure of an acetaldehyde (ethanal) molecule is expressed by a combination of the functional groups -CH_3 and -CHO . In GC methods, the molecule structure is usually defined in a vector whose elements contain the number of occurrence for each functional group. Component properties are then predicted as a function of these functional groups. The contribution of a functional group to the property is expressed by coefficients which are usually regressed using large experimental data sets, e.g., the Dortmund Data Bank (Gmehling, 2009). Generally, the segmentation of molecular structures reduces the required property data since thermodynamic properties are not required for all components but only for present functional groups.

A classic example for GC methods is the UNIFAC ('Universal Quasichemical Functional Group Activity Coefficients') method (Fredenslund et al., 1975; Gmehling et al., 2002) which has been used in the earliest CAMD/CAMPD approaches (Gani and Brignole, 1983; Odele and Macchietto, 1993; Joback and Stephanopoulos, 1995) and is still popular today. UNIFAC is able to predict phase equilibrium behavior for multicomponent systems and is well established throughout research and industrial application (Gmehling, 2009). However, UNIFAC is not applicable to supercritical fluids and properties such as densities or heat capacities cannot be calculated (Gmehling, 2003). More recently, advanced GC methods based on the Statistical Associating Fluid Theory (SAFT) have been integrated into CAMPD (Schilling et al., 2017a,b; Burger et al., 2015; Lampe et al., 2015; Pereira et al., 2011). In comparison to UNIFAC, SAFT-based models are able to predict non-ideal gas and liquid phase behavior thermodynamically consistent and are applicable for a large range of temperatures and pressures (Mac Dowell et al., 2010). SAFT-based models are thus commonly applied for CAMD/CAMPD problems involving high-pressure processes such as absorption (Gopinath et al., 2016; Papadopoulos et al., 2016) or Organic Rankine Cycles (Linke et al., 2015).

Typically, GC methods are fast and easy to apply to large sets of molecules. In addition, GC method predictions can be very accurate when parametrized to a specific problem (Satola et al., 2017). However, the application of GC methods in CAMPD is usually limited to simple first-order group contribution methods which strongly simplify the underlying thermodynamics. In general, the accuracy of GC methods decreases for molecules with more than one strongly interacting functional group (Deliovich et al., 2014). In addition, poor results are obtained for the solubility of very hydrophobic compounds, such as alkanes, alkenes, or cycloalkanes in water (Gmehling et al., 2015). For GC methods, lack of experimental data is particularly severe for

novel classes of compounds, for which usually little or no parameters are available; examples are missing UNIFAC group interaction parameters for solvent design in extractive distillation (Chen et al., 2005) or lack of data for the design of ionic liquids (IL) (Song et al., 2015).

Notably, in first-order GC methods, proximity effects by strongly polar neighboring groups are neglected and structural isomers cannot be distinguished (Gmehling et al., 2015). To resolve proximity effects, higher-order GC methods (Marrero and Gani, 2002, 2001) can be employed. Typically, higher-order GC methods include additional sets of (larger) functional groups to provide a more comprehensive picture of structural features such as aromatic rings or cis/trans-isomers. Alternative concepts to improve the structural resolution of molecules in CAMD/CAMPD approaches use topological indices (TI) (McLeese et al., 2010; Eslick et al., 2009) or structural descriptors (SD) (Chemmanurgattuvalappil et al., 2010b; Weis and Visco, 2010). Both TI and SD employ molecular descriptors from graph theory (Balasubramanian, 1985) which allow to distinguish very similar molecular structures. However, group-based prediction methods with improved structural resolution are more difficult to include in CAMD/CAMPD optimization procedures due to the combinatorial explosion of structural descriptors as well as required experimental group interaction parameters (Samudra and Sahinidis, 2013). These drawbacks strongly limit the application in CAMD/CAMPD as well as the available molecular design space.

2.3.2 Quantum Mechanics-based Models

In contrast to group-contribution methods, QM-based property prediction methods can predict thermodynamic properties with little need for component-specific experimental data (Deglmann et al., 2015). QM-based property prediction methods rely on (approximate) solutions to the Schrödinger equation, which describes the state of matter on a quantum level. For a detailed review of QM, the reader is referred to exhaustive literature on this field, e.g., Cancès et al. (2003). In this section, only a brief outline is given of QM-based property prediction methods relevant in CAMD/CAMPD approaches. Common QM-based methods can be subdivided in a hierarchy of methods with different degrees of assumptions for numerical solutions to the Schrödinger equation. Two prominent types of QM methods are ab-initio methods and density functional theory (DFT) methods.

Ab-initio methods are the most fundamental solution approaches to the Schrödinger equation. In these methods, the electronic structure and the energy of a system are calculated solely based on fundamental natural constants (Lin et al., 2005; Wu and

Sandler, 1991). However, the numerical expense of accurate calculations is tremendous and usually limits the application for practical engineering problems such as properties of a few single molecules (Leonhard and Deiters, 2000).

A good compromise between accuracy and computational costs are **DFT methods** (Kohn et al., 1996). DFT methods do not directly solve the Schrödinger equation but one-particle electron densities (Hohenberg and Kohn, 1964). These one-particle electron density solutions require the introduction of a functional that accounts for multi-body effects in the system. These functionals are based on the Kohn-Sham formalism (Kohn and Sham, 1965), e.g., functionals in local density approximation (LDA), generalized gradient approximation (GGA) or hybrid functionals (Weiss et al., 2016). Practically, DFT methods such as BP86 (GGA functional) and B3LYP (hybrid-GGA functional) are standard tools in quantum chemistry (Deglmann et al., 2015). The gap from DFT quantum chemistry gas phase calculations to condensed-phase systems is bridged, e.g., by continuum solvation models (CSM). In CSM, a solute molecule is virtually placed in a void cavity within a continuous dielectric medium that mimics the solvent (Tomasi et al., 2005). Hereby, the solute is described on a QM level whereas the solute-solvent interactions are derived from electrostatic charges. The charge distribution of the solute polarizes the dielectric continuum which in turn polarizes the solute charge distribution. This implicit solute-solvent interaction leads to a nested electrostatic problem which can be tackled by solving the implicit solute-solvent interaction numerically (Tomasi and Persico, 1994).

Presently, CSM put QM-based property prediction for liquid-phase problems within reach for CAMD/CAMPD applications: QM-based molecular design approaches have been proposed for polymers (Weiss et al., 2016), (extraction) solvents (Sheldon et al., 2006; Lehmann and Maranas, 2004) or reactions (Struebing et al., 2013; Stanescu and Achenie, 2006). These QM-based design approaches overcome the limits of GC method-based design approaches due to their general applicability to yet unknown molecules. However, large computational expense for QM calculations currently still limits the exploration of the full molecular design space.

A good trade-off between applicability and computational efficiency is provided by COSMO-RS ('COnductor like Screening MOdel for Real Solvents', Klamt (1995)). COSMO-RS is based on COSMO, a CSM that considers solute molecules virtually placed in a perfect conductor. Thereby, COSMO allows for efficient solution procedures for pure component molecule surface charges using QM-calculations (Eckert and Klamt, 2002). Based on the molecular surface charges, thermodynamic pure component and mixture properties (COSMO-RS) are then calculated using fast statistical thermodynamics (Klamt et al., 2010). Currently, COSMO-RS is mainly used

as prediction tool for thermodynamic properties for flowsheet simulators, e.g., for vapor-liquid equilibria (Spuhl and Arlt, 2004) or liquid-liquid equilibria (Jasperson et al., 2017; Freire et al., 2007)

Recently, COSMO-RS has been successfully applied to solvent selection in a vast number of applications. Here, COSMO-RS is employed for property prediction in solvent selection, e.g., the prediction of concentration-dependent partitioning equilibrium (Spieß et al., 2008) or solubility calculation (Lapkin et al., 2010). Similarly, COSMO-RS is used in the screening of tailor-made ionic liquids (Garcia-Chavez et al., 2012; Song et al., 2015; Zhou et al., 2012). Further applications of COSMO-RS to knowledge-based molecule selection are liquid-liquid extraction (Burghoff et al., 2008), thermomorphic systems (McBride et al., 2016) or reactive systems (Kröger et al., 2017; Deglmann et al., 2017). Furthermore, systematic large-scale COSMO-RS database screenings have been proposed based on property targets (Blumenthal et al., 2016; Zhou et al., 2014). Similar approaches are clustering techniques that create clusters of solvents based on COSMO-RS prediction to identify desired properties, e.g., benign environmental properties (Moity et al., 2012). Recent contributions combine rigorous process model simulations with expert solvent selection using COSMO-RS, e.g., for pharmaceutical components (Ferro et al., 2012) as well as in process design for compounds without existing database entries (Ferro et al., 2015) or solvent selection for IL (Larriba et al., 2018; Song et al., 2017; Lyu et al., 2014). Similarly, large-scale database screenings have been presented for Organic Rankine Cycles (Preißinger et al., 2017; Schwöbel et al., 2017) or synthesis of formic acid derivatives (Jens et al., 2016). However, database screenings are limited to components existing in databases and thus to known molecules. To overcome these limitations, molecular modifications (e.g., functional group addition or chain length variation) have been explored using heuristics with COSMO-RS-assisted experiments (Burghoff et al., 2008).

To go beyond the limitation of pre-stored database molecules and heuristic molecule modifications, only recently a few CAMD/CAMPD applications of COSMO-RS have been reported. The group of Sundmacher used COSMO-RS in an integrated solvent and process design for a Diels-Alder reaction (Zhou et al., 2015b) and the design of reaction solvents (Zhou et al., 2015a). Similarly, the group of Sahinidis proposed COSMO-RS-based molecular design for solvent mixtures (Austin et al., 2016a) and reactions (Austin et al., 2018). All current CAMD/CAMPD applications of COSMO-RS rely on empirical descriptors which approximate the surface charge of designed molecules by first-order group contribution methods. Designed molecules thus still lie within the space of pre-defined groups and rely on the inherent assumption of group additivity. Highly desirable is thus a method that overcomes limitations of

empirical descriptors or pre-defined molecular databases and enables for COSMO-RS-based molecular design while taking into account process models.

2.4 Process Performance Models

2.4.1 Property-based Models

Besides sound thermodynamic property prediction, the quality of the selected molecules in CAMD/CAMPD depends strongly on the (process) performance assessment criteria used (Adjiman et al., 2014). Commonly, the use of rigorous process models in CAMPD is prohibitive due to large computational costs and tedious convergence procedures (Cremaschi, 2015). Thus, many CAMPD approaches rely on simplified process models to pre-select molecules and to limit the vast molecular design space (limited scope, Fig. 2.2). Most prominent are design strategies based on heuristically defined property targets, e.g., simplified structure-property relations (Phillips et al., 2017; Moity et al., 2016, 2012) or expert knowledge (Le Nhien et al., 2016). Based on these property targets, a few molecules are preselected for rigorous process simulations. E.g., molecules are selected based on single (Kossack et al., 2008) or multi-objective (Burger et al., 2015; Papadopoulos and Linke, 2006b) pre-selection criteria. These molecule selection criteria are usually simplified process performance indicators based on physical molecule properties such as selectivity, solvent loss or phase distribution coefficients (Pretel et al., 1994).

The choice of a selection criterion for pre-selection is critical: Papadopoulos and Linke (2006b) show in a CAMPD approach for solvent design that several targets for solvent properties exist simultaneously in process flowsheets. In addition, desired solvent properties inherently include trade-offs. E.g., high affinity of the solute to the solvent is desired for efficient extraction, whereas a low affinity between solvent and solute helps to reduce the energy demand in solvent recovery by distillation. Papadopoulos and Linke (Papadopoulos and Linke, 2006b) show by multi-objective optimization (MOO) that these trade-offs in desired solvent properties cannot be captured by evaluating single solvent properties. Preferentially, solvent performance is directly evaluated on the process-level to fully capture the relevant trade-offs in solvent properties (Papadopoulos and Linke, 2005).

2.4.2 Process Models

Various approaches for targeting solvent performance on process-level evaluation have been proposed (large scope, Fig. 2.2). First approaches for integrated CAMPD stem from the pioneering work of (Eden et al., 2004) who invert process models to target properties based on derived property clusters (cf. Section 2.2). This property clustering technique has been adopted in various approaches for integrated CAMPD (Chemmangattualappil et al., 2010a; Eljack et al., 2007).

Bardow et al. (2010) identify a hypothetical optimal molecule during process optimization by relaxation of molecular properties. The resulting so-called CoMT-CAMD approach has been successfully applied to identify promising solvent candidates for physical absorption (Lampe et al., 2015; Stavrou et al., 2014) or working fluids in Organic Rankine Cycles (Schilling et al., 2017a,b; Lampe et al., 2014). Similarly, Pereira et al. (2011) integrate solvent and process design for methane recovery from carbon dioxide. Solution approaches addressing the full CAMPD problem (Eq. (2.2)) have been employed, e.g., hybrid-stochastic optimization to identify solvent candidates for a coupled absorption-desorption process (Zhou et al., 2017). A comprehensive review of current CAMPD approaches is given by, e.g., Ng et al. (2015). These studies show that the quality of the solvent selection critically depends on the quality of the process models.

Rigorous process models lead to highly accurate results but are laborious to solve and difficult to automate (Brüggemann and Marquardt, 2004). In particular, the initialization and thus the convergence limits the practical application to large sets of solvents. Thus, current conceptual process design approaches that apply rigorous process models first reduce the numbers of solvent candidates in pre-selection steps (Kossack et al., 2008). E.g., Burger et al. (2015) extend the approach of Papadopoulos and Linke (2006b) using simplified process models in the solvent pre-selection stage. Simplified process models are typically classic shortcut methods such as the well-known equations of Smith and Brinkley (1960) or Underwood (1949). A drawback of these classic shortcut methods is that they strongly simplify the underlying process model or thermodynamics which can cause inaccuracies, especially for non-ideal separations (Skiborowski et al., 2013).

In contrast, advanced pinch-based process models provide a thermodynamically sound objective (Skiborowski et al., 2013; Marquardt et al., 2008). These pinch-based process models exploit the concept of vanishing thermodynamic driving force in the so-called pinch points. The pinch points are calculated using the full non-ideal thermodynamics and thus overcome limits of conventional shortcut methods

leading to reliable process modeling even for non-ideal systems (Redepenning et al., 2016; Skiborowski et al., 2013). The pinch-based process models assume an infinite time for heat and mass transfer in separation unit operations and thus operate at a point of minimum reflux or minimum solvent demand (Brüggemann and Marquardt, 2011). Accordingly, pinch-based process models give a tight lower bound of the *minimum energy demand* of the process. This class of advanced pinch-based process models was used successfully in various applications, e.g., conceptual process design (Marquardt et al., 2008), reaction-separation process design (Recker et al., 2015), extractive distillation (Brüggemann and Marquardt, 2004), reactive rectification (Lee et al., 2003), hybrid extraction-distillation (Krämer et al., 2011a), multi-component extraction (Redepenning et al., 2016) and multicomponent absorption (Redepenning and Marquardt, 2017). The efficiency and accuracy of pinch-based process models was also successfully shown in screening approaches, e.g., for biorefinery processing pathways (Ulonska et al., 2016).

Overall, large-scope process models in CAMPD are currently mostly combined with limited-scope molecular selection approaches (cf. Fig. 2.2), e.g., simple database or experiment selection (Le Nhien et al., 2016; Kossack et al., 2008). Often, a limited set of solvents are pre-selected (Papadopoulos and Linke, 2005) for complex process models. In turn, large-scope molecular design is commonly performed for simplified design targets (Struebing et al., 2013; Austin et al., 2016a). Thus, highly desirable would be to transfer process information from large-scope process models to large-scope molecular design while being computationally efficient. Here, pinch-based process models seem to provide a promising trade-off in efficiency and accuracy.

2.5 Contribution of this Thesis

The review of current CAMD/CAMPD applications in the previous Sections 2.2, 2.3 and 2.4 reveals that there is a need for direct integration of QM-based property prediction into CAMPD. This integration should take into account the full structural molecular information to overcome limitations of group-additivity while still allowing for computationally efficient calculations. Similarly, molecules in the design procedure should be evaluated on the process level providing comprehensive information about process-inherent trade-offs while still being numerically robust and efficient. In this thesis, a large-scope CAMPD approach is thus developed that allows for integrated molecular and process design using QM-based property prediction with COSMO-RS. The large-scope CAMPD approach is stepwise established by integrating COSMO-RS into CAMD/CAMPD approaches with increasing scopes (Fig. 2.3). The CAMD/

CAMPD approaches are exemplified for solvent and process design.

In **Chapter 3** a first step is taken towards integration of COSMO-RS in CAMPD by addressing the CAMPD problem (Eq. (2.2)) in a process-level evaluation of a comprehensive solvent database. For this purpose, pinch-based process models for separation processes are employed, i.e., for distillation (Bausa et al., 1998) and extraction (Redepenning et al., 2016). Thereby, a reliable evaluation of the minimum energy demand of the process is possible within seconds. The evaluation of the minimum energy demand is combined with COSMO-RS property prediction in a fully automated solvent screening. Hereby, large-scope QM-based property prediction is established for large-scope process model evaluation using limited-scope molecular structure selection (Fig. 2.3, dashed line). The proposed approach is applied to a case study for a hybrid extraction-distillation.

To overcome limited-scope molecular selection, **Chapter 4** introduces COSMO-CAMD, a framework for optimization-based solvent design with COSMO-RS. COSMO-CAMD integrates COSMO-RS calculations into molecular design based on property

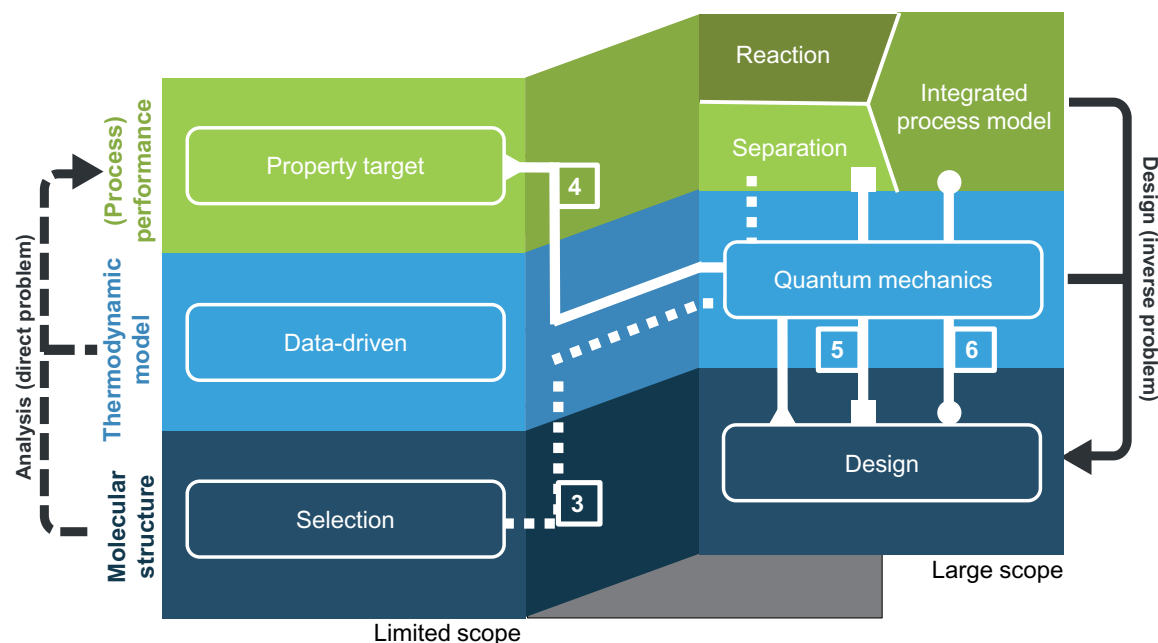


Figure 2.3: Content of this thesis in the broader CAMD/CAMPD context. Lines with chapter numbers (square boxes) indicate the classification of the CAMD/CAMPD approaches in the respective chapters: Chapter 3 (dashed line), Chapter 4 (line with triangles), Chapter 5 (line with squares) and Chapter 6 (line with dots).

targets and thus establishes large-scope molecular structure exploration for large-scope QM-based property prediction (Fig. 2.3, line with triangles). In this chapter, the COSMO-CAMD framework is presented and applied to two case studies for solvent design in liquid-liquid extraction.

In **Chapter 5**, COSMO-RS property prediction is fully integrated in CAMPD by presenting COSMO-CAMPD. COSMO-CAMPD further extends the scope of COSMO-RS-based CAMPD by using large-scope approaches for all three key elements (Fig. 2.3, line with squares): COSMO-CAMPD builds upon COSMO-CAMD (Chapter 4), which allows for molecular design with COSMO-RS based on thermodynamic property targets. In this chapter, COSMO-CAMD is combined with pinch-based process models for separation unit operations (Redepenning et al., 2016; Bausa et al., 1998) that have been applied for large-scale process evaluation of solvents in Chapter 3. The resulting COSMO-CAMPD framework is applied to a case study for process and solvent design in a hybrid extraction-distillation process. Finally, predictions of COSMO-CAMPD are challenged by performing liquid-liquid equilibrium experiments for top designed commercially available solvents.

In Chapter 5, COSMO-CAMPD is applied to the design of separation sequences. However, reaction steps, e.g., in the production of CO from CO₂ and H₂, require the consideration of complex solvent effects in reactions. In **Chapter 6**, the COSMO-CAMPD framework is thus extended to consider complex process concepts with multiphase equilibrium reactions in an integrated process model (Fig. 2.3, line with dots). The extended COSMO-CAMPD framework is applied to identify optimal solvents and process flowsheets in a reaction-separation process.

Massive, Automated Solvent Screening for Minimum Energy Demand in Hybrid Extraction-Distillation using COSMO-RS

3.1 Introduction

In this chapter, a first step is taken towards integrating QM-based property prediction into computer-aided molecular and process design. For this purpose, a molecular screening approach is developed which builds on two key aspects: First, thermodynamic properties of molecules are predicted by COSMO-RS (Klamt et al., 2010) which allows for an efficient evaluation of a large space of molecular structures without experimentally determined GC parameters (large-scope thermodynamic model, Fig. 3.1). Second, pinch-based process models for distillation (Bausa et al., 1998) and extraction (Redepenning et al., 2016) are integrated in a process flowsheet (large-scope process performance model, Fig. 3.1). In this chapter, pinch-based process models and COSMO-RS are combined into an automated molecular screening approach which allows for large-scale solvent assessment on process-level. The proposed molecular screening approach is presented and challenged in a case study for the hybrid extraction-distillation of the bio-based platform chemical γ -valerolactone (GVL) (Alonso et al., 2013a).

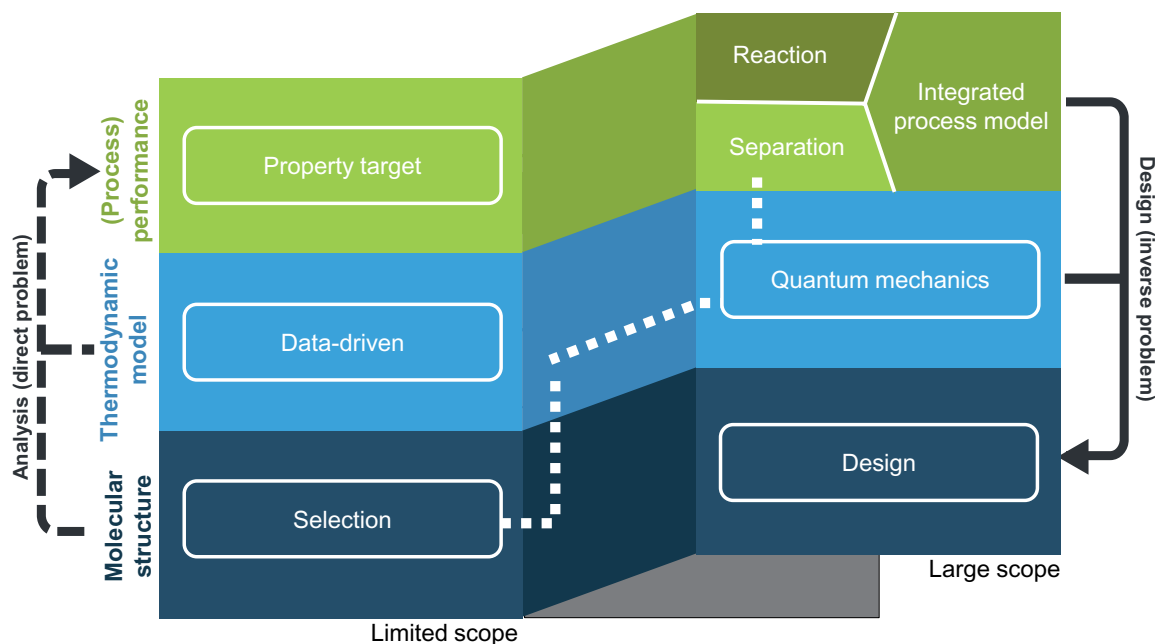


Figure 3.1: Content of this chapter in the broader CAMD/CAMPD context. Dashed line indicates the classification of the automated molecular screening approach presented in this chapter. The molecular screening approach (limited-scope molecular structure exploration method) combines QM-based COSMO-RS property prediction (large-scope thermodynamic model) with pinch-based separation process models (large-scope process performance model).

Executable routines of pinch-based process models employed in this chapter have been provided by Christian Redepenning. These executable routines are published under GNU General Public License GPL v3.0. Major parts of this chapter have been published in:

J. Scheffczyk, C. Redepenning, C. M. Jens, B. Winter, K. Leonhard, W. Marquardt and A. Bardow. Massive, Automated Solvent Screening for Minimum Energy Demand in Hybrid Extraction-Distillation using COSMO-RS, *Chemical Engineering Research and Design*, 2016, 115 Part B, 433-442.

3.2 Automated Screening Approach

The integrated molecular and process design problem (Eq. (2.2)) is approached in this chapter by enumerating a fixed set of solvents from a databank. In the following section, all three key elements of the CAMPD approach (cf. Chapter 2.1) are presented: thermodynamic model, process performance model and molecular structure exploration method.

3.2.1 Thermodynamic Model

In this thesis, all thermodynamic properties are predicted with COSMO-RS (Klamt et al., 2010). COSMO-RS is based on the quantum-mechanic (QM) COSMO method and allows for thermodynamic property predictions for any component accessible to QM. In particular, a large variety of thermodynamic properties for systems consisting of pure components or mixtures in liquid or vapor phases can be predicted without regression to system-specific experimental data. COSMO-RS calculations can be generally divided in two steps (For further information see (Klamt et al., 2010)):

1. For each pure component in the system, one QM COSMO calculation (usually DFT) is performed. COSMO is a continuum solvation model (CSM), which extends the range of QM-calculations to molecules dissolved in a continuum. COSMO is a special type of CSM in which the pure component is placed in an ideal conductor. This assumption simplifies the boundary conditions for the isolated molecule in comparison to dielectric continuum and allows for efficient calculations. Based on 3D structure information of the molecule, DFT calculations are performed and the molecule geometry is optimized until a self-consistent state (COSMO state) is reached. The result of the DFT/COSMO optimization is the so-called screening charge density (SCD), i.e., charges of discretized surface segments of the embedded molecule. The SCD is transformed in a histogram (σ -profile) for the next step.
2. Based on the σ -profiles of pure components from COSMO calculations, mixture properties are calculated by an interaction model and statistical thermodynamics (COSMO-RS). The underlying theory removes the ideal conductor between molecules in a mixture. The resulting local contact energies of COSMO molecules are combined to the so-called σ -potential. As a result, the chemical potential of the mixture components is derived. Notably, COSMO-RS calculations are performed within seconds from pure component σ -profiles. Thus, computationally demanding COSMO calculations are re-usable and need to be

performed only once for each component.

The computational effort of COSMO/COSMO-RS calculations depends strongly on the underlying QM in COSMO and the so-called parametrization in COSMO-RS. Different accuracy levels/parametrizations are available for COSMO-RS and provided within the software COSMOtherm (COSMOlogic, 2015d). The standard accuracy supported by COSMOtherm is BP-TZVP (in the following 'TZVP'), which uses the Becke-Perdew functional and the TZVP basis set (triple zeta valence plus polarization function) for geometry optimization and calculation of the SCD. Variations of the TZVP-level exist and are further discussed in Chapter 4. Large databanks (COSMObase, (COSMOlogic, 2015a)) are provided with pre-calculated COSMO files for common solvents and can be used for quick COSMO-RS calculations. For molecules not in these databanks, entirely new COSMO calculations need to be performed using COSMOconf (COSMOlogic, 2015b). In this chapter, COSMO-RS calculations for pre-calculated solvents from COSMObase are performed in COSMOthermX15 (COSMOlogic, 2015d).

3.2.2 Process Performance Model

In this chapter, the purification of a diluted solute a from a carrier c is considered in a hybrid-extraction distillation process (Skiborowski et al., 2013). A fixed flowsheet structure is used for the hybrid extraction-distillation. In this flowsheet, the process is limited to a single distillation step and the solute a is recovered as a heavy boiler (Fig. 3.2).

In the hybrid extraction-distillation process, a feed stream F with the molar composition $\mathbf{z}_F$ enters an extraction column where a solvent stream S is used to extract the solute a into the extract stream E . The raffinate stream R , deprived of solute a , leaves the extraction column and can further be processed, e.g., by wastewater treatment. The extract stream E is subsequently fed to a distillation column for further purification. In the distillation column, the solute a is separated from the solvent y and co-extracted carrier c . Solvent y and co-extracted carrier c are phase separated in a decanter stage and the solvent rich-phase is recycled to the extraction column. The carrier-rich phase is sent to wastewater treatment. Solvent loss is replaced by solvent make-up.

The process flowsheet is modeled by advanced pinch-based process models for extraction (Redepenning et al., 2016) and distillation (Bausa et al., 1998). For extraction, isothermal separation is assumed and for the distillation isobaric separation. The pinch-based process methods assume an infinite number of stages which allows for a

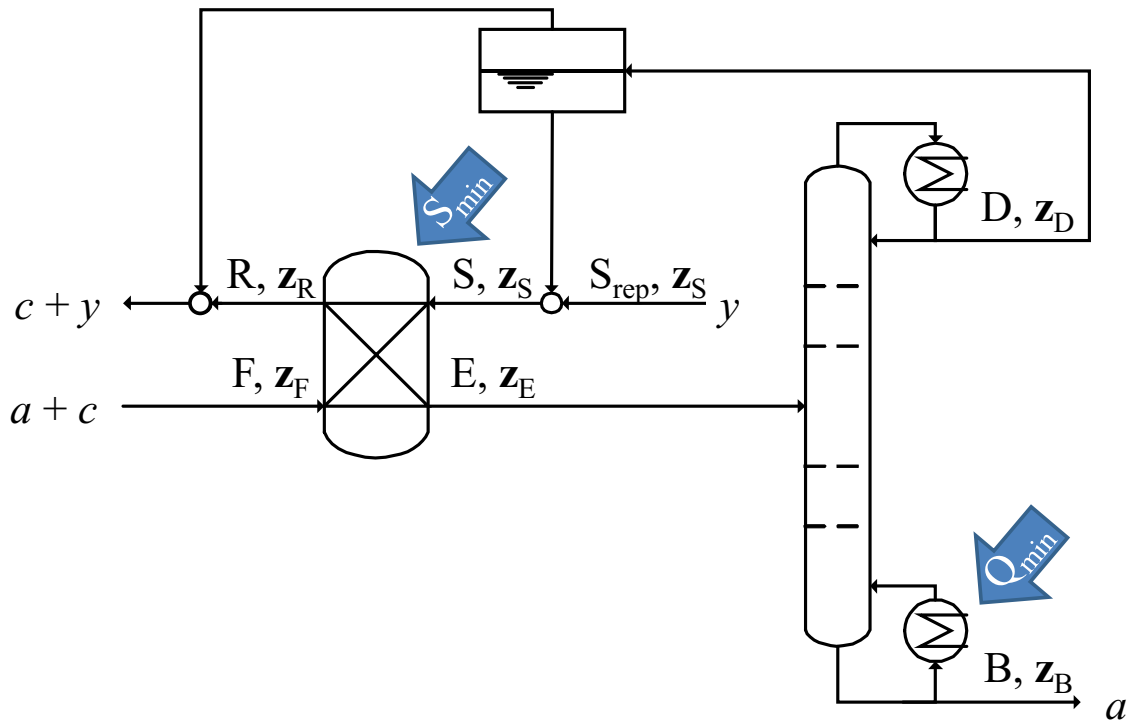


Figure 3.2: Process flowsheet for hybrid extraction-distillation. Feed stream F containing solute a and carrier c , solvent stream S containing solvent y , raffinate stream R , extract stream E , distillate stream D , bottom stream B , solvent make-up stream S_{rep} with corresponding compositions z_F , z_S , z_R , z_E , z_D , z_B . S_{min} is the minimum amount of solvent required for the specified separation task, Q_{min} is the minimum energy demand required for the specified separation task.

sharp split of the components in the separation steps. Thus, solute a is completely extracted into the extract stream E and the raffinate is free of solute a . Similarly, the solute a is completely recovered as pure product in the distillation column. Consequently, the carrier-rich phase from the phase separation in the decanter stage has the identical composition as the raffinate stream R.

The energy demand in the process is mostly determined by the heat demand of the reboiler in the distillation column. Other heating devices (e.g., preheating of the feed stream to boiling temperature) are assumed to be negligible. Thus, in this work, the process energy demand is represented by the energy requirement for distillation which is expressed by the minimum energy demand $Q_{\min}$. This minimum energy demand $Q_{\min}$ depends on the minimum amount of solvent $S_{\min}$ required for solute extraction in the extraction column. Thus, the performance of a solvent in the process is expressed by a single target function: The minimum energy demand $Q_{\min}$.

3.2.3 Molecular Structure Exploration Method

In the proposed solvent screening approach, the solvent y in the hybrid extraction-distillation process in Fig. 3.2 is considered as degree of freedom. Thus, the aim of the screening is to identify the solvent candidate y with the lowest minimum energy demand $Q_{\min}$ from a specified set of solvents. The solvent screening consists of the following steps (cf. Fig. 3.3):

1. Process specifications need to be set. Thus, the solute a and the carrier solution c are specified. In addition, the feed composition $\mathbf{z}_F$ needs to be specified. For extraction, temperature T is specified; for distillation, pressure p is set.
2. Design specifications on the solvent candidates y are imposed. A possible set of solvent candidates is defined by imposing a desired range for molecular size or specifying molecular constituents. E.g., certain groups of atoms such as halogens can be excluded from consideration. All solvent candidates meeting these design specifications are selected from a comprehensive COSMO-RS databank (COSMObase-1501-BP-TZVP) for screening. The COSMO databank contains pre-calculated, so-called COSMO files that result from quantum-mechanic structure optimization. COSMO-RS uses these COSMO files to predict thermodynamic properties. For a detailed description of the COSMO-RS method, see Klamt et al. (2010).

For each selected solvent candidate y in the COSMO-RS databank, relevant thermodynamic properties are predicted by COSMO-RS:

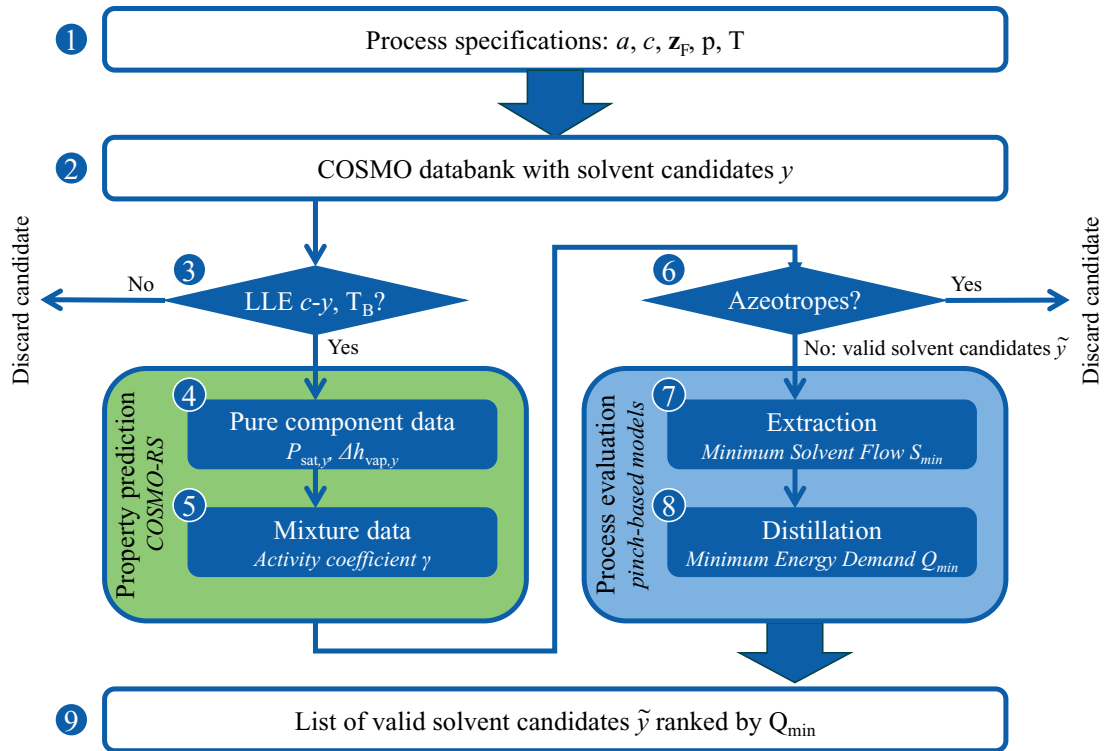


Figure 3.3: Schematic procedure of the proposed solvent screening approach.

3. The solvent candidates are tested to meet property constraints that prohibit their application as solvents in a hybrid extraction-distillation process. For this purpose, a preliminary evaluation of specific solvent properties is performed. The existence of a liquid-liquid equilibrium (LLE) in the binary mixture of the solvent y and the carrier c is mandatory for the use of the solvent candidate as an extraction agent. Thus, all solvent candidates are tested for the existence of this LLE using the binary LLE calculation in COSMOtherm (COSMOlogic, 2015d). This calculation checks whether an LLE exists. At this point, no constraint is added on the composition of the phases. Solvent candidates without a LLE are discarded. Additionally, solvent candidates are tested to be liquid at ambient temperature and pressure by calculating the boiling point temperature $T_{B,y}$ for all solvent candidates using COSMOtherm. Two limits are imposed on $T_{B,y}$: First, solvent candidates that are not liquid at ambient temperature with a safety margin of 15 K ($T_{B,y} \leq 313.15$ K) are discarded. Second, to obtain pure solute in the bottom product of the distillation column (cf. Section 3.2.2), all solvents that are heavy keys are discarded.
4. For the solvent candidates meeting all requirements from Step 3, pure component data is calculated. In particular, Antoine parameters to predict the vapor pressure $p_{\text{sat},y}$ and the molar enthalpy of vaporization $\Delta h_{\text{vap},y}$ are calculated using COSMO-RS (cf. Appendix A). This pure component data is used to estimate boiling temperature and the enthalpy of vaporization in the pinch-based process models.
5. In the next step, mixture data are calculated. Non-idealities in liquid-liquid systems and vapor-liquid systems are described using activity coefficients γ . Isothermal activity coefficients γ can directly be calculated using COSMO-RS. The mixture behavior of the ternary system a - c - y is described in the pinch-based process models using parameters of the *non-random-two-liquid* (NRTL) (Renon and Prausnitz, 1968) model. NRTL parameters of the binary systems a - c , a - y and c - y are used to predict phase equilibria in the ternary system a - c - y over a specified range of temperature. Thus, a regression is performed for NRTL parameters based on isothermal activity coefficients γ from COSMO-RS. (cf. Appendix A).
6. To limit the purification by distillation to a single distillation column, no ternary azeotropes and azeotropes between solvent and solute can be present. Thus, all azeotropes in the ternary system are calculated based on the algorithm proposed by Fidkowski et al. (1993). Solvent candidates with ternary azeotropes and azeotropes between solvent and solute are discarded.

All solvent candidates y that meet the requirements from Step 1 to Step 6 are considered valid solvent candidates $\tilde{y}$. For all valid solvent candidates, the process model is evaluated:

7. For each valid solvent candidate, the extraction unit operation is evaluated. The extraction unit operation is modeled using the pinch-based process model and solved using the procedure proposed by Redepenning et al. (2016). For a given feed composition, the procedure calculates the minimum amount of solvent $S_{\min}$, as well as all outlet compositions and flow rates (cf. Fig. 3.2).
8. Next, for each valid solvent candidate $\tilde{y}$, the distillation unit operation is evaluated. The distillation unit operation is modeled using the rectification body method (RBM) proposed by Bausa et al. (1998). Limitations by continuous molecular overflow assumption are overcome by considering the enthalpy of vaporization in the RBM pinch-based process model. Further contributions to the energy demand due to temperature changes in vapor and liquid phases are assumed to be negligible. Thus, no molar heat capacities are required in the calculation. The RBM pinch-based process model returns the minimum energy demand $Q_{\min}$. All process streams for the hybrid extraction-distillation process with all corresponding compositions (cf. Fig. 3.2) are now determined.
9. As a result, a list is returned containing all valid solvent candidates $\tilde{y}$ ranked by their respective minimum energy demand $Q_{\min}$.

3.3 Case Study: Purification of the Platform Chemical γ -Valerolactone

In this section, the proposed screening approach (cf. Section 3.2.3) is exemplified for the identification of novel solvents to purify γ -valerolactone (GVL). GVL is an intermediate in the production of bio-based value products and is a promising precursor for fuel and commodity chemicals (Alonso et al., 2013a). GVL can be produced from lignocellulosic biomass and needs to be recovered and purified from aqueous solutions for further processing. A recent approach for the recovery of GVL is a hybrid extraction-distillation process proposed by Murat Sen et al. (2012). In their work, Murat Sen et al. use n-butyl acetate as extraction solvent which is considered as benchmark in this case study.

A massive databank screening of more than 4600 solvent candidates is performed to identify novel promising solvents for GVL purification. The minimum energy demand

$Q_{\min}$ is used as the target function. For the solvent screening, the following process specifications are set (Step 1): A feed composition of $\mathbf{z}_F = (0.05, 0.95, 0.00)$ for the ternary system GVL-water- y is assumed; pressure and temperature in the process are set to $p = 1$ bar and $T = 25$ °C. Bio-compatibility of the solvent candidates y is desirable, thus, possible solvents candidates y are limited to molecules only containing carbon, hydrogen and oxygen atoms (Step 2).

All considered solvent candidates are tested to meet the required property constraints (Step 2). A binary LLE in the system water- y exists for 4331 solvent candidates y in the databank. Additionally, the boiling temperature $T_{B,y}$ is evaluated. Solvents with a boiling temperature higher than the predicted boiling temperature of GVL ($T_{B,\text{GVL}} = 507$ K) are discarded. Overall, 1715 solvent candidates also meet the required property constraint on the boiling temperature $T_{B,y}$ and are further evaluated. For those solvent candidates, pure- and mixture-data are computed (Step 4 and 5). The computational time to evaluate all 1715 solvent candidates y is ≈ 14 h on a 3.2 GHz desktop computer using 4 parallel cores. This computational time demand is almost exclusively due to the computation demand of COSMO-RS property prediction. In particular, the generation of (temperature-dependent) NRTL parameters is computationally expensive. The calculations can easily be accelerated using parallel computing. In contrast to COSMO-RS property prediction, the pinch-based process models are evaluated within minutes (≈ 5 min). The calculated NRTL parameters are stored and can be re-used. E.g., for the ternary system a - c - y , three sets of binary NRTL parameters are calculated: a - c , a - y and c - y . Binary NRTL parameters for the system a - c are independent from the solvent candidates y and have to be calculated only once. In addition, binary NRTL parameters for the systems c - y are independent from the solute a . Since the binary NRTL parameters for the systems c - y have already been calculated in this work for $C = \text{water}$, screening for a different solute a in water would require $\approx 50\%$ less time: Only binary NRTL parameters for the system a - y would have to be re-calculated for all solvent candidates y .

For all 1715 solvent candidates y , azeotropes are calculated and evaluated (Step 6). Overall, 1439 solvent candidates y meet the requirements from (Step 1 to 6) and thus are regarded valid solvent candidates $\tilde{y}$. For all valid solvent candidates $\tilde{y}$, the hybrid extraction-distillation process is evaluated (Step 7 and 8). As a result, all 1439 valid solvent candidates are ranked by the corresponding minimum energy demand $Q_{\min}$.

Fig. 3.4 shows the result of the solvent screening. Promising solvent candidates are identified in Tab. 3.1. In particular, 155 solvents are predicted to have a lower minimum energy demand $Q_{\min}$ than the literature benchmark n-butyl acetate. Tab. 3.1 shows that the top 5 identified solvents are aliphatic alkynes. The molecule with the

lowest minimum energy demand is 1,5-hexadiyne with 63 % reduction of $Q_{\min}$ (67 % reduction of $S_{\min}$) in comparison to the benchmark solvent n-butyl acetate. Thus, 1,5-hexadiyne is highly promising in terms of the energy saving potential.

All top solvents identified in Tab. 3.1 have both lower $Q_{\min}$ and lower $S_{\min}$ compared to the benchmark. This indicates that the extraction contributes significantly to the good performance of the top solvents in the distillation. A comparison of the ternary diagram of n-butyl acetate (benchmark) and 1,5-hexadiyne (best solvent identified) reveals the excellent extraction agent properties of 1,5-hexadiyne. Fig. 3.5 shows a larger miscibility gap and steeper tie lines for 1,5-hexadiyne compared to the benchmark. Thus, a high selectivity and a high capacity for the solute GVL are achieved which results in a low minimum amount of solvent $S_{\min}$. In turn, a low $S_{\min}$ conse-

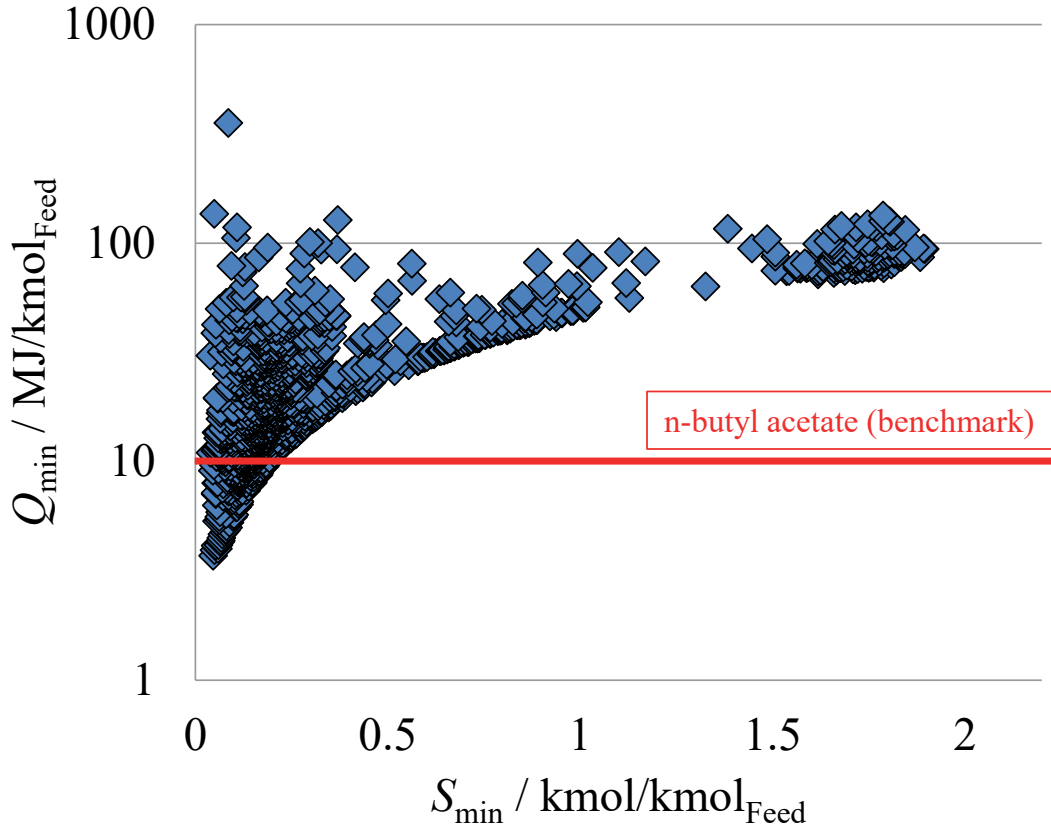


Figure 3.4: Minimum energy demand $Q_{\min}$ and minimum amount of solvent $S_{\min}$ for 1439 valid solvent candidates $\tilde{y}$ (blue diamonds) and $Q_{\min}$ for benchmark solvent n-butyl acetate (red line).

Table 3.1: Selected results for solvent screening using pinch-based process model and reduction of minimum energy demand compared to benchmark (BM). Available experimental boiling point temperature $T_{B,y}$ (exp.) and melting point $T_{\text{melt},y}$ (exp.) are provided for further solvent assessment.

Solvent	S_{min} kmol/kmol _{feed}	Q_{min} MJ/kmol _{feed}	Q_{min} reduct. %	Rank	$T_{\text{melt},y}$ (exp.) °C	Ref.	$T_{B,y}$ (exp.) °C	Ref.
1,5-hexadiyne	0.05	3.70	63	1	-4	(Rappoport, 1967)	88	(Rappoport, 1967)
1,6-heptydiyne	0.05	3.94	61	2	-85	(Rappoport, 1967)	112	(Rappoport, 1967)
1,3-hexadien-5-yne	0.06	3.99	61	3			82	(Sondheimer et al., 1961)
1,7-octadiyne	0.05	4.12	59	4			136	(Eyerett and Kon, 1950)
1-penten-4-yne	0.06	4.14	59	5			42	(Bubnov et al., 1980)
furan	0.07	4.64	54	9	-86	(Gontrani et al., 2006)	31	(Gontrani et al., 2006)
3-methylfuran	0.09	5.26	48	15			65	(Okabe et al., 2006)
2-methylfuran	0.11	6.36	37	27			64	(Rappoport, 1967)
2,3-hexanedione	0.10	6.98	31	34			128	(Heilbron et al., 1946)
3,4-hexanedione	0.11	7.15	29	36				
n-butylacetate	0.14	10.11	-	156	-74	(Rappoport, 1967)	126	(Feng et al., 1998)

quentially leads to a reduced extract stream E which reduces the minimum energy demand in the distillation $Q_{\min}$.

Rigorous process model calculations using Aspen Plus V8.4 are performed to validate the results from pinch-based process models. The minimum energy demand in Aspen Plus $Q_{\min, \text{ASPEN}}$ is calculated for a flowsheet according to Fig. 3.2. The flowsheet is initialized with results from the pinch-based process models and converged to minimum solvent flow and minimum reboiler heat duty (for calculation details see Appendix A).

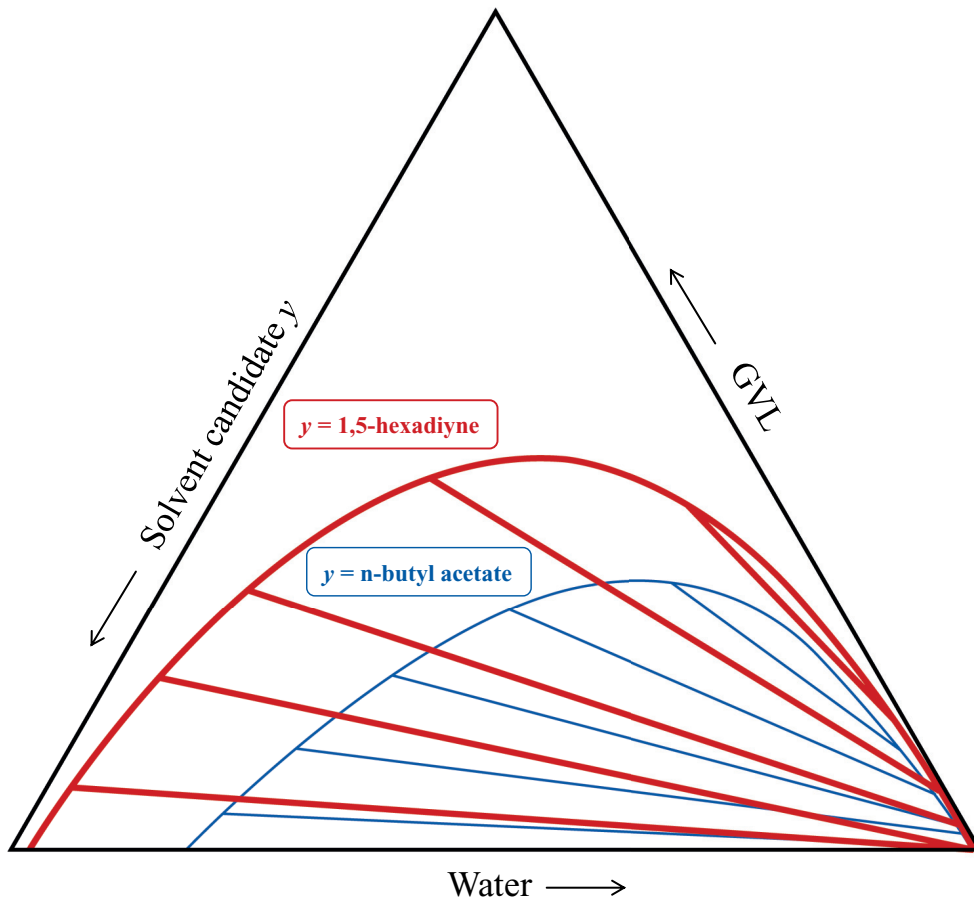


Figure 3.5: Ternary diagram of literature benchmark n-butyl acetate (blue thin line) and best solvent identified by solvent screening 1,5-hexadiyne (red thick line).

Overall, the results of the pinch-based process models are in good agreement with the rigorous process calculations. The mean average percentage error for the top 50 solvents is +9 %. In particular, the minimum energy demand in Aspen Plus for the best solvent identified 1,5-hexadiyne is $Q_{\min, \text{ASPEN}} = 3.97 \text{ MJ/kmol}_{\text{Feed}}$ (+7 %) and for the benchmark n-butyl acetate $Q_{\min, \text{ASPEN}} = 11.10 \text{ MJ/kmol}_{\text{Feed}}$ (+9 %). Additionally, the high selectivity and capacity of 1,5-hexadiyne for GVL is confirmed by the low minimum solvent demand in Aspen Plus ($S_{\min, \text{ASPEN}} = 0.05 \text{ kmol/kmol}_{\text{Feed}}$). Thus, rigorous process model calculations in Aspen Plus validate the predicted potential to reduce the minimum energy demand and determine a reduction by 64 % with 1,5-hexadiyne. This validation supports the presented approach for identification of promising solvents by the pinch-based process models. An extended list with validated results for the top 50 solvents is provided in Appendix A. The solvent ranking by minimum energy demand generated by the screening approach and the Aspen Plus simulations are in very good agreement with a Spearman's rank correlation coefficient of 0.93, where an ideal correlation corresponds to a rank correlation coefficient of 1 (Spearman, 1987). However, the Aspen Plus simulations also show that two components, trans-3-penten-1-yne and furan, ranked #7 and #9 by the screening (Tab. 3.1) require even less energy demand than 1,5-hexadiyne according to the Aspen Plus simulations. This effect is due to the different settings in the flowsheets (lower convergence tolerance and finite number of stages in Aspen Plus, see Appendix A). This finding also highlights the importance to not focus only on the top candidate from the screening but to consider the list of top candidates generated in the approach.

Only 59 of the 100 best solvents could have been evaluated using the modified UNIFAC (Dortmund) group-contribution method (Jakob et al., 2006). This is mainly caused by the lack of parametrized groups for promising furanes and missing interaction parameters for alkynes with water. This highlights a key feature of the presented approach that is independent of experimentally parametrized group interaction parameters.

To be a solvent of practical relevance, further criteria need to be evaluated next to minimum energy demand, e.g., design limits on melting point temperature or boiling point temperature as well as toxicity and chemical stability. These criteria are currently evaluated by human post-processing. For the present case study, all available experimental melting point temperatures indicate that the selected solvents are liquids at room temperature (Tab. 3.1). Similarly, experimental boiling points in Tab. 3.1 show that all components are within the desired design limits on boiling point temperature. Furthermore, alkynes have a tendency to decompose and are usually stored under low temperatures to prevent degradation. Thus, the alkynes ranked as top

solvents candidates in Tab. 3.1 seem questionable for practical application. However, an advantage of the proposed solvent screening approach is that a ranked list of solvents is generated (extended list of all top 155 solvents can be found in Appendix A). Further criteria for practical relevance can be evaluated by the design engineer.

For practical relevance, promising solvents are diketones and furanes which are already discussed in the literature as fuel candidates and solvents (Delidovich et al., 2014). Tab. 3.1 shows that promising furanes (#9, #15, #27) are identified in the solvent screening, which are stable under relevant process conditions according to manufacturer data (Sigma-Aldrich, 2017). The identified furanes reduce the minimum energy demand $Q_{\min}$ by $\approx 40\text{-}50\%$. The experimental boiling point temperature of furan is $T_{B,\text{furan}} = 31\text{ }^{\circ}\text{C}$ (Tab. 3.1) which is close to ambient temperature. Here, solvent loss by evaporation should be considered for practical applications. Future screening studies could add this constraint explicitly to the design problem. If further criteria such as toxicity are applied, 2,3-hexanedione (#34) seems to be a very promising candidate for GVL extraction. In contrast to the furanes, this diketone is non-toxic according to manufacturer data. 2,3-hexanedione reduces the minimum energy demand $Q_{\min}$ by 31% in comparison to the benchmark and is thus proposed as the most promising solvent from a practical perspective. Additionally, 2,3-hexanedione is commercially available and can further be evaluated experimentally.

3.4 Comparison to Conventional Process Performance Indicators

In this section, the results from solvent screening based on minimum energy demand are compared to results obtained using conventional process performance indicators. All identified 1439 valid solvent candidates $\tilde{y}$ are considered for the comparison. The following conventional screening criteria are evaluated (for calculation details see Appendix A):

1. Phase distribution coefficient P (at infinite dilution of solute a) is commonly used to assess the solvent extraction selectivity (e.g., Hostrup et al. (1999))
2. Relative volatility α is commonly used to assess the energy demand in the distillation (e.g., Pretel et al. (1994))

To quantify the correlation between conventional process indicators and the results from the pinch-based process models, the Pearson correlation coefficient $r_{m,n}$ is evaluated. The Pearson correlation coefficient determines the correlation of sample values

m_i and n_i . The Pearson correlation coefficient $r_{m,n}$ can take on values from $r_{m,n} = -1$ to $r_{m,n} = 1$. Pearson correlation coefficients of $r_{m,n} = -1$ and $r_{m,n} = 1$ correspond to an ideal linear correlation of m_i and n_i . In contrast, $r_{m,n} = 0$ indicates no correlation.

Fig. 3.6 shows a comparison of conventional process performance indicators to the results obtained from the screening approach based on minimum energy demand.

A very good correlation is found between phase distribution coefficient P and $S_{\min}$ (Fig. 3.6A). This is reflected in the high Pearson correlation coefficient $r_{\log P, \log S_{\min}} = -0.97$. Accordingly, all top solvents (green triangles in Fig. 3.6A) achieve high P

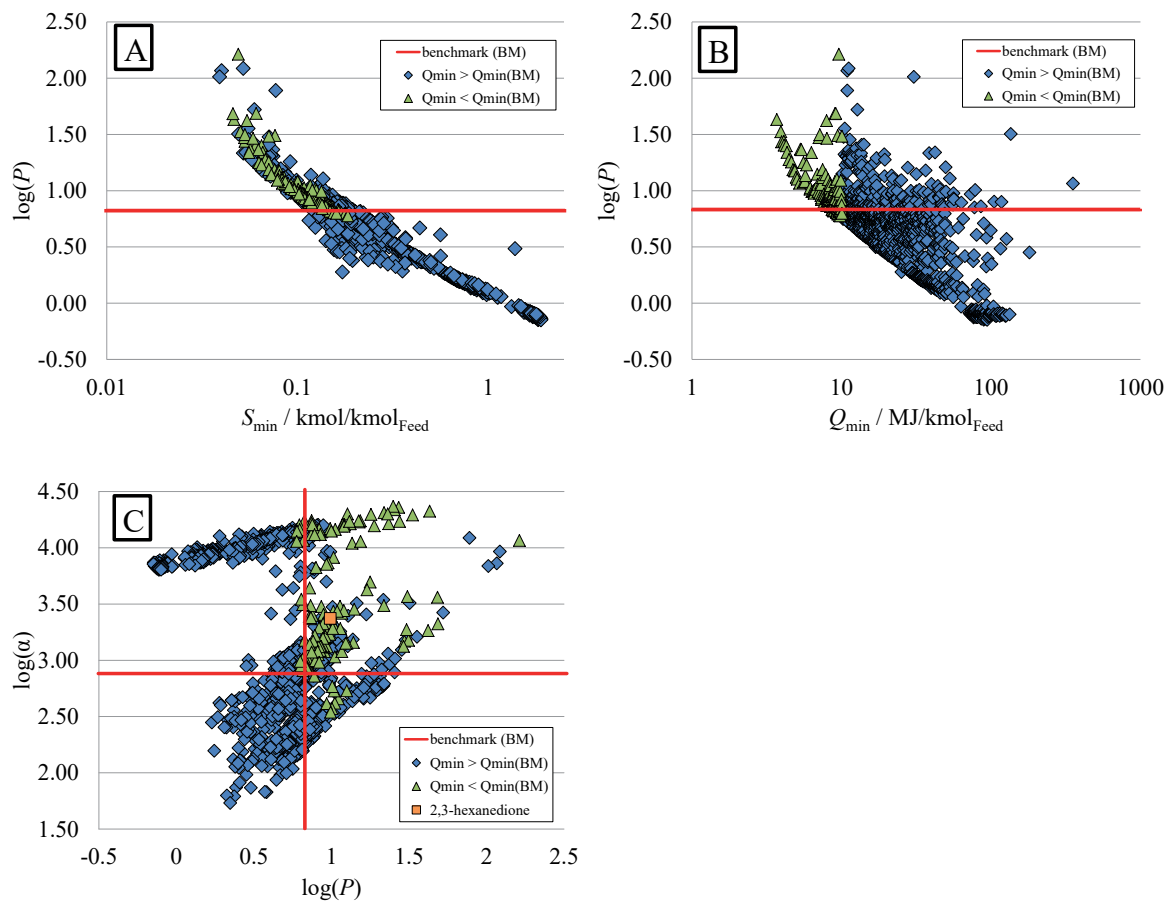


Figure 3.6: Comparison of phase distribution coefficient P to minimum amount of solvent $S_{\min}$ (A) and minimum energy demand $Q_{\min}$ (B). Solvent screening based on phase distribution coefficient P and relative volatility α (C). Red lines: benchmark (BM), green triangles: promising solvents ($Q_{\min} < Q_{\min}(\text{BM})$), blue diamonds: other solvents ($Q_{\min} > Q_{\min}(\text{BM})$), orange square: most promising identified solvent 2,3-hexanedione.

values. Similarly, Fig. 3.6B shows that P is moderately good correlated to the minimum energy demand $Q_{\min}$ ($r_{\log P, \log Q_{\min}} = -0.76$). A solvent screening based on phase distribution coefficient P would identify 2-methylphenol as the best solvent. 2-methylphenol can be considered as a good solvent candidate since it has a lower minimum energy demand $Q_{\min}$ than the benchmark molecule. However, a solvent assessment based on the phase distribution coefficient P has significant drawbacks. Firstly, the best solvent identified based on the phase distribution coefficient P has a $Q_{\min}$ of 9.63 MJ/kmol_{Feed} which is $\approx 160\%$ higher than the best solvent found in the screening based on pinch-based process models. The reason for the high minimum energy demand is the comparably high boiling point temperature predicted for 2-methylphenol $T_{B,2\text{-methylphenol}} = 201^\circ\text{C}$, which is in agreement with experimental findings (experimental $T_{B,2\text{-methylphenol}} = 192^\circ\text{C}$ (Brittain et al., 1981)). Secondly, Fig. 3.6B shows that a screening based on the phase distribution coefficient P does not identify *all* solvents with a lower minimum energy demand $Q_{\min}$ than the benchmark. Thus, promising solvents with a low phase distribution coefficient P would be missed in a solvent selection based on the phase distribution coefficient P .

Further refinement of solvent selection based on conventional process indicators commonly employs the relative volatility α (Pretel et al., 1994). The best identified solvent based on α is 2-methyl-1-buten-3-yne with a $Q_{\min} = 4.28$ MJ/kmol_{Feed}, which is 16 % higher than the best solvent found in the screening based on pinch-based process models. The good performance of 2-methyl-1-buten-3-yne can be explained by its low boiling point ($T_{B,2\text{-methyl-1-buten-3-yne}} = 51^\circ\text{C}$) and low enthalpy of vaporization ($\Delta h_{\text{vap},2\text{-methyl-1-buten-3-yne}} = 31.8$ kJ mol⁻¹). The low correlation of the relative volatility α and the minimum energy demand $Q_{\min}$ (correlation coefficient $r_{\log \alpha, \log Q_{\min}} = 0.24$) does not allow for an evaluation of the solvent process performance solely based on α . Yet, relative volatility α seems to be a good heuristic process performance indicator for further solvent selection in combination with distribution coefficient P (Fig. 3.6C). In a solvent screening based on both conventional process indicators P and α , most solvents with lower minimum energy demand $Q_{\min}$ than the benchmark are properly identified as promising. However, a solvent selection based on process indicators P and α still comes with major drawbacks: First, not all solvents with lower minimum energy demand $Q_{\min}$ than the benchmark are identified: Overall, 16 % of all promising solvent candidates are missed in a solvent selection based on P and α . Moreover, 116 components with a higher minimum energy demand $Q_{\min}$ than the benchmark are falsely identified as promising. Further conventional process performance indicators, i.e., constraints on the boiling point temperature, would thus be necessary to rule out more candidates which comes at the risk of losing further

promising solvent candidates. Second, there is no direct correlation between the conventional process indicators and $Q_{\min}$, i.e., it is not known whether favorable α is more important than favorable P . Importantly, many promising solvents with low $Q_{\min}$ are not located at extreme values of P or α but at intermediate values of P or α (cf. Fig. 3.6C). In particular, the solvents of practical relevance discussed above, e.g., 2,3-hexanedione with $\log P = 0.99$ and $\log \alpha = 3.37$, are identified at intermediate values of P and α (orange square in Fig. 3.6C). Thus, for a reliable identification of all promising solvents, a quantitative ranking of the solvents is necessary. This quantitative ranking by minimum energy demand $Q_{\min}$ is provided by the pinch-based process models.

3.5 Conclusions

A massive solvent screening approach based on COSMO-RS and pinch-based process models is presented. Automated evaluation of pinch-based process models for extraction and distillation devices was established to screen solvents by the minimum energy demand. The approach is not limited to a reduced number of components which is highlighted for a fully automated solvent screening of a large-scale database.

Results show that more than 4600 solvents can be screened fast and efficiently with the presented screening approach. Novel promising solvents for the hybrid extraction-distillation of GVL are identified with predicted better performance than the literature benchmark n-butyl acetate. A theoretical reduction of the minimum energy demand $Q_{\min}$ of 63 % is predicted for 1,5-hexadiyne and 31 % for 2,3-hexanedione. The most promising solvent candidate for practical application, 2,3-hexanedione, is commercially available and can further be evaluated experimentally.

A comparison to conventional screening criteria provides insight in the inherent trade-off of desired solvent properties in the process whereas heuristic criteria based on phase distribution coefficient or relative volatility lead to suboptimal solutions. The proposed screening approach captures this trade-off and yields quantitative information on process performance. Thus, the presented screening approach provides a comprehensive process-level assessment of the screened solvents and successfully takes into account inherent process trade-offs. Overall, the proposed solvent screening approach efficiently combines COSMO-RS property prediction with a comprehensive process-level assessment and thus significantly enlarges the range of current solvent selection approaches.

COSMO-CAMD: A Framework for Optimization-Based Computer-Aided Molecular Design using COSMO-RS

4.1 Introduction

In this chapter, the next step is taken towards integrating QM-based property prediction into computer-aided molecular and process design. To overcome limited-scope molecular screening (cf. Chapter 3), COSMO-CAMD is introduced, a framework for optimization-based molecular design with COSMO-RS. COSMO-CAMD integrates COSMO-RS calculations into molecular design based on thermodynamic property targets. Thereby, COSMO-CAMD links large-scope molecular structure design with large-scope thermodynamic modeling in using limited-scope property targets (Fig. 4.1). The COSMO-CAMD framework is presented and applied to two case studies for solvent design in liquid-liquid extraction. The first case study is the extraction of phenol from water, which illustrates the basic application of COSMO-CAMD. In the second case study, the extraction of hydroxymethylfurfural (HMF) from water is considered to show the application of the framework to a larger molecular design space.

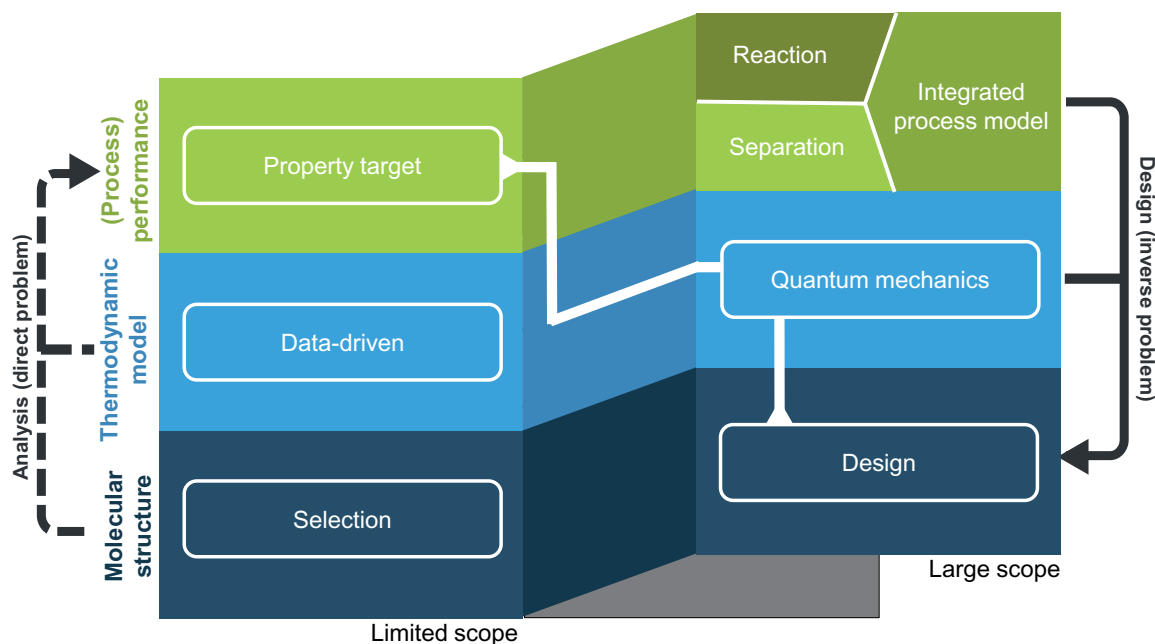


Figure 4.1: Content of this chapter in the broader CAMD/CAMPD context. Line with triangles indicates the classification of the COSMO-CAMD approach presented in this chapter. COSMO-CAMD combines molecular design (large-scope molecular structure exploration method) with QM-based COSMO-RS property prediction (large-scope thermodynamic model) and property design targets (limited-scope process performance model).

Major parts of this chapter have been published in:

J. Scheffczyk, L. Fleitmann, A. Schwarz, M. Lampe, A. Bardow and K. Leonhard. COSMO-CAMD: A Framework for Optimization-Based Computer-Aided Molecular Design using COSMO-RS, *Chemical Engineering Science*, 2017, 159, 84-92.

J. Scheffczyk, L. Fleitmann, A. Schwarz, A. Bardow, K. Leonhard, 2016. Computer-aided molecular design by combining genetic algorithms and COSMO-RS. In: Zdravko Kravanja, Milos Bogataj (Eds.), *Proceedings of the 26th European Symposium on Computer Aided Process Engineering*, volume 34 of *Computer Aided Chemical Engineering*, pages 115-120. Elsevier, Amsterdam.

4.2 COSMO-CAMD Framework

In this section, the COSMO-CAMD framework is introduced. COSMO-CAMD tackles the generic CAMD optimization problem (Eq. (2.1)) using COSMO-RS property prediction to evaluate the CAMD objective function $F(y)$ depending on the molecular structure y . To optimize these molecular structures, the genetic algorithm LEA3D (Douguet et al., 2005) is employed to overcome limits of molecule selection from prefixed databases. In the following sections, accuracy levels in COSMO-RS are discussed and selected for the optimization procedure (Section 4.2.1). Details on the genetic algorithm LEA3D are given (Section 4.2.2) and the full COSMO-CAMD procedure is introduced (Section 4.2.3).

4.2.1 Accuracy Levels in COSMO-RS

As outlined in Chapter 3, the COSMO-RS method consists of two main steps:

First, COSMO/DFT (Density Functional Theory) QM calculations are performed for each pure component i . Here, a molecule of species i is embedded in a virtual conductor. Through DFT calculations, the screening charge density σ for molecule i is calculated and the molecule is iteratively converged to its energetically optimal COSMO state. The resulting pure component σ -surface is stored in a so-called COSMO file and thus re-useable for calculating mixture properties.

Second, statistical thermodynamics methods are used to compute mixture properties (COSMO-RS). Based on the σ -surface, a histogram (σ -profile) is used to calculate the chemical potential (σ -potential) of a surface segment. From this σ -potential, thermodynamic properties, e.g., phase equilibrium data are computed. For a detailed description of COSMO and COSMO-RS see Klamt et al. (2001).

COSMO-RS has two approaches to generate σ -surfaces: The first approach is the generation of σ -surfaces using QM calculations. Here, different QM levels can be chosen to trade off computational effort and accuracy (Klamt et al., 2010):

- The standard QM level in COSMO is BP-TZVP (cf. Section 3.2.1), which is suitable for problems where number and size of compounds allows for full DFT geometry optimization;
- The QM calculation detail and accuracy can be increased by using BP-TZVPD-FINE: This QM level includes DFT geometry optimization on BP-TZVP level, with a consecutive single point COSMO calculation performed with a larger QM basis set BP-def2-TZVPD. Additionally, FINE-cavities are used for the single

point COSMO calculations, which improves concave surfaced molecules. BP-TZVPD-FINE represents the most accurate QM level available in COSMO-RS;

- In contrast, a reduction of computational demand from BP-TZVP is possible using BP-TZVP-MF. Here, σ -surfaces are generated on BP-TZVP level with initial conformer generation using semi-empirical methods, i.e., MOPAC (Stewart, James J. P., 2012).

The second approach to generate σ -surfaces is to bypass the computationally expensive QM calculations with the shortcut method COSMOfrag (Klamt et al., 2005). Here, no QM calculation is performed but σ -surfaces are generated from fragments of pre-calculated molecules in a database.

Depending on the selected approach to generate σ -surfaces and, accordingly, the chosen QM level, a trade-off exists between accuracy and computational time. To find a suitable trade-off, two QM levels are combined in this work: In the *Design Phase* (cf. Section 4.2.3), COSMO calculations are performed on BP-TZVP-MF level, which provides intermediate accuracy and computationally efficient generation of novel molecular structures. In the *Refinement Phase* (cf. Section 4.2.3) a selected set of molecular structures is calculated on BP-TZVPD-FINE, the most accurate QM level available in COSMO-RS.

This two-level procedure was chosen based on the evaluation of the accuracy and the computational performance of different σ -surface generation approaches for a representative set of solvent candidates. The ACS Green Chemical Institute (ACS GCI) Solvent Selection Guide (ACS Green Chemistry Institute, 2016) covers the most common and important solvents in pharmaceutical solvent selection and contains 64 molecules from which 31 form a liquid-liquid-equilibrium (LLE) with water. The distribution coefficients for the extraction of phenol and HMF from water were chosen as test cases. Phenol and HMF are the solutes in the case studies in the liquid-liquid extraction and the distribution coefficient is a key performance indicator in the objective functions for the optimization (cf. Section 4.3 and 4.4). The distribution coefficient P_i for a solute i between two LLE phases, extract E and raffinate R, is defined as

$$P_i = \frac{\mathbf{z}_{i,E}}{\mathbf{z}_{i,R}} = \frac{\gamma_{i,R}}{\gamma_{i,E}}. \quad (4.1)$$

The activity coefficients γ_i are calculated with COSMO-RS at the LLE concentration $\mathbf{z}_{i,E}$ and $\mathbf{z}_{i,R}$ of a binary water-solvent mixture with the solute at infinite dilution and $T = 25^\circ\text{C}$. In this work, COSMO calculations are performed using COSMOconf 4.0 (COSMOlogic, 2015b) with TURBOMOLE 6.5 for DFT geometry optimization;

COSMOtherm version C30_1501 is used for COSMO-RS calculations (COSMOlogic, 2015d). COSMOquick15 is used for COSMOfrag calculations (COSMOlogic, 2015c). Fig. 4.2 shows the trade-off between the average time needed per COSMO calculation for a molecule and the root mean square error (RMSE) (Eq. (4.2)) calculated for the distribution coefficient $\log P_i$ using different σ -surface generation approaches and the most accurate QM level BP-TZVPD-FINE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log P_{i,\text{BP-TZVPD-FINE}} - \log P_{i,X})^2} \quad (4.2)$$

Here, n is the number of solvents that form an LLE with water in BP-TZVPD-FINE level and $P_{i,X}$ is the distribution coefficient calculated by the different considered σ -surfaces generation approach X, where $X = \{\text{COSMOfrag}, \text{BP-TZVP-MF}, \text{BP-TZVP}\}$

The average calculation time for a COSMO calculation is ≈ 10 min per molecule in BP-TZVPD-FINE. Fig. 4.2 shows that the average calculation time per $\log P_i$ is reduced only slightly in BP-TZVP, intermediately in BP-TZVP-MF (factor ≈ 2) and significantly in COSMOfrag in comparison to BP-TZVPD-FINE. The error bars indicate that the maximum calculation time per COSMO calculation can significantly be larger than the average calculation time, in particular for BP-TZVPD-FINE and BP-TZVP calculations. Eckert and Klamt (2002) report the computation time for BP-TZVP COSMO calculations on a single CPU (800 MHz, PentiumIII) for water (30 s) and octylbenzene (2 h). Thus, these findings are in good agreement with reported literature data. The RMSE for BP-TZVP-MF and BP-TZVP is in the range of the known COSMO-RS accuracy (RMSE of COSMO-RS BP-TZVP level for experimental distribution coefficients $\log P_i$ has been stated as 0.3 (Klamt et al., 1998)). In contrast, the RMSE for COSMOfrag is significantly higher. Accordingly, BP-TZVP-MF is chosen for the *Design Phase* in COSMO-CAMD. To validate this choice, it is challenged and validated within the case studies with a larger set of data (cf. Section 4.3 and 4.4).

4.2.2 Genetic Algorithm LEA3D

In this work, molecular structures are generated by LEA3D, a genetic algorithm for molecular structure optimization (Douguet et al., 2005). LEA3D is based on LEA ('Ligand by Evolutionary Algorithm') that was developed for novel drug design (Douguet et al., 2000). The concept of LEA3D is based on a fragment-based molecular description, where molecular structures are decomposed into basic fragments and

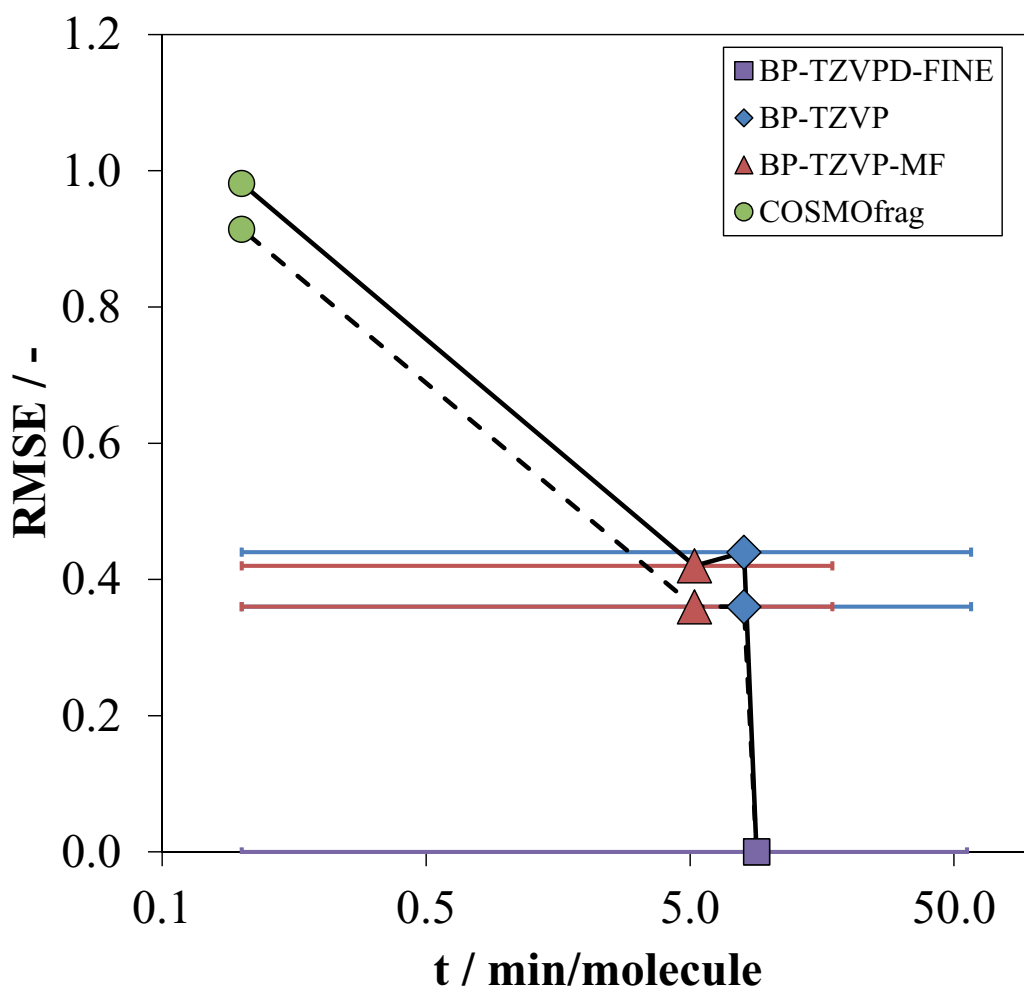


Figure 4.2: Root mean square error (RMSE) for calculating the distribution coefficient $\log P_i$ for a representative set of 31 molecules (ACS Green Chemistry Institute, 2016) from COSMO-RS vs. average time t for a COSMO calculation using different QM methods and COSMOfrag in comparison to BP-TZVPD-FINE. Calculations are performed on a 3.2 GHz desktop PC using 4 cores. Phenol: solid line, HMF: dashed line. Error bars indicate minimum/maximum time for a COSMO calculation.

each fragment is described by a unique representation (Bemis and Murcko, 1996). In LEA3D, genetic operators (cf. Section 2.2) search, extract and recombine molecular fragments in a *parent population* and generate a novel *breeding population*. The first generation is randomly initialized from a specified library of molecular fragments. Depending on the fragments in the initial library, large numbers of possible molecular structures can be designed and evaluated (e.g., a possible virtual library of 10^{14} molecular structures in Douguet et al. (2005)). The library fragments and parameters in this work for the optimization are included in Appendix B. For a detailed description of LEA3D, see Douguet et al. (2005).

4.2.3 COSMO-CAMD Procedure

The general procedure of COSMO-CAMD consists of the following steps (Fig. 4.3):

1. In the specification step, the LEA3D fragment library is initialized with the set of

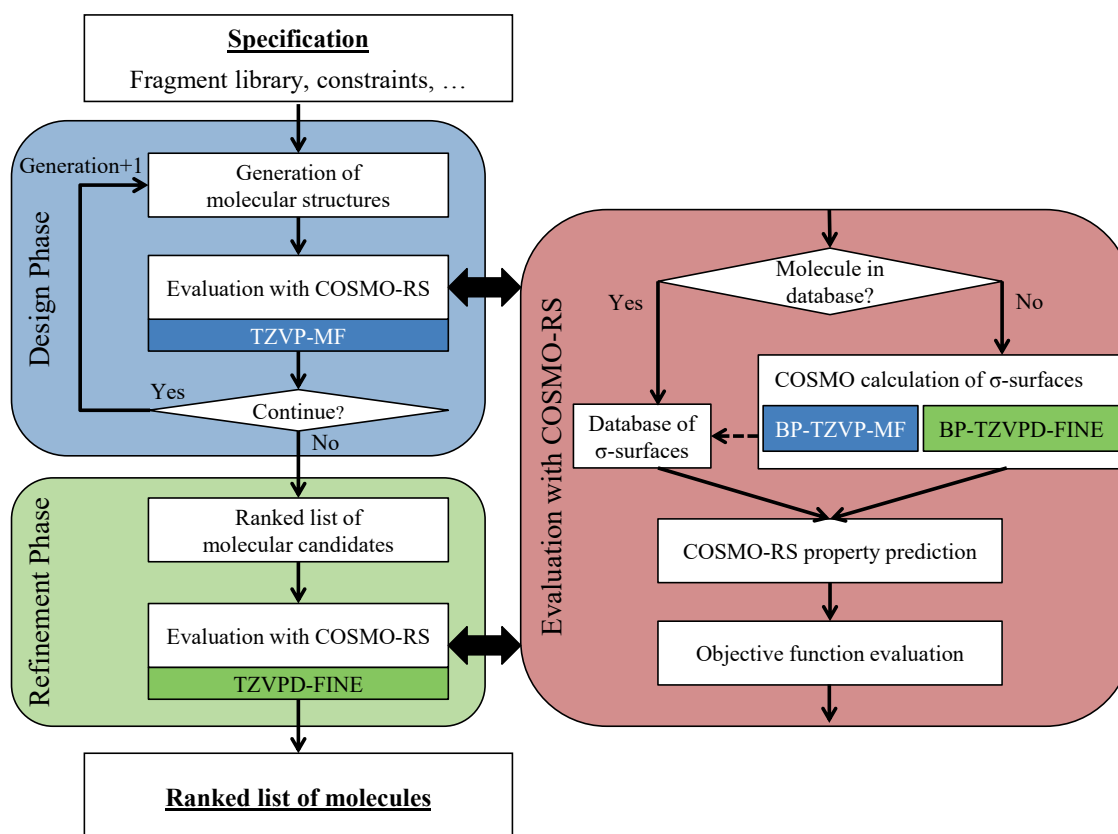


Figure 4.3: Schematic procedure of the COSMO-CAMD framework.

molecular fragments that can be combined by the genetic algorithm. Constraints are specified for molecular properties (e.g., molecule size) and parameters for the optimization are specified (# of generations, # of molecular structures per generation, probabilities for genetic operators).

2. In the *Design Phase*, the GA generates new molecular structures based on the current set of molecular structures using genetic operators (crossover and mutation). The first population is automatically randomly initialized from library fragments.
3. For each molecular structure, the objective function $F(y)$ (Eq. (2.1)) is evaluated based on thermodynamic property prediction from COSMO-RS using a medium accuracy BP-TZVP-MF QM level (cf. Section 4.2.1). While COSMO-RS provides mixture-specific properties depending on a specific solvent-solute system, σ -surfaces from COSMO calculations are properties of pure components (cf. Section 4.2.1). To increase computational efficiency, a database for σ -surfaces of novel molecular structures is generated during the optimization run. If the σ -surface of a molecular structure exists in the database, σ -surfaces are used from the database. If not, σ -surfaces are calculated on BP-TZVP-MF level. From σ -surfaces, COSMO-RS calculates the thermodynamic properties of the molecular structures in the specific mixture. Therefore, time-consuming COSMO calculations are performed only once for each molecular structure. Additionally, the σ -surface database can be re-used for following case studies, thus, further increasing computational efficiency. A generic optimization using specifications from the HMF extraction case study (cf. Section 4.4.1) was performed on an Intel Xeon CPU E5-1660 v3 @ 3.00 GHz using parallel computation of COSMO files on 8 cores. This allows for the design of ≈ 1000 novel molecules in 4 days leading to an average calculation time of 5.9 min per molecule.
4. The current generation is updated until the genetic algorithm reaches a termination criterion (e.g., after a specified number of generations). The result of the *Design Phase* is a preliminary ranked list of solvent candidates based on the evaluation of $F(y)$ for all molecules designed by the GA during the optimization.
5. In the *Refinement Phase*, the top solvent candidates from the *Design Phase* are selected. The objective function of the top solvent candidates is re-evaluated on BP-TZVPD-FINE level to obtain results with the highest possible accuracy in COSMO-RS. The final result is a ranked list of solvent candidates for further investigation (e.g., experiments for LLE (Dechambre et al., 2014b)).

4.3 Case Study 1: Solvent Design for Phenol Extraction

An industrially important extraction problem is the removal of phenol from wastewater (Kiezyk and Mackay, 1973). A screening for extracting agents by Gmehling and Schedemann (2014) has been carried out using an experimental database (Dortmund Data Bank, DDB). In this present work, novel solvents are designed for phenol extraction using the COSMO-CAMD framework introduced.

4.3.1 Problem Specification

Objective function $F_1(y)$ from Gmehling and Schedemann (2014) is used to evaluate the extraction-agent performance based on selectivity, capacity and solvent loss:

$$F_1(y) = \log \left(\underbrace{\left[\frac{P_{\text{phenol}}}{P_{\text{water}}} \right]^{1.0}}_{\text{selectivity}} \cdot \underbrace{[P_{\text{phenol}}]^{1.5}}_{\text{capacity}} \cdot \underbrace{(1 - \mathbf{z}_{\text{s,R}})^{3.0}}_{\text{solvent loss}} \right). \quad (4.3)$$

Here, P_i is the distribution coefficient for solute i between two LLE phases E and R according to Eq. (4.1). $\mathbf{z}_{\text{s,R}}$ is the concentration of solvent s in the raffinate phase R.

LEA3D only allows for unconstrained optimization, thus, constraints from Eq. (2.1) need to be incorporated in the objective function $F(y)$:

- For molecular structures, structural feasibility needs to be ensured $\rightarrow$ this is ensured by LEA3D,
- A LLE must exist between water and the solvent. This binary LLE is evaluated using the LLE calculation option in COSMOtherm $\rightarrow$ else $F_1(y)=0$.

Large molecules naturally are more likely to phase separate from water, which positively affects $F_1(y)$ (Eq. (4.3)). However, larger components have high melting points. To avoid high melting points, for each molecule, the number of heavy atoms n , i.e., all atoms except hydrogen, is restricted. For this purpose, LEA3D has built-in penalty functions to impose boundaries on structural properties (Douguet et al., 2005). In the LEA3D version used in this work (Version 2011v2), the available penalty function used to restrict molecular size is $F_2(y)$:

$$F_2(y) = \begin{cases} 1, & \text{if } n_{\min} \leq n \leq n_{\max} \\ \frac{(0.9n_X)^2 - (n - n_{\min})^2}{(0.9n_X)^2}, & \text{if } n < n_{\min} \vee n_{\max} < n \end{cases} \quad (4.4)$$

with $n_X = n_{\min}$, if $n < n_{\min}$ and $n_X = n_{\max}$, if $n_{\max} < n$.

Penalties for the molecular size are imposed by setting the number of heavy atoms to $[n_{\min} = 5, n_{\max} = 12]$.

If further case-specific solvent requirements are known, corresponding constraints could be added to the problem formulation to limit the molecular design space. E.g., restrictions on boiling point or viscosity can be evaluated using COSMO-RS and applied as constraints in LEA3D (Douguet et al., 2005). However, this chapter focuses on presenting the molecular design framework and employ target properties from the literature to validate the framework. Thus, in this generic case study, no further constraints are employed.

The final objective function is:

$$F(y) = -\frac{F_1(y) + F_2(y)}{2}. \quad (4.5)$$

Please note that the objective function $\frac{F_1(y) + F_2(y)}{2}$ in Eq. (4.5) is desired to be maximized while the general CAMD problem (Eq. (2.1)) is introduced as minimization problem. Thus, objective function in Eq. (4.5) is posed as minimization problem with inverted sign.

The initial fragment library consists of methyl-, ethyl-, propyl-, butyl-, phenyl-, carbonyl-, carboxyl-, ether-, hydroxy- and cyclohexylgroup fragments (cf. Appendix B). The number of generations is set to 20 with 40 molecules per generation based on pre-tests showing good performance. The top 25 molecules from the *Design Phase* are evaluated in the *Refinement Phase*.

4.3.2 Results

Tab. 4.1 shows the most promising solvents ranked by their BP-TZVPD-FINE objective function value $F(y)$. For all solvents in the top 50, the penalty function $F_2(y) = 1$. While a product form of the objective function is in general more flexible to combine several objectives, the findings of the case study show the successful application of the penalty function and support the problem formulation Eq. (4.5). In the database

solvent screening by Gmehling and Schedemann (2014), the most promising molecule identified is methylisobutylketone (MIBK). MIBK results in an objective function $F(y)$ of 4.24 ($F_1(y) = 7.48$) on BP-TZVPD-FINE level. A comparison to the solvents designed by COSMO-CAMD in Tab. 4.1 shows that the newly designed solvents are predicted to achieve higher objective function values $F(y)$ than MIBK. COSMO-CAMD designs a molecule with a maximum objective function value $F(y)$ of 4.76 ($F_1(y) = 8.52$), thus leading to an increase in extraction-agent performance $F_1(y)$ of 14 %. This increase is larger than the observed deviation of $F_1(y)$ on BP-TZVPD-FINE level from experimental data (Mean Absolute Percentage Error = 10 %) for the 10 solvents published in Gmehling and Schedemann (2014). Thus, the improvement is regarded as significant.

Generally, molecules designed by COSMO-CAMD achieving high performance are aprotic aromatic compounds (Tab. 4.1). Remarkably, benzaldehydes show even higher performance than acyclic ketones and aldehydes which have been identified by the database search in Gmehling and Schedemann (2014). Due to the physical basis of COSMO-RS, σ -profiles can be used to study polarity characteristics of the most promising molecular compounds (Fig. 4.4): The σ -profiles are divided into 3 regions: a non-polar region with $-0.01 \text{ e}/\text{\AA}^2 < \sigma < 0.01 \text{ e}/\text{\AA}^2$, the hydrogen bond (HB) donor region with $\sigma < -0.01 \text{ e}/\text{\AA}^2$ and the HB acceptor region with $\sigma > 0.01 \text{ e}/\text{\AA}^2$. The HB donor/acceptor strength corresponds to the peaks beyond $\sigma = -0.01 \text{ e}/\text{\AA}^2$ and $0.01 \text{ e}/\text{\AA}^2$, respectively. The larger the peak, the more σ -surface segments are present at a given charge and accordingly, the number of possible HB interactions is increased; the further the peak is located away from center, the stronger the HB interactions. In contrast, non-polar segments have peaks centered at $\sigma = 0 \text{ e}/\text{\AA}^2$.

It can be seen from Fig. 4.4 that the molecule designed by COSMO-CAMD (2-methyl-4-(propan-2-yl)benzaldehyde) with the highest objective function $F(y)$ shows similar polarity to MIBK, the best solvent in Gmehling and Schedemann (2014). Both solvents are aprotic solvents and HB acceptors, which is reflected in the high probability density around $\sigma = 0.015 \text{ e}/\text{\AA}^2$. However, longer molecules tend to increase

Table 4.1: Ranked list of solvents designed by COSMO-CAMD in BP-TZVP-MF (MF) and BP-TZVPD-FINE (FINE).

Rank	component	$F(y)$ MF	$F(y)$ FINE
1	2-methyl-4-(propan-2-yl)benzaldehyde	4.31	4.76
2	2-methyl-5-propylbenzaldehyde	4.27	4.75
3	4-methyl-2-propylbenzaldehyde	4.29	4.74
4	2-methyl-5-(propan-2-yl)benzaldehyde	4.26	4.72

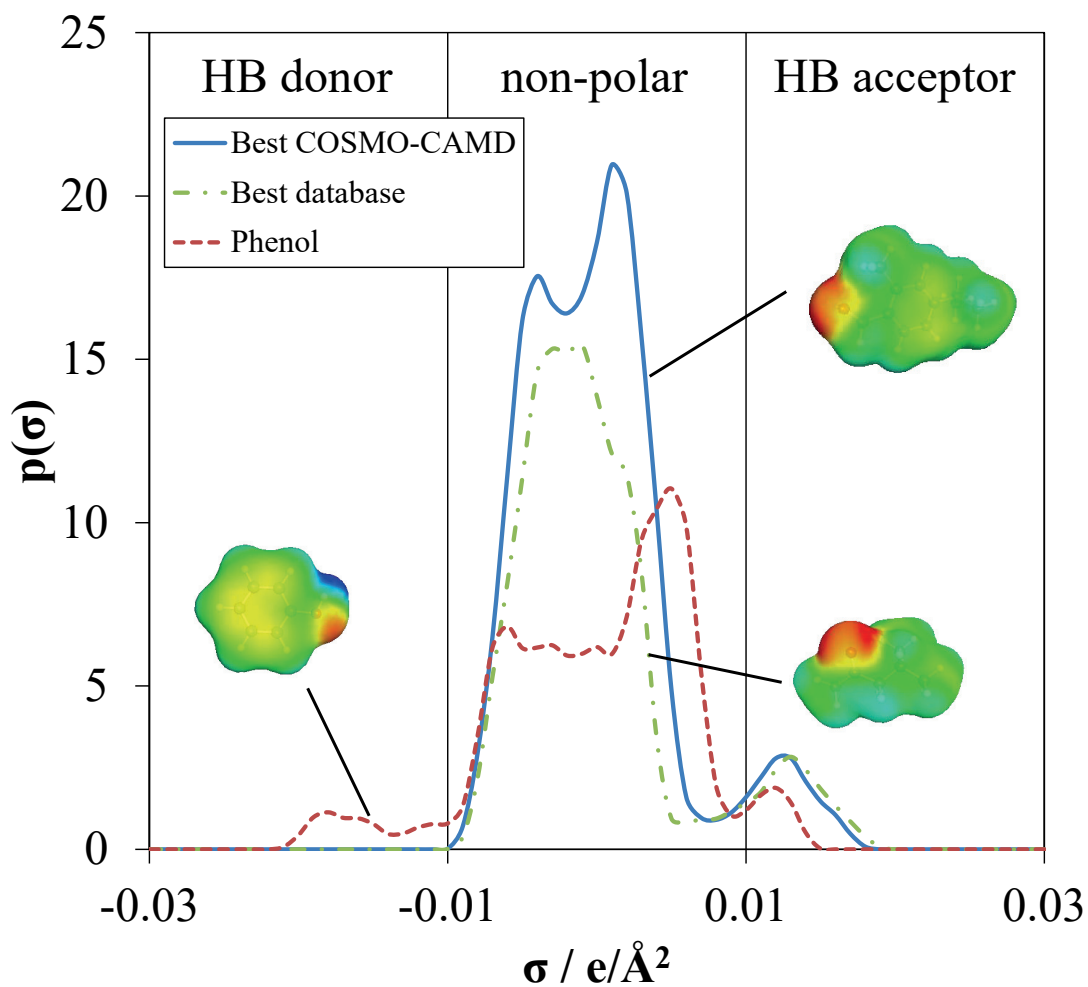


Figure 4.4: σ -profiles of best COSMO-CAMD molecule 2-methyl-4-(propan-2-yl)benzaldehyde (solid blue line), best molecule from database search MIBK (Gmehling and Schedemann, 2014) (dashed-dotted green line) and solute phenol (dotted red line).

the miscibility gap due non-polar surface interaction formed by the alkyl side chains. This hydrophobic region (centered at $\sigma = 0 \text{ e}/\text{\AA}^2$) is beneficial, since it reduces the amount of co-extracted water (Garcia-Chavez et al., 2012) and has a direct influence on the objective function $F(y)$ (Eq.(4.3)). This region is increased for 2-methyl-4-(propan-2-yl)benzaldehyde in comparison to MIBK, thus, leading to a higher objective function $F(y)$.

To validate the convergence of the GA optimization, the optimization run is repeated. Fig. 4.5 shows the mean and maximum objective function values of three optimization runs. The maximum objective function values show an increase in the

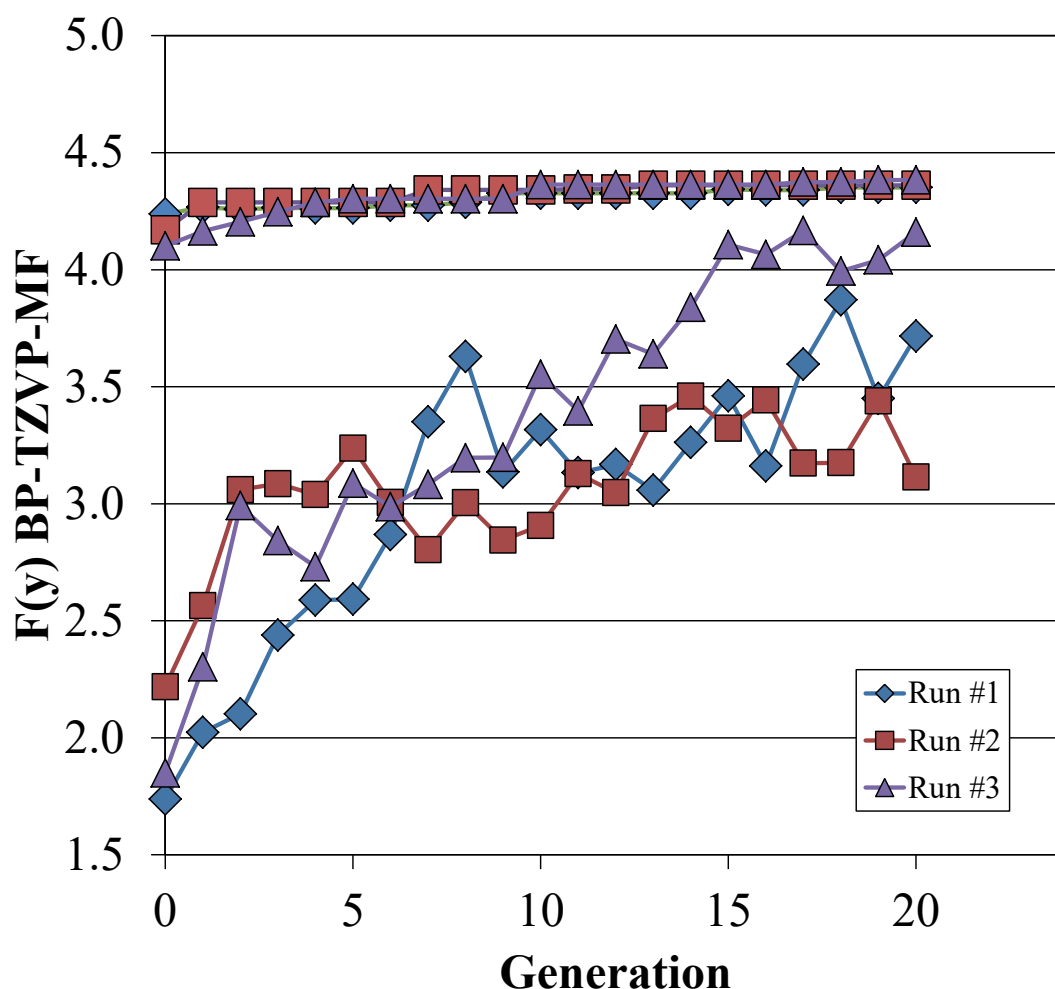


Figure 4.5: Evolution of the mean (lower curves) and maximum (upper curves) objective function values $F(y)$ based on BP-TZVP-MF in three runs of the genetic optimization algorithm.

first generations and converge to a steady final value. The molecules with the highest objective function values designed in all three optimization runs are similar structural isomers with a difference in objective function values of less than 1%. Due to the stochastic nature of the GA optimization, no guarantee on (global) optimality of the solutions can be given. However, the results in Fig. 4.5 indicate that no major further improvement of the objective function is expected if the number of generations is increased.

As discussed in Section 4.2.1, BP-TZVP-MF is chosen as an intermediate accuracy QM level in the *Design Phase* of the COSMO-CAMD framework. To challenge and validate the use of BP-TZVP-MF for this case study, a comparison for two optimization runs is shown in Fig. 4.6 using either COSMOfrag or BP-TZVP-MF. Here, the objective function $F(y)$ values is compared using both QM levels for all designed molecules, which exist in the BP-TZVPD-FINE database or were calculated in the *Refinement Phase*.

The results show considerable deviation ($\text{RMSE} = 0.61$) between COSMOfrag and BP-TZVPD-FINE (Fig. 4.6A) which is in agreement with the initial findings in Section 4.2.1. In general, COSMOfrag is able to give a qualitative picture about promising solvent groups (phenols, ketones and aldehydes). However, the use of COSMOfrag leads to the identification of suboptimal molecules in comparison to the database. The top molecule designed in COSMOfrag is di-n-butylether (DNBE) with an objective function in COSMOfrag of $F(y) = 5.14$. However, the *Refinement Phase* reveals that DNBE has an objective function of $F(y) = 4.19$ in BP-TZVPD-FINE whereas the molecule with the largest objective function value in BP-TZVPD-FINE is cyclohexanecarboxaldehyde ($F(y) = 4.61$). Thus, despite significant time-saving potential (cf. Section 4.2.1), COSMOfrag is not sufficiently accurate for the present case study.

In contrast, BP-TZVPD-FINE values are in good agreement with BP-TZVP-MF (Fig. 4.6B). The calculated RMSE value between BP-TZVP-MF and BP-TZVPD-FINE values is 0.27. Accordingly, the deviation between BP-TZVP-MF and BP-TZVPD-FINE is in the range of the initially estimated accuracy (cf. Section 4.2.1). Due to the known accuracy of COSMO-RS (cf. Section 4.2.1), the candidate list should in any case serve as an input for experiments to verify the predicted performance.

Overall, novel solvents are designed by COSMO-CAMD: 18 of the most promising 25 molecules designed by COSMO-CAMD are not contained in the COSMO-RS database (Version COSMObase-1501-BP-TZVPD-FINE). Therefore, COSMO-CAMD successfully goes beyond the established screening methods based on COSMO-RS database search.

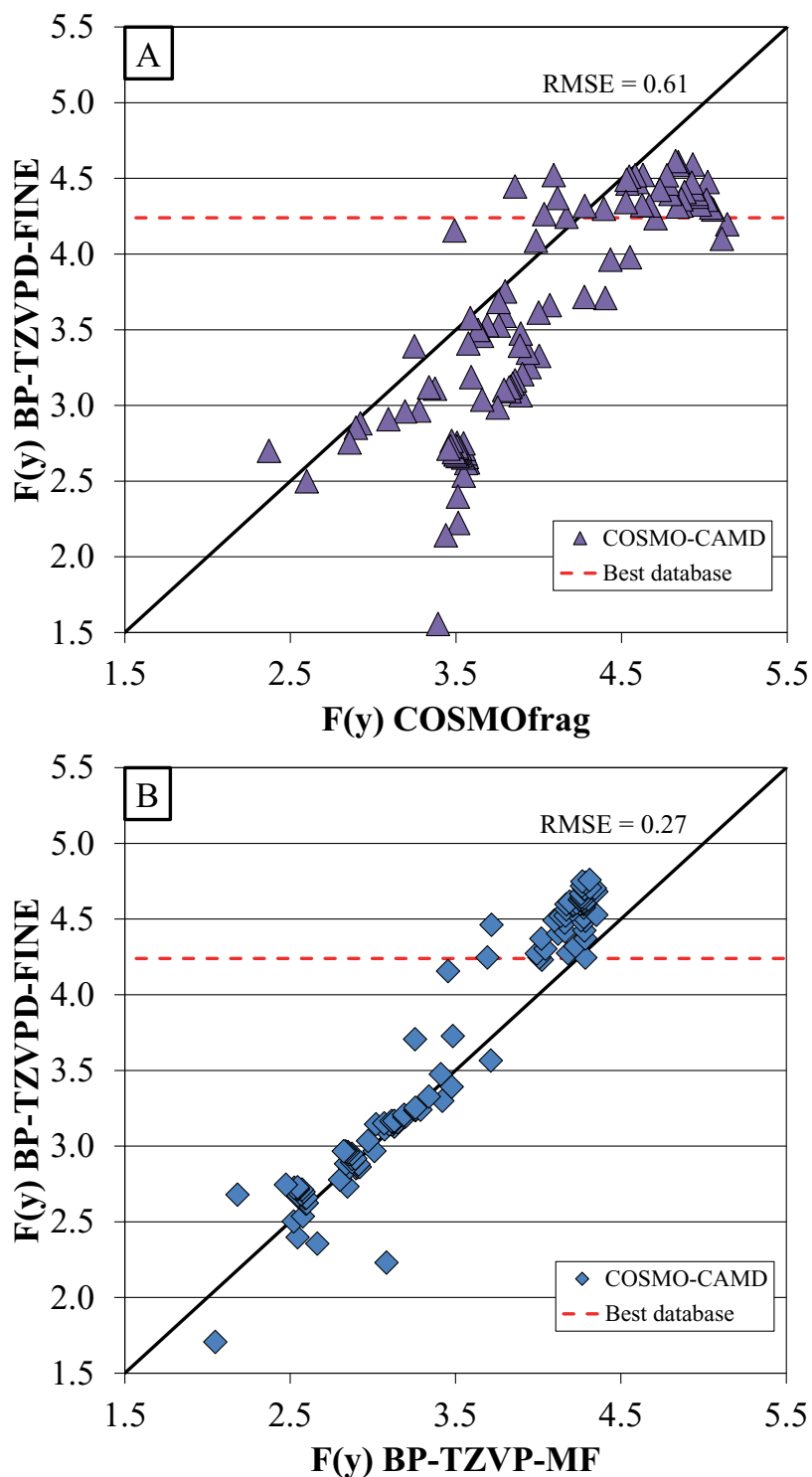


Figure 4.6: Molecules generated in COSMO-CAMD using COSMOfrag (A) and BP-TZVP-MF (B) in comparison to refined molecules on BP-TZVPD-FINE level. Dashed line indicates best solvent from database screening MIBK (Gmehling and Schedemann, 2014). The RMSE (Eq. (4.2)) is calculated between COSMOfrag/BP-TZVP-MF and BP-TZVPD-FINE for all available BP-TZVPD-FINE values.

4.4 Case Study 2: Solvent Design for HMF Extraction

HMF is a highly promising chemical as intermediate in bio-based fuels and chemical production (Alonso et al., 2013a). However, a bottleneck in the large-scale production of HMF is the liquid-liquid extraction due to low product concentrations after the reaction (van Putten et al., 2013). Continuous HMF extraction during the reaction can be used for in-situ removal to significantly increase selectivity suppressing the formation of undesired humins and side-products (Román-Leshkov et al., 2007).

A systematic solvent screening for HMF extraction based on COSMO-RS database components has been carried out by Blumenthal et al. (2016). In their work, experimental data are in very good agreement with COSMO-RS predictions, which shows the applicability of COSMO-RS to HMF extraction. In the present work, novel solvents are designed for HMF extraction using the COSMO-CAMD framework introduced.

4.4.1 Problem Specification

For the evaluation of extraction efficiency, Blumenthal et al. (2016) use a mass-based distribution coefficient K_{HMF} , which can be derived from the mole-based distribution coefficient P_i (Eq. (4.1)).

$$K_{\text{HMF}} = \frac{\mathbf{w}_{\text{HMF,E}}}{\mathbf{w}_{\text{HMF,R}}} = \frac{\gamma_{\text{HMF,R}} \cdot \sum_{i=1}^2 (\mathbf{z}_{i,\text{R}} \cdot M_{i,\text{R}})}{\gamma_{\text{HMF,E}} \cdot \sum_{i=1}^2 (\mathbf{z}_{i,\text{E}} \cdot M_{i,\text{E}})} \quad (4.6)$$

Here, $\mathbf{w}_{\text{HMF,E}}$ and $\mathbf{w}_{\text{HMF,R}}$ are weight fractions of the solute HMF in the extract E and raffinate R phases respectively. M_i are molecular weights of the components. The binary water-solvent mixture is assumed to be present with the solute at infinite dilution and $T = 25^\circ\text{C}$. Due to the assumption of infinite dilution, mole fractions of the solute HMF are $\mathbf{z}_{\text{HMF,E}} = \mathbf{z}_{\text{HMF,R}} \approx 0$. For this case study, the objective function $F(y)$ is defined based on K_{HMF} (Eq. (4.6)):

$$F(y) = -\log K_{\text{HMF}}. \quad (4.7)$$

As discussed in Section 4.3.1, constraints for the optimization $h(y, g)$ and $c(y)$ from Eq. (2.1) are included as penalty functions in the objective function $F(y)$ (Eq. (4.7)). In the first case study, a soft constraint penalty function is employed (Eq. (4.4)). This approach may still require the evaluation of large and complex molecules. For

such molecules, the time for COSMO calculations can be considerably longer than average calculation times (cf. Section 4.2.1). To further increase computational speed, molecular size is restricted with a hard constraint in this case study.

- For molecular structures, structural feasibility needs to be ensured $\rightarrow$ this is ensured by LEA3D,
- A LLE must exist between water and the solvent. This binary LLE is evaluated using the LLE calculation option in COSMOtherm $\rightarrow$ else $F(y)=0$,
- Molecular structures with undesired properties (e.g., unstable) are excluded with a restriction on the molecular structures in SMILES notation (Weininger, 1990): Peroxides, di-ketones, anhydrides und oxalates as well as oxygenfluorides. For all molecular structures containing corresponding SMILES strings $\rightarrow F(y)=0$,
- The number of heavy atoms n in the molecular structures n is constrained to $[n_{min}=1, n_{max}=12]$ to avoid larger components with high melting points. Thus, for molecules outside the target range $\rightarrow F(y)=0$.

For this case study, the initial library from the phenol case study (cf. Section 4.3.1) is complemented with additional fragments to include ester-, fluorine- and amingroups (cf. Appendix B). Hereby, the effect of more functional groups, especially highly polar fluorine groups is investigated.

An additional challenge for property prediction in HMF extraction is that besides K_{HMF} , the distribution coefficient of fructose K_{fructose} plays an important role (Blumenthal et al., 2016): In HMF synthesis, the co-extraction of fructose leads to loss of the reactant and is therefore undesirable. Fructose is a challenging molecule for property prediction, since it has many so-called conformers (energetically stable molecular geometries). In COSMO-RS, these conformers occur with/without intramolecular hydrogen bonds depending on the surrounding medium (nonpolar/polar respectively). A COSMO-RS fructose model was generated based on the work of Blumenthal et al. (2016) using COSMOconfX13. The distribution coefficient of fructose in the extraction of HMF is accounted for in the *Refinement Phase*. To account for a larger search space in comparison to Section 4.3.1, the number of generations is set to 100 generations with 40 molecules in each generation, 50 molecules are refined in the *Refinement Phase*.

4.4.2 Results

In the work of Blumenthal et al. (2016), solvents are selected based on their ability to enhance the extraction of the desired product HMF (high K_{HMF}), while still providing

Table 4.2: Ranked list of solvents designed by COSMO-CAMD with primary objective function $\log(K_{\text{HMF}})$ and secondary objective function $\log(K_{\text{fructose}})$

Rank	component	$\log(K_{\text{HMF}})$	$\log(K_{\text{fructose}})$
1	2-ethenyl-4,5-difluorophenol	1.41	-0.29
2	2-ethenyl-5-fluorophenol	1.38	-0.38
3	2,4-diethenyl-5-fluorophenol	1.37	-0.64
11	2,5-diethenyl-4-fluorophenol	1.30	-0.80
18	3-fluoro-6-(propan-2-yl)benzene-1,2-diol	1.24	-0.72

a low co-extraction of fructose (low K_{fructose}). Blumenthal et al. (2016) find that phenolic compounds show the most promising behavior. The proposed candidate for HMF extraction is o-propylphenol with $K_{\text{HMF}} = 10.02$ which corresponds to an objective function $F(y) = 1.00$ in BP-TZVPD-FINE.

Tab. 4.2 shows that the most promising components designed by COSMO-CAMD with the highest objective function $F(y)$ are fluorinated phenolic compounds. This is in very good agreement with findings from the database screening of Blumenthal et al. (2016), where phenolic compounds are identified as top molecules. Adding highly polar fluorine atoms as functional groups favorably influences the extraction behavior of HMF (Tab. 4.2, #1 and #2). This effect can be explained by the withdrawal of negative charge from the phenol hydroxyl-group, which increases its ability to act as a protic HB donor, thus, stabilizing HMF in the extraction phase.

Inherently similar characteristics of HMF and fructose consequently lead to a co-extraction of fructose, which is reflected in high distribution coefficients K_{fructose} for the top molecules. Here, functional group substitution of a fluorine group by an alkenyl-sidechain (Tab. 4.2, #11) or alkyl-sidechains (Tab. 4.2, #18) leads to significantly lower co-extraction of fructose, yet still offering a promising high distribution coefficient $\log K_{\text{HMF}}$. Thus, for practical purposes the most promising solvent identified by COSMO-CAMD is 2,5-diethenyl-4-fluorophenol (Tab. 4.2, #11), which allows for a high distribution coefficient of HMF ($\log K_{\text{HMF}} = 1.30$) and comparable low undesired co-extraction of fructose ($\log K_{\text{fructose}} = -0.80$). Future work should consider the distribution coefficient for fructose directly in the *Design Phase*. Further questions of practical relevance could then also be added to the design problem.

Additional effects, e.g., finite solute concentrations, could be investigated for the most promising molecules. Such effects have been shown to influence quantities predicted by COSMO-RS such as the distribution coefficient (Blumenthal et al., 2016). However, Kossack et al. (2008) showed that the assessment of solvent performance based on the distribution coefficient at infinite dilution can provide a reasonable esti-

mate for early-stage process design. In addition, COSMO-CAMD not only provides one single best molecule that could be influenced by finite solute concentration effects but a large set of promising molecules. Overall, it is expected that the promising molecule candidate identified by COSMO-CAMD will also show good performance at finite concentrations.

A comparison of BP-TZVP-MF to BP-TZVPD-FINE level for this case study

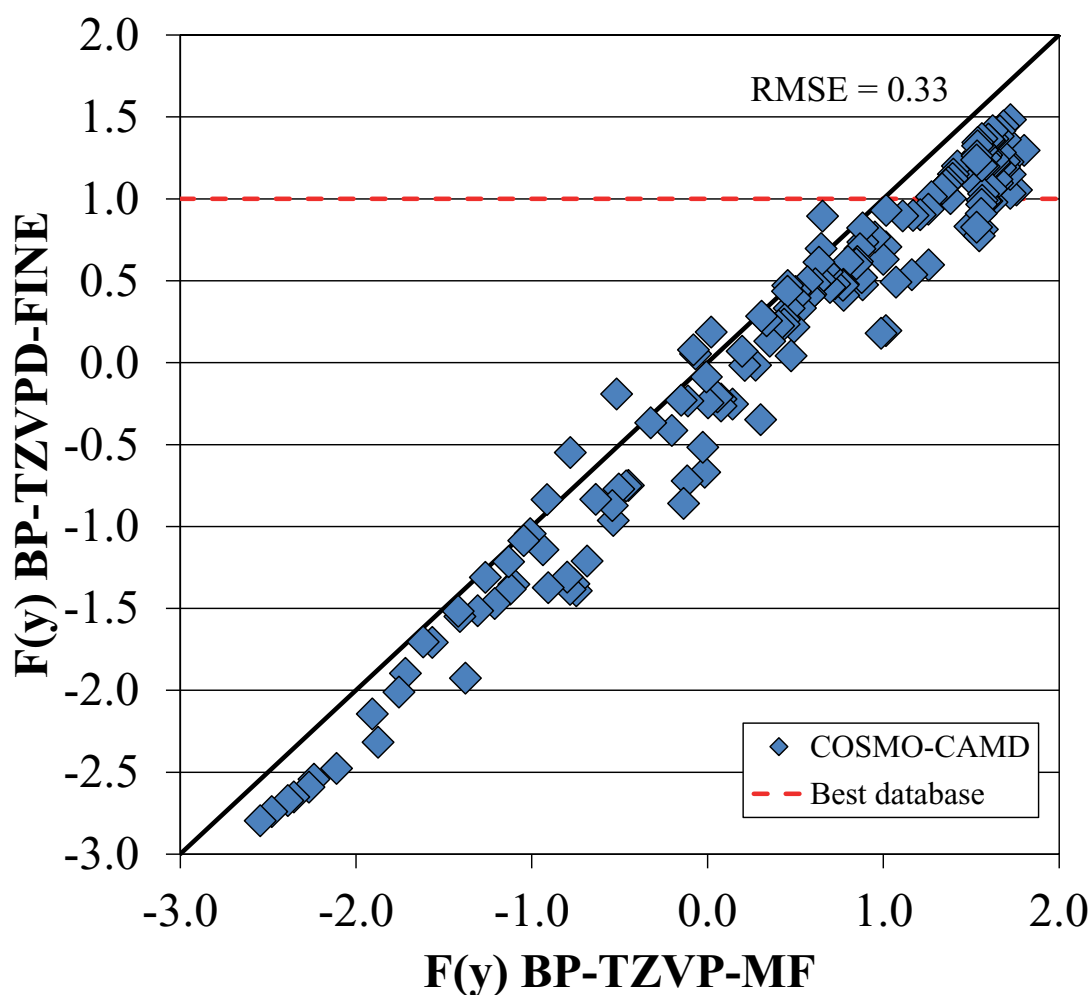


Figure 4.7: Objective function values $F(y)$ for molecules generated in COSMO-CAMD using BP-TZVP-MF in comparison to refined molecules on BP-TZVPD-FINE level. Dashed line indicates best solvent from database screening o-propylphenol (Blumenthal et al., 2016). The RMSE (Eq. (4.2)) is calculated between BP-TZVP-MF and BP-TZVPD-FINE for all available BP-TZVPD-FINE values.

(Fig. 4.7) shows good agreement of the chosen intermediate QM level to BP-TZVPD-FINE. The RMSE = 0.33 between BP-TZVP-MF and BP-TZVPD-FINE is similar to the RMSE for the solvents evaluated in the ACS GCI set of common solvent candidates (cf. Section 4.2.1). Similar to the findings in the extraction of phenol (cf. Section 4.3.2), the good agreement between BP-TZVP-MF and BP-TZVPD-FINE supports the initial choice of BP-TZVP-MF as a medium accuracy QM method. Comparing the top 50 molecules to the COSMO-RS database (Version COSMObase-1501-BP-TZVPD-FINE) shows that all top 50 molecules generated by COSMO-CAMD are not included in the database. Thus, COSMO-CAMD successfully designs novel molecules with predicted high objective function values $F(y)$. In addition, the *Refinement Phase* allows for the evaluation of secondary objectives, e.g., the co-extraction of fructose.

4.5 Conclusions

In this chapter, COSMO-CAMD is presented, an optimization-based framework for designing novel solvents based on property prediction with COSMO-RS. The framework combines a computationally efficient *Design Phase* using medium accuracy QM methods with most accurate property prediction for promising molecules in a *Refinement Phase*. The generation of a re-usable σ -surface database during the optimization further increases computational efficiency.

Novel molecules are generated by direct manipulation of the chemical structure within the genetic algorithm LEA3D. The resulting framework allows to explore a large search space (+1000 molecules within days) by optimization. The use of COSMO-RS facilitates property prediction without the need for experimental data.

The case studies for phenol and HMF liquid-liquid extraction from water show that COSMO-CAMD successfully designs promising novel solvents, predicted to have better performance than molecules contained in comprehensive databases (experimental and COSMO-RS): 18 of the most promising 25 designed molecules in phenol extraction and all top 50 molecules in HMF extraction are novel molecules, which had not been in the COSMO-RS BP-TZVPD-FINE database. Thus, COSMO-CAMD successfully overcomes limitations of both, experimental and COSMO-RS-based database screenings.

COSMO-CAMPD: A Framework for Integrated Design of Molecules and Separation Processes based on COSMO-RS

5.1 Introduction

In this chapter, the last step towards integrating COSMO-RS in CAMPD is taken. For this purpose, COSMO-CAMPD is presented, a framework for computer-aided molecular and process design based on COSMO-RS. For COSMO-CAMPD, COSMO-CAMD (cf. Chapter 4) is combined with pinch-based process models that allow for large-scale, automated process evaluation of molecules (cf. Chapter 3). Thereby, large-scale process performance models are linked to large-scale thermodynamic prediction models in a large-scale design approach (Fig. 6.1). The resulting COSMO-CAMPD framework is a hybrid stochastic-deterministic optimization method for the integrated design of molecules and processes. In this chapter, the COSMO-CAMPD framework is presented and subsequently applied to a case study for integrated solvent and process design in a hybrid extraction-distillation process. The results of COSMO-CAMPD are discussed and finally, the predicted process performance of promising designed solvents is experimentally validated by liquid-liquid measurements.

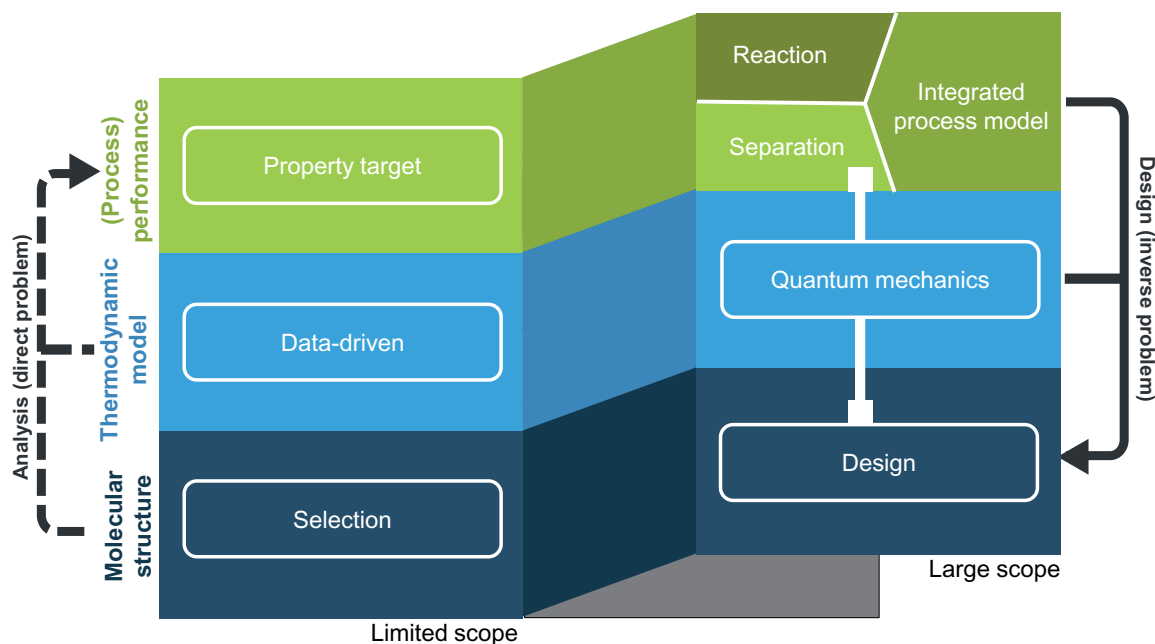


Figure 5.1: Content of this chapter in the broader CAMD/CAMPD context. Line with squares indicates the classification of the COSMO-CAMPD approach presented in this chapter. COSMO-CAMPD combines molecular design (large-scope molecular structure exploration method) with QM-based COSMO-RS property prediction (large-scope thermodynamic model) and pinch-based separation process models (large-scope process performance model).

Experimental measurements in this chapter have been performed by Julia Thien. Major parts of this chapter have been published in:

J. Scheffczyk, P. Schäfer, L. Fleitmann, J. Thien, K. Leonhard and A. Bardow. COSMO-CAMPD: A framework for integrated design of molecules and processes based on COSMO-RS, *Molecular Systems Design and Engineering*, 2018 - Reproduced by permission of The Royal Society of Chemistry.

5.2 COSMO-CAMPD Framework

5.2.1 Integrated Molecular and Process Design Approach

In this section, COSMO-CAMPD is presented which addresses the full CAMPD MINLP optimization problem (Eq. (2.2)). For this purpose, pinch-based process models introduced in Chapter 3 are combined with molecular optimization by COSMO-CAMD (Chapter 4). The resulting hybrid stochastic-deterministic algorithm solves the CAMPD problem by exploiting a natural problem decomposition: integer variables (molecular variables) are optimized by a stochastic algorithm and continuous variables (process variables) are optimized using a deterministic non-linear programming (NLP) algorithm. The resulting COSMO-CAMPD optimization algorithm can be divided in 4 general steps (Fig. 5.2):

1. The optimization procedure is initialized by specifying, e.g., degrees of freedom for the optimization: process variables, process operating limits $h(x, y)$ in Eq. (2.2) and an initial set of molecular fragments from which molecules are built (cf. Appendix C).
2. In COSMO-CAMPD, all discrete degrees of freedom correspond to the molecular structure y . Molecular structures y are optimized as in the previously proposed COSMO-CAMD framework (Chapter 4): The genetic optimization tool LEA3D (Douguet et al., 2005) creates explicit 3D molecular structures of a set of molecules $y^{(i)}$ called a generation i . In contrast to COSMO-CAMD, molecules in COSMO-CAMPD are evaluated based on the process function $F(x, y)$ (Eq. (2.2)). The molecular design space is then explored by genetic operations on the molecular structure and a next generation $i = i + 1$ of molecules is created until a specified termination criterion is reached, e.g., the number of generation $i = i_{\max}$. The evaluation of each molecule is performed in steps 2.1 and 2.2 described next:

- 2.1 The molecular structures y generated by LEA3D serve as input to property prediction with COSMO-RS ($g_1(x, y) = 0$, Eq. (2.2)). For each molecular structure y , a COSMO file is required (Section 3.2.1).

A design with COSMO-RS calculations on highest accuracy TZVPD-FINE requires large computational costs compared to the expected increase in accuracy (cf. Section 4.2.1). Thus, COSMO-CAMPD uses a computationally efficient approach based on two elements (Section 4.2.3): Two accuracy levels are employed for COSMO-RS calculations in the two phases of the

COSMO-CAMPD approach: In the *Design Phase*, molecules are designed on TZVP-MF level (further referenced as TZVP for simplicity). From the design on TZVP level a set of promising molecules is re-calculated after the optimization with highest possible accuracy using TZVPD-FINE in the *Refinement Phase* (cf. Step 4). Additionally, an automated storage system creates a COSMO database of novel designed molecules during any optimization run (Section 4.2.3). Thereby, COSMO calculations in Step 2.1 need only be performed once for a specific designed molecule which results in large time-saving and allows for efficient calculation. Based on the COSMO calculations, all thermodynamic properties for pure components and mixtures are calculated with COSMO-RS. The candidate molecules can then already be tested in a first feasibility check if criteria based on thermodynamic properties ($c(y)$ and $h(x, y)$ in Eq. (2.2)) can be formulated (e.g., the existence of a liquid-liquid phase split). These requirements are specific to the process considered and exemplified in the case study. The remaining candidate solvents are indicated by $y_{\text{TD}}^{(i)}$ in Fig. 5.2.

- 2.2 For the candidate solvents $y_{\text{TD}}^{(i)}$, the process flowsheet is optimized with respect to the continuous process variables x . Large-scale, automated evaluation of rigorous process simulations, e.g., delivered by Aspen Plus, for all designed molecules is challenging due to convergence problems and thus avoided in the *Design Phase*. For this purpose, the process is modeled using pinch-based process models (Redepenning et al., 2016; Bausa et al., 1998) ($g_2(x, y) = 0$ in Eq. (2.2)). The pinch-based process models calculate the process minimum energy demand $F(x, y_{\text{TD}}^{(i)})$ (cf. Section 3.2.2) for a specified operating point, i.e., a fixed value of continuous process variables x . The values of the continuous process variables x are then optimized by an NLP solver (such as *MATLAB fmincon* used in this work) with respect to minimum energy demand. For each candidate molecular structure in generation i , $y_{\text{TD}}^{(i)}$, the performance in terms of the minimum energy demand $F(x^*, y_{\text{TD}}^{(i)})$ and the optimal operating point x^* are returned to COSMO-CAMPD to build a new generation $i = i + 1$ of molecules.
3. When the termination criterion for molecular optimization is reached (e.g., number of iterations), an optimized solution $F(x^*, y^*)$ is obtained. This optimized solution consists of the final generation of molecular structures y^* each in its optimized process operation point denoted by optimized process variables x^* . Additionally, all molecules created during optimization are forwarded as a ranked list to the *Refinement Phase*.

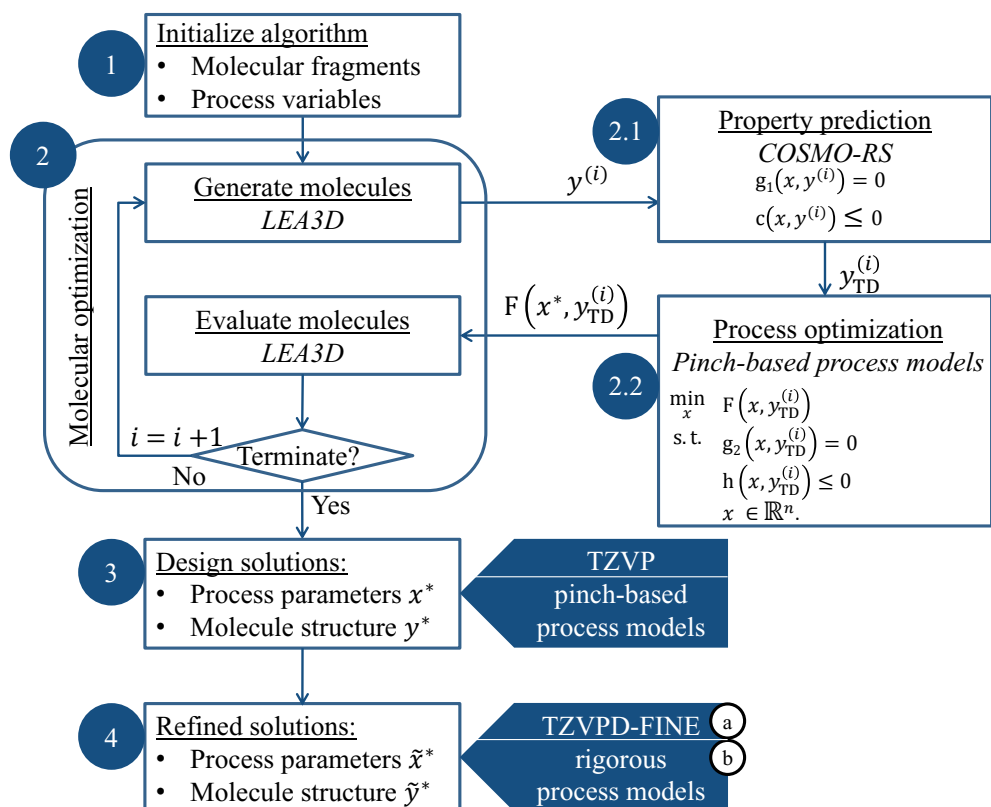


Figure 5.2: Integrated process and molecular design approach COSMO-CAMPD.

4. The *Refinement Phase* in COSMO-CAMPD has *two* aspects: In Step 4a, for the most promising molecules from the *Design Phase*, COSMO-RS thermodynamic predictions are re-calculated on the highest possible accuracy TZVPD-FINE level. With TZVPD-FINE level thermodynamic properties, the pinch-based process models are re-optimized and the optimized process parameters x^* are re-determined. In Step 4b, pinch-based process models are replaced by rigorous process models in Aspen Plus (V8.4) to determine the objective function for optimized process variables $\tilde{x}^*$. This rigorous calculation is fully automated and benefits from the excellent initialization by the pinch-based process models. The rigorous calculations return a good estimate of the minimum energy demand, if a sufficiently large finite number of stages (e.g., 50, used in this work) is used. They also provide the internal stage to stage flows which can be input to future cost optimization. The output of COSMO-CAMPD is thus a final ranked list of molecules $\tilde{y}^*$ on TZVPD-FINE level and process energy demand in Aspen Plus $F(\tilde{x}^*, \tilde{y}^*)$. The results can be easily integrated into existing process flowsheets and can be used by the design engineer for experimental validation

or for applying further criteria of practical relevance.

5.3 Case study: Integrated Solvent and Process Design for Hybrid Extraction-Distillation of γ -Valerolactone

5.3.1 Process Formulation

In Chapter 3, solvents for a hybrid extraction-distillation process (cf. Fig. 3.2) have been identified in a large database screening. Here, the solvent has two tasks: first extract GVL from the water-rich product stream and afterwards facilitate the purification of GVL by distillation. A high affinity of the solvent for GVL enhances extraction performance, whereas it might increase the energy demand for distillation. Optimal solvents will balance these two effects. The performance of the hybrid extraction-distillation process thus crucially depends on the choice of the right extraction solvent (cf. Chapter 3). In addition, extraction temperature plays a crucial role: At low temperatures, the miscibility gaps between water and organic solvents are known to be largest, i.e., the mutual solubilities are smallest. One would therefore expect to find the best performances at low temperatures. However, higher temperatures might cause steeper tie lines and thus can decrease the solvent demand. The optimal extraction temperature should balance this trade-off. In this chapter, the process proposed in Chapter 3 is extended to include extraction temperature as a degree of freedom. Here, COSMO-CAMPD is applied for integrated process and solvent design. In the following, n-butyl acetate is referred to as the literature benchmark solvent (Murat Sen et al., 2012).

The process considered in this chapter consists of a fixed flowsheet introduced in Chapter 3 with an extraction and a distillation unit (Fig. 5.3). A feed stream F with composition $\mathbf{z}_F$ containing water and GVL enters the extraction column in which a solvent stream S (consisting of solvents with molecular structure y) is used to extract GVL in the extract stream E . The extract stream E is separated in the distillation column into GVL (recovered as bottom product) and a stream of solvent and water (recovered as distillate). The distillate is separated in a decanter and the solvent-rich phase is recycled to the extraction column. The employed pinch-based process models assume an infinite number of stages which results in sharp splits in the extraction and distillation columns which are therefore imposed as purity constraints: GVL is fully recovered in the extract E and obtained as pure component product in the bottom

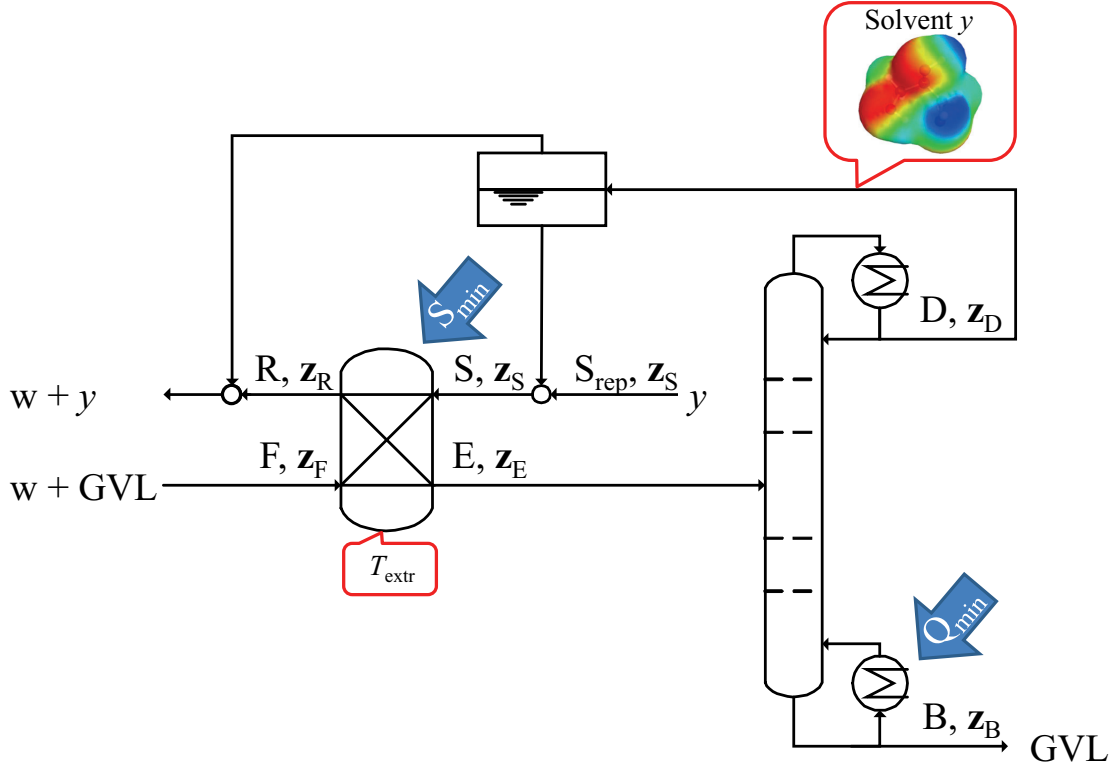


Figure 5.3: Hybrid extraction-distillation process for the purification of GVL. Streams: Feed stream F containing solute GVL and carrier water (w), solvent stream S containing solvent y , raffinate stream R , extract stream E , distillate stream D , bottom stream B , solvent make-up stream S_{rep} with corresponding compositions $z_F, z_S, z_R, z_E, z_D, z_B$. S_{min} is the minimum amount of solvent required for the specified separation task, Q_{min} is the minimum energy demand required for the specified separation task. Red boxes: Optimization variables (Eq. (2.2)) solvent y and temperature of the extraction process T_{extr} . Process specifications can be found in Appendix C.

B by distillation. The overall energy demand of the process is represented by the minimum energy demand for the distillation Q_{min} which serves as the the objective function $F(x, y)$ (Eq. (2.2)) in this chapter. The process is modeled on two different accuracy levels (cf. Section 5.2.1):

- In the *Design Phase* (Step 1 - 3), advanced pinch-based process models (Redepenning et al., 2016; Bausa et al., 1998) are used which allow for efficient calculations while still being thermodynamically accurate. Process specifications

and assumptions for the pinch-based process models are taken from Chapter 3. The pinch-based process models inherently assume an infinite number of separation stages and sharp component splits. Pinch-based process models thus give a lower bound on process streams, in particular the minimum solvent demand $S_{\min}$ and minimum energy demand $Q_{\min}$. In the *Design Phase*, the process operating point, i.e., the extraction temperature T_{extr} , is considered as a degree of freedom and optimized.

- In the *Refinement Phase* (Step 4b), the flowsheet simulator Aspen Plus (V8.4) is used to model the process with rigorous stage-to-stage calculations. The Aspen Plus models are initialized fully automated with results from the pinch-based process models. The optimized extraction temperature T_{extr} determined in the *Design Phase* by pinch-based process model optimization is fixed for the Aspen Plus calculations. For Aspen Plus simulations, a fixed number of stages (50 in extraction and distillation) is assumed to approximate the minimum solvent demand and minimum energy demand. The assumption of sharp component splits is modeled with high purity specifications for component recovery (for further process specifications, cf. Appendix C). The Aspen Plus simulations yield the process minimum solvent demand S_{rig} and minimum energy demand Q_{rig} .

5.3.2 Application of COSMO-CAMPD

The COSMO-CAMPD approach is applied to the case study to find optimal solvents and optimize the operating point of the process. The algorithm is started by setting the following process specifications (Step 1): A feed composition of $\mathbf{z}_F = (0.05, 0.95, 0.00)$ is assumed for the ternary system GVL-water-solvent. Complete recovery is required for the value component GVL in the distillation bottom B (Fig. 5.3). Pressure in the distillation is set to $p = 1$ bar whereas the temperature in the extraction column T_{extr} is considered as free process variable x (Eq. (2.2)) to be optimized in the *Design Phase*. Bounds for the extraction column temperature are set to $25^\circ\text{C} \leq T_{\text{extr}} \leq 80^\circ\text{C}$. In addition, T_{extr} is constrained by the boiling temperature of the solvent $T_{\text{boil,solvent}}$ with a safety margin of 15 K. Designed solvents are limited to molecules only containing carbon, hydrogen and oxygen atoms (Step 2). Furthermore, alkynes are excluded due to stability concerns (cf. Chapter 3). The fragment library for the optimization includes alkyl-, carbonyl-, carboxyl-, ether-, hydroxy- and cyclic groups which can be found in Appendix C. In total, three consecutive COSMO-CAMPD optimization runs are performed, each with a termination criterion of $i_{\max} = 50$ generations and 40 solvents per generation (= 2040 generated

molecular structures in total). Calculations are performed using an Intel Xeon CPU E5-1660v3@3.00GHz workstation for COSMO calculations on 8 parallel cores. For all designed solvents $y^{(i)}$, pure compound and mixture properties are predicted with COSMO-RS and tested to meet required property constraints $h(x, y)$ (Eq. (2.2)) (Step 2a):

- The number of heavy atoms in the solvents is (heuristically) constrained to a maximum of $n_{\max} = 12$ ($c(y)$, Eq. (2.2)) to avoid larger compounds with high melting points.
- The existence of a miscibility gap between water and the solvent is mandatory for a solvent to be used in extraction. Therefore, all solvent molecular structures $y^{(i)}$ are tested for the existence of a binary liquid-liquid equilibrium (LLE) with water in a preliminary feasibility test. This test is performed with the binary LLE calculation option in COSMOtherm (COSMOlogic, 2015d) at low computational expense. Solvents that do not exhibit an LLE are discarded.
- For solvents exhibiting an LLE, the pure component vapor pressure and heat of vaporization are predicted by COSMO-RS with the procedure proposed in Chapter 3. Limits on the boiling point are imposed by $T = 25^\circ\text{C} < T_{\text{boil, solvent}} < T_{\text{boil, GVL}}$ to ensure that the extraction solvent is liquid at room temperature and to obtain pure GVL in the bottom of the distillation column. Otherwise, solvents are discarded.
- For solvents with suitable boiling points, mixture properties are calculated: For liquid-phase non-ideal behavior, activity coefficients γ are used in this work, which are calculated by COSMO-RS. Parameters of the Non-Random Two Liquid (NRTL) model (Renon and Prausnitz, 1968) are regressed for each solvent $y^{(i)}$ to express the liquid-phase non-ideal behavior (cf. Appendix A). In this work, the search is confined to solvents with distillation topologies that allow for simple distillation. Thus, for the ternary system GVL-water-solvent, azeotropes are determined using the algorithm proposed by Fidkowski et al. (1993). If no ternary azeotropes and no binary azeotropes between the solvent and GVL occur, the solvent is considered to be suitable for the process.

For all suitable solvents $y_{\text{TD}}^{(i)}$, the process model is optimized with predicted thermodynamic data (Step 2b). Using the COSMO-CAMPD algorithm, solvent structures are optimized iteratively until the termination criterion ($i_{\max} = 50$ generations) is reached and a ranked list of solvents is returned (Step 3). For the top 50 most promising designed solvents, thermodynamic properties are re-calculated on TZVPD-FINE level in the *Refinement Phase* (Step 4) and automated rigorous process model calculations

in Aspen Plus are performed. As a result, a final set of designed solvents is obtained ranked by minimum process energy demand Q_{rig} on TZVPD-FINE level.

5.3.3 Results of the Design Phase

On average, 1804 solvents out of 2040 total generated solvents in each COSMO-CAMPD optimization run meet the required specification on molecular size ($n_{\text{max}} = 12$, $c(y)$, Eq. (2.2)) and are further evaluated in COSMOtherm. In an average run, 153 novel solvent are created and 1651 solvents are evaluated from the COSMO-CAMPD databank (including chemical structures suggested repeatedly by the stochastic algorithm). Further details on the optimization runs are given in Appendix C. The solution time for one COSMO-CAMPD optimization run is in the range of one day leading to an average evaluation time of 1 min per solvent and 12 min per COSMO calculation. Out of 6120 overall designed solvents in three COSMO-CAMPD optimization runs, 693 solvents are unique (i.e., not counting chemical structures suggested repeatedly by the stochastic algorithm) and fulfill the constraints $c(y)$ and $h(x, y)$ (Eq. (2.2)). Thus, the average calculation time is 7.5 min for any unique solvent. These solvents are evaluated on the process level (blue triangles in Fig. 5.4). The solvents designed by COSMO-CAMPD are compared to existing COSMObase solvents (COSMO-RS database version COSMObase-1501-BP-TZVP, COSMOlogic (2015a)) to evaluate the performance of the COSMO-CAMPD optimization. In COSMObase, 1390 solvents meet the required criteria $c(y)$ and $h(x, y)$ (Eq. (2.2)). The COSMObase solvents are shown as grey diamonds in Fig. 5.4 .

The best solvent designed by COSMO-CAMPD in the *Design Phase* is 1-(ethenyloxy)buta-1,3-diene. It reduces the minimum energy demand Q_{min} by 48 % compared to the literature benchmark n-butyl acetate. 1-(Ethenyloxy)buta-1,3-diene reduces the minimum energy demand Q_{min} by 17 % compared to the best database molecule furan. Fig. 5.4 shows the influence of minimum solvent demand S_{min} on the minimum energy demand Q_{min} . Generally, low S_{min} leads to low Q_{min} . Notably, a lower bound for the minimum energy demand Q_{min} exists which scales almost linearly with minimum solvent demand S_{min} (note: log-scale for Q_{min} in Fig. 5.4). This lower bound is formed by solvents, which are very easy to separate by distillation, e.g., solvents with low boiling point and low enthalpy of vaporization. For these solvents, very small internal reflux ratios are calculated and accordingly small minimum energy demand Q_{min} . However, even for these solvents, the minimum solvent demand S_{min} has to be evaporated which leads to the almost linear relation between S_{min} and Q_{min} . For all other solvents, no distinct correlation between minimum solvent demand S_{min} and

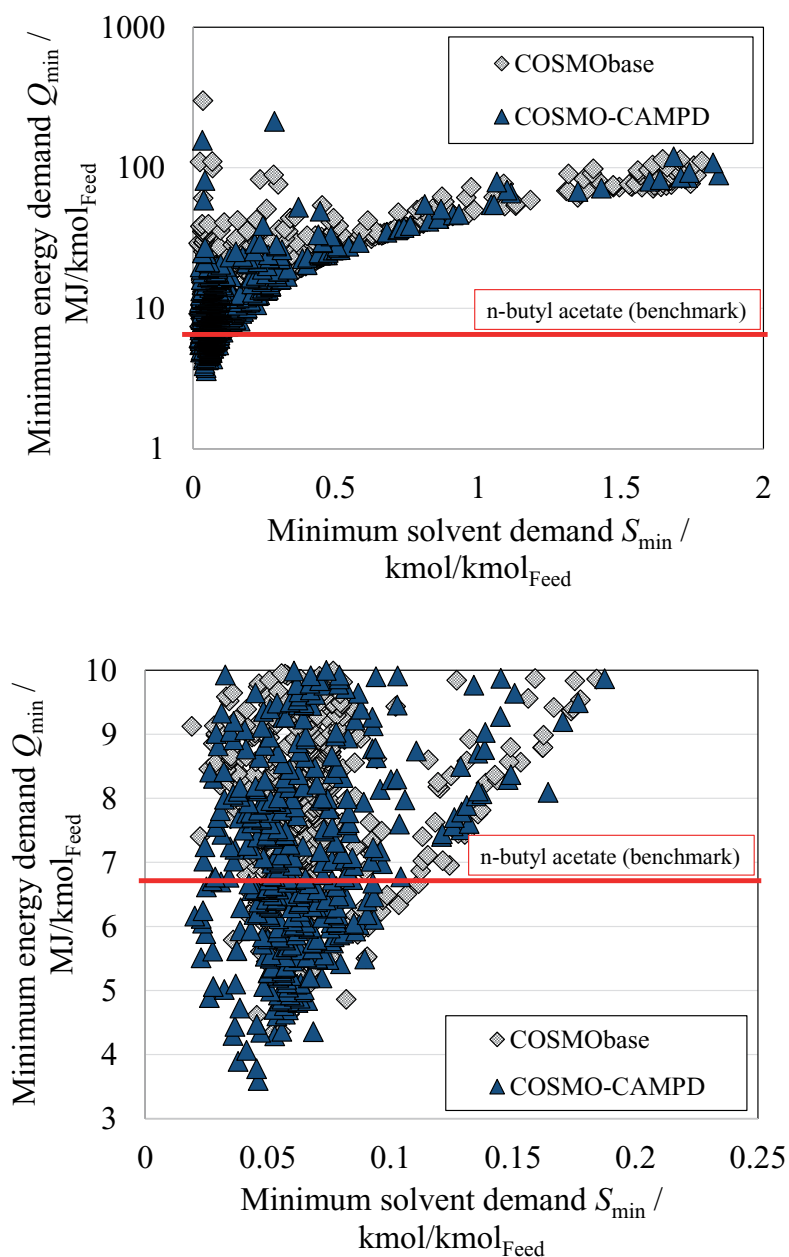


Figure 5.4: Top: Minimum energy demand $Q_{\min}$ vs. minimum solvent demand $S_{\min}$ for all solvents in the database COSMObase (grey) and COSMO-CAMPD optimization (blue). Best molecule from COSMObase: furan, literature benchmark: n-butyl acetate, best solvent from COSMO-CAMPD: 1-(ethenyloxy)buta-1,3-diene. Bottom: Zoom.

the minimum energy demand $Q_{\min}$ is observed which supports the need for an integrated molecule and process design to identify optimal solvents on a process-level. Most importantly, the optimum in minimum energy demand $Q_{\min}$ is not located at the lowest minimum solvent demand $S_{\min}$ (Fig. 5.4). Thus, a molecular solvent design based solely on extraction would not identify the optimal solvent for the overall process and an integrated process optimization is required to find the desired optimum in minimum energy demand $Q_{\min}$.

5.3.4 Results of the Refinement Phase

The *Refinement Phase* (Step 4) has two steps (5.3): first (Step 4a), thermodynamic predictions are refined on TZVPD-FINE level in COSMO-RS for the most promising solvents. Second (Step 4b), pinch-based process models are refined by rigorous process models in Aspen Plus (V8.4). As a result, a final list of ranked solvents on TZVPD-FINE level and evaluated in rigorous process models is presented.

Refinement of parametrization (Step 4a). The top 50 designed solvents are refined on the highest accuracy level (TZVPD-FINE) in COSMO-RS. In addition, all designed solvents that are already available in the COSMO-RS database COSMObase on TZVPD-FINE level are also included in the refined list of solvents to provide a larger set of promising molecules for the design engineer. In total, 213 molecules are considered in the final list of refined solvents on TZVPD-FINE level. The correlation between the different COSMO-RS accuracy levels can be measured in terms of absolute values (mean average percentage error MAPE or Pearson correlation coefficient ρ_{pearson}) or rank (Spearman correlation coefficient ρ_{spearman}). These correlation coefficients ρ determine if a correlation between two variables is linear with $\rho = +1$ for strictly positive linear relation and $\rho = 0$ for no linear correlation (cf. Chapter 3).

A comparison between TZVP vs. TZVPD-FINE shows good correlations for both rank and value of $\rho_{\text{pearson}} = 0.87$ and $\rho_{\text{spearman}} = 0.84$. Yet, the mean average percentage error (MAPE) between TZVP and TZVPD-FINE is 50 % which is significant. This rather high MAPE in this specific ternary system is caused by overestimating the GVL solubility in water on TZVP level (cf. Section 5.4). However, the high linear correlation between TZVP and TZVPD-FINE shows that this overestimation is systematic. Thus, the overestimation can be systematically corrected by higher accuracy prediction on TZVPD-FINE level. Notably, the overestimated binary system GVL-water is always present in the ternary system GVL-water-solvent and its binary NRTL parameters are unchanged by the design. The prediction of GVL solubility in water can thus be already improved in the *Design Phase*, e.g., by replacing COSMO-

RS NRTL parameters by either TZVPD-FINE calculations or experimental NRTL parameters for the binary GVL-water system (Zaitseva et al., 2016). Zaitseva et al. (2016) observed noticeable deviations of predictive methods (GC and COSMO-RS) from vapor-liquid-equilibrium measurements due to the complexity of modeling GVL. However, Zaitseva et al. (2016) found that COSMO-RS showed better performance than UNIFAC Dortmund in predicting VLE data which supports the use of COSMO-RS in the present work. In the comparison of TZVP vs. TZVPD-FINE, the correlation coefficients improve significantly using experimental binary NRTL parameters for the binary GVL-water system (further referenced to as TZVP*). Most importantly, the mean average percent error MAPE in predicting the minimum process energy demand $Q_{\min}$ is decreased by over 20 %. Thus, if experimental NRTL parameters for the solute/water system are already available, prediction quality in the *Design Phase* can be improved. For this work, TZVP* is thus considered the best possible prediction in the *Design Phase*. Yet, the comparably high correlation coefficients between TZVP and TZVPD-FINE show that experimental data are not mandatory for the proposed design method which can run fully predictive. If sufficient computing power and/or time is available, COSMO-CAMPD could also be applied only on the TZVPD-FINE level.

Refinement with rigorous process models (Step 4b). While an evaluation of all designed solvents during the COSMO-CAMPD design with rigorous process models is usually prohibitive in practice due to convergence problems, Aspen Plus calculations can be performed automatically for the final set of TZVPD-FINE refined solvents. For this purpose, rigorous process models are initialized by results from pinch-based process models. Fig. 5.5 shows the comparison between pinch-based and rigorous process models. (To allow for most general comparison, process model predictions are compared in Fig. 5.5 on TZVP level where a larger data-set is available).

The agreement between rigorous and pinch-based process models is particularly good for the most promising solvents, i.e., for solvents with a low minimum energy demand. If only promising solvents with a minimum energy demand smaller than 10 MJ kmol^{-1} feed are considered, a rank correlation or Spearman’s coefficient $\rho_{\text{spearman}} = 0.87$ can be found. Thus, the relative order of the solvents evaluated with pinch-based process models in the *Design Phase* corresponds well to the relative order of solvents evaluated with rigorous models in the *Refinement Phase*. The root-mean-square error $\text{RMSE} = 1.24 \text{ MJ kmol}^{-1}$ is also small compared to the magnitude of calculated minimum energy demands. The reason for the increasing deviations for solvents with higher minimum energy demand is the use of a finite number of stages in the rigorous model. For solvents with high minimum energy demand, the approx-

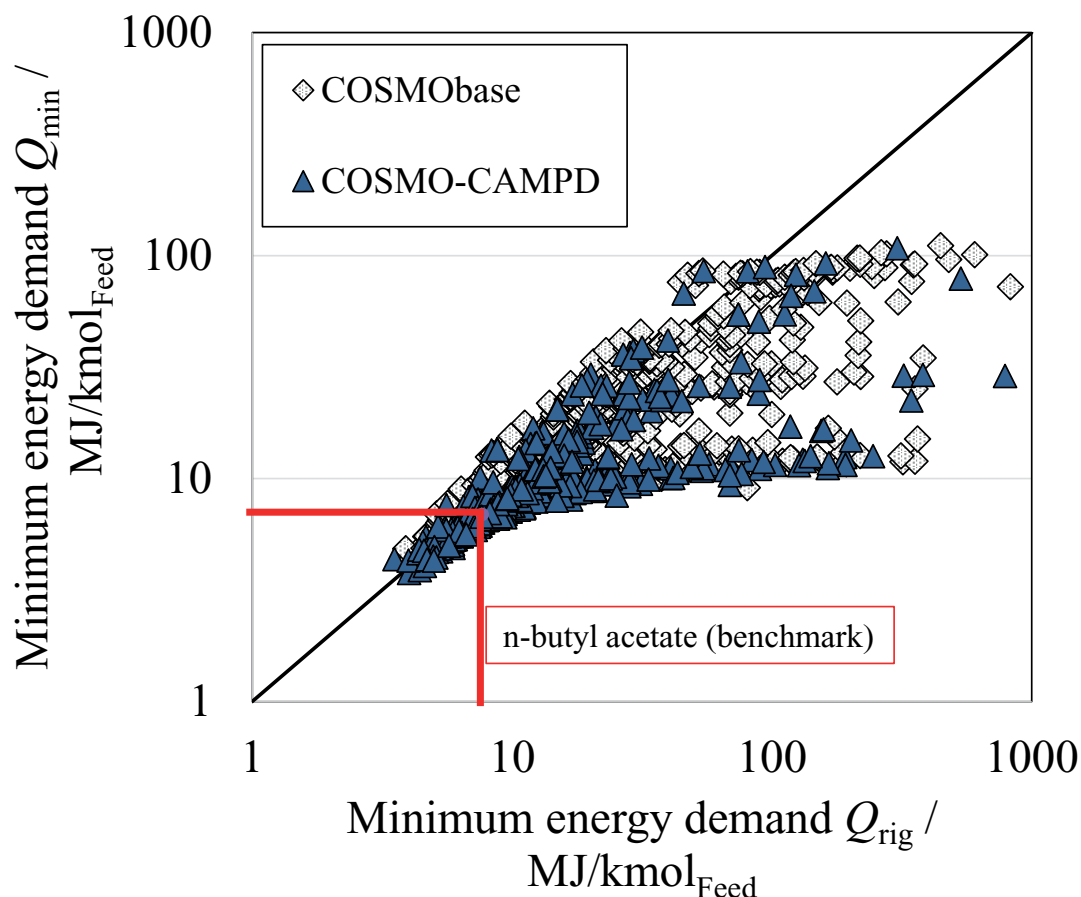


Figure 5.5: Comparison of the minimum energy demand $Q_{\min}$ calculated with the pinch-based process models to minimum energy demand Q_{rig} from rigorous simulations assuming 50 stages in extraction and distillation. Dotted diamonds: solvents from database screening (1144 solvents converged in Aspen Plus), solid triangles: solvents from COSMO-CAMPD optimization (560 solvents converged in Aspen Plus). Red line: Literature benchmark n-butyl acetate.

imation of infinitely many stages by 50 stages is not sufficient. However, since this is only relevant for solvents with poor separability and therefore poor process performance, these deviations seem to be not crucial for the solvent selection in this case study. Nevertheless, the agreement could be easily improved using more stages in the rigorous model. Due to the same reason, the pinch-based process models present a lower bound for the rigorous process simulations. Overall, pinch-based process models are considered sufficiently accurate and thus suitable for the solvent design.

A list of refined solvents designed on TZVPD-FINE level and rigorous process simulations is the final output of the COSMO-CAMPD framework. Tab. 5.1 shows the top 10 designed solvents and further selected solvents with practical relevance (commercially available, non-toxic according to manufacturer data (Sigma-Aldrich, 2017)) after the *Refinement Phase*. Solvents that are designed by the COSMO-CAMPD framework and also contained in the COSMO-RS databank are indicated by 'DB'.

Table 5.1: Ranked list of solvents designed by COSMO-CAMPD with minimum energy demand Q_{rig} on TZVPD-FINE level, optimized temperature $T_{\text{extr}}^{\text{opt}}$, rank in the design and reduction to benchmark (BM). Solvents that are also available in COSMObase are indicated by (DB).

Rank	Solvent	T_{boil} °C	$T_{\text{extr}}^{\text{opt}}$ °C	Q_{rig} MJ/kmol _{Feed}	Reduction to BM %
1	divinylether (DB)	39	25	2.66	−63
2	furan (DB)	48	25	2.99	−59
3	2-vinylfuran (DB)	104	36	3.44	−52
4	3-ethenylfuran	110	29	3.46	−52
5	3-methylfuran (DB)	72	43	3.62	−50
6	furan-3-yl formate	142	25	3.73	−48
7	furan-2-yl formate	149	25	3.75	−48
8	1-(ethenyloxy)prop-1-ene	72	47	3.75	−48
9	3-methoxyfuran	122	31	3.75	−48
10	2-methylfuran (DB)	76	47	3.85	−47
50	toluene (DB)	116	63	5.47	−24
BM	n-butyl acetate (DB)	141	53	7.21	

The best solvent designed by COSMO-CAMPD, divinylether reduces the minimum energy demand Q_{rig} by 63 % compared to n-butyl acetate from Murat Sen et al. (2012). 50 % of the solvents in the top 10 and 58 % of the top 50 are designed molecules not contained in COSMObase. Thus, COSMO-CAMPD significantly enlarges the space of promising solvents by design. In order to analyze the importance of process settings, the process calculations are repeated for a fixed extraction temperature. By using optimized temperatures in the extraction, minimum energy demand Q_{rig} is reduced by 5 % on average for the refined set of 213 solvents. The effect of temperature optimization on minimum energy demand Q_{rig} is largest for solvents with high minimum energy demand Q_{rig} . In particular, higher temperatures are beneficial for solvents of practical relevance 2-methylfuran (2MF, rank # 10 in Tab. 5.1, 3 % reduction to $T_{\text{extr}} = 25$ °C), toluene (rank #50, 12 % reduction to $T_{\text{extr}} = 25$ °C) and the benchmark n-butyl acetate (5 % reduction to $T_{\text{extr}} = 25$ °C). Notably, increased temperature has a significant impact on toluene, i.e., the ranking improves by 9 ranks with optimized

extraction temperature.

5.4 Experimental Validation of COSMO-RS Prediction

The results of COSMO-CAMPD are compared to experimental LLE measurements to validate the prediction of COSMO-RS thermodynamics. The experimental LLE measurements are performed for the most promising non-toxic and commercially available solvents 2-methylfuran (#10 Tab.5.1) and toluene (#50 Tab.5.1) as well as the benchmark solvent n-butyl acetate. The experimental LLE measurements are performed using a highly automated setup combining an autosampler and gas chromatography (Dechambre et al., 2014a). The list of chemicals and details on the used gas chromatography method are given in Appendix C. For the three experimentally validated solvents, elevated temperatures are predicted to be beneficial in the design and the *Refinement Phase* (cf. Tab.5.1). Thus, solvents are compared at ambient level ($T_{\text{extr}} = 25^\circ\text{C}$) and at elevated temperature ($T_{\text{extr}} = 40^\circ\text{C}$) to investigate the predicted beneficial effect of temperature increase on minimum solvent and energy demand. Fig. 5.6 shows the comparison between experimental data (red dashed) and COSMO-RS predictions (blue solid) for LLE of GVL-water-2MF at $T_{\text{extr}} = 25^\circ\text{C}$ for the two COSMO-RS levels TZVP and TZVPD-FINE.

Overall, the shape of the LLE and the organic-phase compositions are well predicted with COSMO-RS on both TZVP and TZVPD-FINE level. For TZVP (Fig. 5.6), the GVL concentrations in the water phase are predicted higher than experimentally determined. On the process level, this overestimation of GVL concentration in the water phase leads to an overestimation of the required minimum solvent demand S_{rig} . The reason for the overestimation of S_{rig} is the high sensitivity of the pinch-based process model to GVL concentration in the water phase due to high purity requirements by sharp splits. Thus, predicting larger solubility of GVL in the water phase leads to an increased demand of solvent to completely recover GVL from the aqueous phase. In addition, tie lines are less steep in comparison to experimental values which also contributes to the overestimating of the required minimum solvent demand S_{rig} . In contrast, for TZVPD-FINE parametrization, the prediction of GVL concentration in the aqueous phase improves with lower GVL concentrations in the water phase and steeper tie-lines are predicted (Fig. 5.6). This prediction leads to a lower solvent demands S_{rig} . The prediction quality of the GVL-water solubility can be improved by using experimental parameters (TZVP*) as discussed in Section

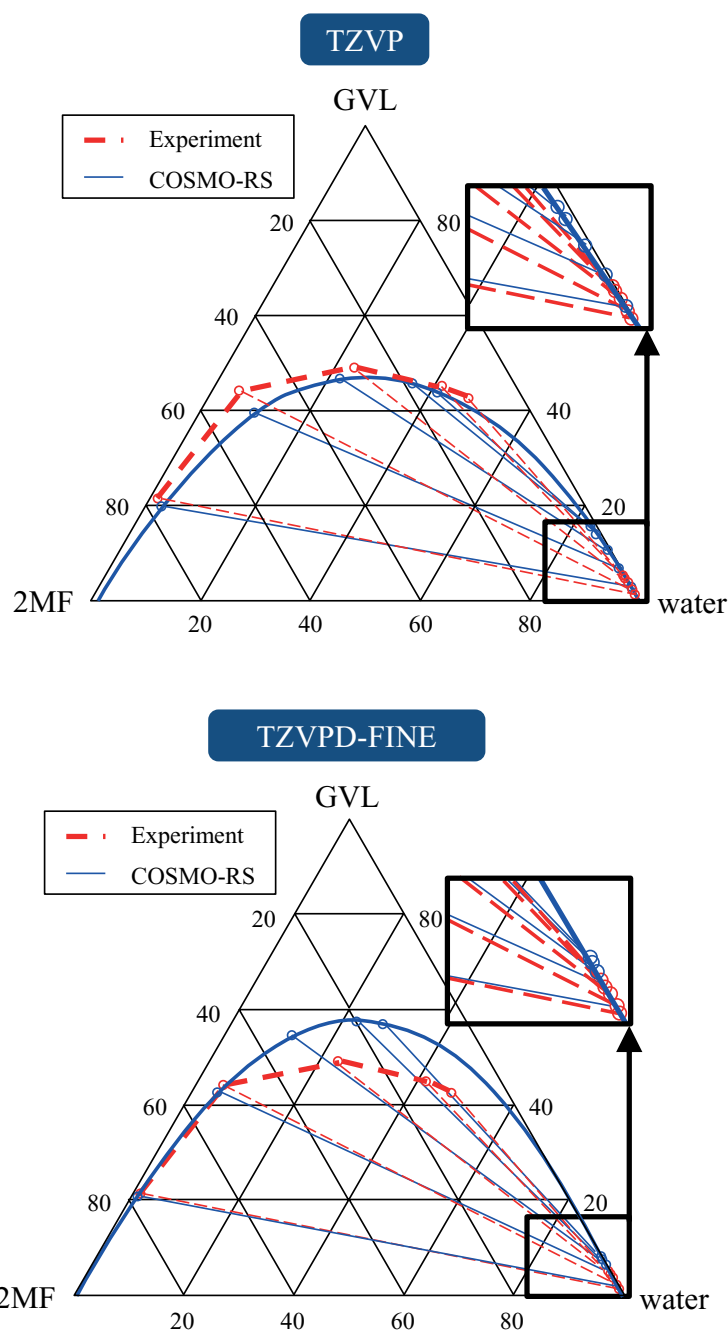


Figure 5.6: Experimental validation for designed solvent 2-methylfuran (2MF) with COSMO-RS predictions on TZVP level (solid blue) vs. experimental data (dashed red). Top: TZVP, bottom: TZVPD-FINE, both at $T = 25\text{ }^{\circ}\text{C}$.

5.3.4. Since the binary system GVL-water is fixed during COSMO-CAMPD design, it is possible to improve prediction of GVL concentrations in the water phase already in the *Design Phase* without loss of general applicability. The comparison between COSMO-RS predictions and experiments are very similar for the solvents toluene and n-butyl acetate (ternary diagrams in the Appendix C). To validate the approach, the solvents designed by COSMO-CAMPD, 2-methylfuran and toluene and are compared to n-butyl acetate on the process level. Hereby, experimentally determined isothermal NRTL parameters are input to the extraction process while temperature-dependent NRTL parameters are used on the highest accuracy COSMO-RS TZVPD-FINE level in the distillation. All solvents are evaluated in terms of minimum energy demand and compared to results from the *Refinement Phase*. Fig. 5.7 shows the result of this comparison.

A comparison of the COSMO-CAMPD predictions (dark blue back row bars in Fig. 5.7) to experimental data (red front row bars in Fig. 5.7) show very good agreement. The COSMO-CAMPD predictions generally overestimate the actual process energy requirement and thus all solvents perform slightly better than predicted. However, this overestimation is systematic. The largest deviation is found for toluene at $T_{\text{extr}} = 40^\circ\text{C}$. For this particular temperature, the minimum solvent demand for toluene is not predicted quantitatively in the *Refinement Phase*. However, COSMO-CAMPD successfully predicts the qualitative trend of benign elevated temperatures for toluene extraction. Importantly, COSMO-CAMPD accurately predicts the relative ranking of all solvents on all temperature levels. Thus, experimental data validates the predicted performance of solvents designed by COSMO-CAMPD: toluene and 2-methylfuran reduce the minimum energy demand Q_{rig} by up to 50 % in comparison to the benchmark solvent n-butyl acetate (Tab. 5.2).

Table 5.2: Minimum energy demand of experimentally validated solvents

Solvent	S_{rig} kmol/kmol _{Feed}	Q_{rig} MJ/kmol _{Feed}	T_{extr} °C	Reduction to BM %
2-methylfuran	0.04	3.21	25	50
toluene	0.05	3.57	40	44
n-butyl acetate (BM)	0.08	6.41	40	-

Toluene has recently been proposed for GVL-based co-solvent hydrolysis process (Won et al., 2017) which indicates its practical relevance for the system GVL-water. However, toluene is shown here to be outperformed by 2-methylfuran with both lower minimum solvent demand S_{rig} and lower minimum energy Q_{rig} leading to decreased process recycles and higher process efficiency. Additionally, 2-methylfuran can be di-

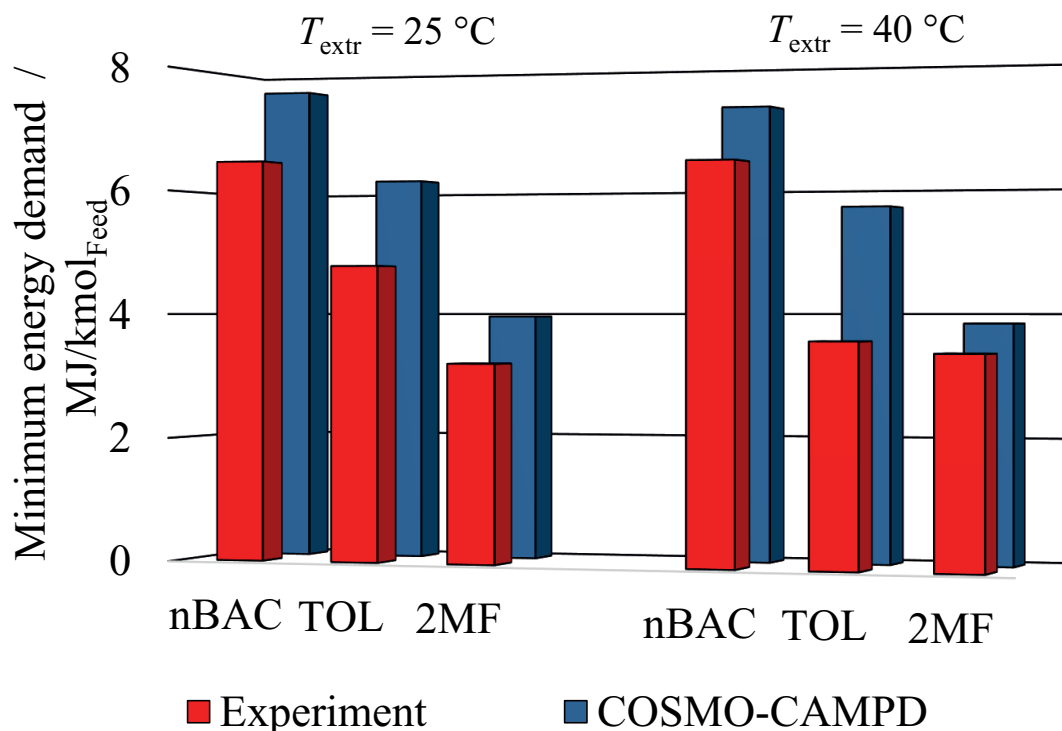


Figure 5.7: Minimum energy demand for COSMO-CAMPD prediction in the *Refinement Phase* and experimental validation for n-butyl acetate (nBAC), toluene (TOL) and 2-methylfuran (2MF). Dark blue rear row bars: Rigorous process models with thermodynamic data on TZVPD-FINE level. Red front row bars: Experimental LLE data for extraction, NRTL parameters on TZVPD-FINE level for distillation. Extraction temperature indicated by $T_{\text{extr}} = 25\text{ }^{\circ}\text{C}$ and $T_{\text{extr}} = 40\text{ }^{\circ}\text{C}$.

rectly produced from lignocellulosic biomass (Alonso et al., 2013b) and thus seems highly promising for an integrated process strategy. Therefore, 2-methylfuran is proposed as the most promising, practically relevant designed solvent by COSMO-CAMPD for the hybrid extraction-distillation of GVL. 2-Methylfuran, toluene and the other top 10 solvents from COSMO-CAMPD (Tab. 5.1) show high potential to reduce the minimum energy demand and should therefore be further investigated for future applications.

5.5 Conclusions

In this chapter, a framework for the integrated molecule and process design is presented based on COSMO-RS with no need for experimental data and no limit to molecular databanks. Computationally efficient design is achieved using two COSMO-RS accuracy levels in two phases of the COSMO-CAMPD framework: In the *Design Phase*, molecules are designed on medium accuracy level TZVP and evaluated on the process level using fast and thermodynamically accurate pinch-based process models. In the *Refinement Phase*, promising molecules are further evaluated using the highest possible QM accuracy TZVPD-FINE and rigorous process models in Aspen Plus.

The case study for hybrid extraction-distillation of γ -valerolactone shows that COSMO-CAMPD successfully designs promising solvents that are not contained in extensive databases: 58 % of the top 50 designed molecules are not contained in the COSMO-RS database. A comparison of pinch-based process models calculations to rigorous process simulation in Aspen Plus shows the high accuracy of the pinch-based process models. An inherent overestimation in COSMO-RS of the GVL solubility in aqueous phases is identified as possible improvement for the *Design Phase*. While the fully predictive approach already yields accurate prediction results and a sound ranking of the solvents, the use of experimental binary NRTL parameters for the system GVL-water leads to significant improvement of the quantitative design results.

Results of the design approach are validated by experimental liquid-liquid measurements; the promising and commercially available solvents 2-methylfuran and toluene achieve 50 % reduction in process energy demand based on experimental LLE data. Overall, COSMO-CAMPD allows for the efficient and successful design of solvents leading to better process performance.

Integrated Process and Molecule Design using COSMO-RS for Reaction-Separation Processes

6.1 Introduction

In the previous chapter (Chapter 5), COSMO-CAMPD has shown convincing results for the integrated design of solvents and separation processes. However, solvents do not only impact separation steps but instead influence other process steps such as chemical reactions. Chemical reactions are a key, e.g., for integrating fluctuating H_2 from renewable energy into the chemical value chain. For this integration, a promising concept is the conversion of CO_2 to CO via chemical storage (Behr et al., 2004). The efficiency of this storage process, however, depends strongly on the solvents employed and the process concept. To consider complex reaction steps in the design of optimal solvents and processes, in this chapter, COSMO-CAMPD is extended towards a fully integrated molecular and process design approach for reaction-separation processes (Fig. 6.1). For this purpose, process models for multiphase equilibrium reactions are integrated in the COSMO-CAMPD framework and combined with pinch-based process models to provide a sound and reasonably accurate design target. The extended COSMO-CAMPD framework for reaction-separation processes is applied to identify optimal solvents and process flowsheets for the production of CO from CO_2 using H_2 .

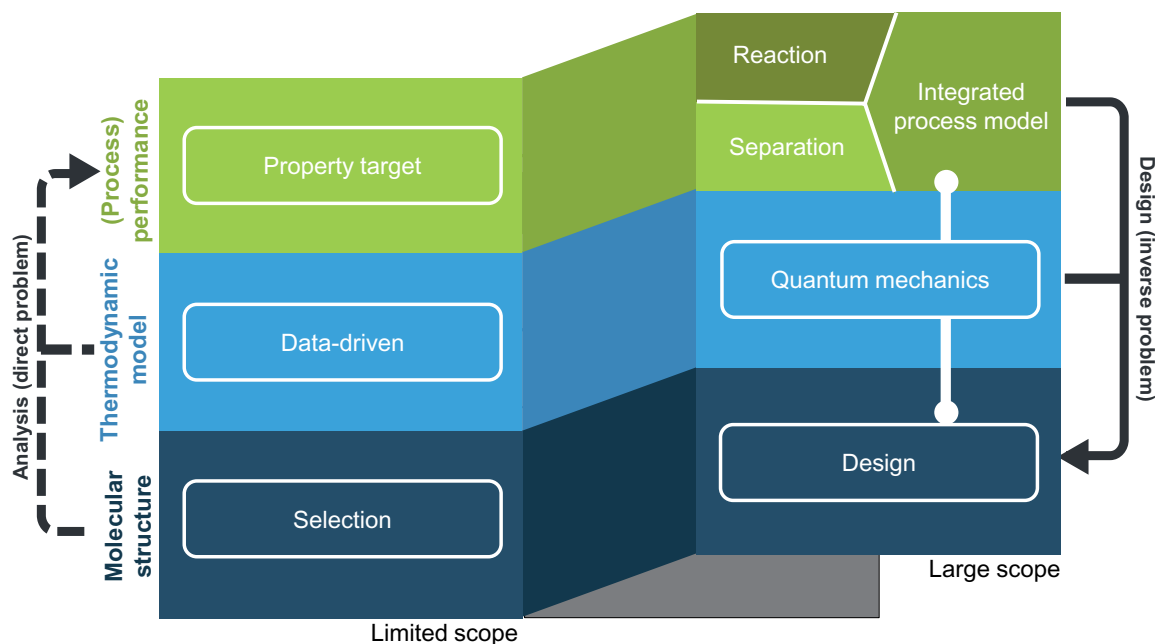


Figure 6.1: Content of this chapter in the broader CAMD/CAMPD context. Line with circles indicates the classification of the COSMO-CAMPD approach for integrated process models presented in this chapter. COSMO-CAMPD combines molecular design (large-scope molecular structure exploration method) with QM-based COSMO-RS property prediction (large-scope thermodynamic model) and pinch-based process models for reaction-separation processes (large-scope process performance model.)

Major parts of this chapter have been published in:

J. Scheffczyk, P. Schäfer, C.M. Jens, K. Leonhard and A. Bardow, 2017. Integrated process and solvent design using COSMO-RS for the production of CO from CO₂ using H₂. In *Antonio Espuña, Moisès Graells and Luis Puigjaner (Eds.), Proceedings of the 27th European Symposium on Computer Aided Process Engineering*, volume 40 of *Computer Aided Chemical Engineering*, pages 1765-1770. Elsevier, Amsterdam.

6.2 Integrated Design of Molecules, Processes and Flowsheets

6.2.1 Problem Specification

For the integrated molecule and process design problem for reaction-separation problems in this chapter, the general mixed-integer nonlinear problem MINLP (Eq. (2.2)) has to be extended to include the process flowsheet variants as degree of freedom:

$$\begin{aligned} & \underset{x, y, z}{\text{minimize}} && F(x, y, z) \\ & \text{subject to} && g_1(x, y) = 0 && (\text{thermodynamic model}), \\ & && g_2(x, y, z) = 0 && (\text{process model}), \\ & && h(x, y, z) \leq 0 && (\text{operating limits}), \\ & && c(y) \leq 0 && (\text{molecular constraints}), \\ & && x \in \mathbb{R}^n, y \in Y, z \in Z. \end{aligned} \tag{6.1}$$

Here, $F(x, y, z)$ is the objective function (e.g., process exergy demand) depending on *three* types of design variables: process variables x , the molecular structure y and the process structure z . In this chapter, the variable x is a vector containing n continuous process variables (e.g., reactor pressure). Molecules y can be either selected from a databank or generated from a defined set of molecular fragments within the design space of all possible molecules Y . To generate molecules y , the proposed COSMO-CAMPD design algorithm is employed (cf. Chapter 5). The process structure z denotes a flowsheet from a set Z of possible flowsheet variants. Equality constraints $g_1(x, y)$ and $g_2(x, y, z)$ encompass thermodynamic models and process models respectively. As a thermodynamic model, COSMO-RS (Klamt, 1995) is used. Constraints on operating limits are represented by $h(x, y, z) \leq 0$ (e.g., maximum reactor pressure or existence of a miscibility gap). Molecular constraints on the designed molecules such as size or structural feasibility are imposed by $c(y)$.

6.2.2 Integrated Process and Molecular Design for Reaction-Separation Processes

The MINLP optimization problem (Eq. (6.1)) is solved by the hybrid stochastic-deterministic algorithm COSMO-CAMPD proposed in Chapter 5. The optimization

algorithm is extended from Chapter 5 to account for process flowsheet structure optimization in reaction-separation processes (Fig. 6.2).

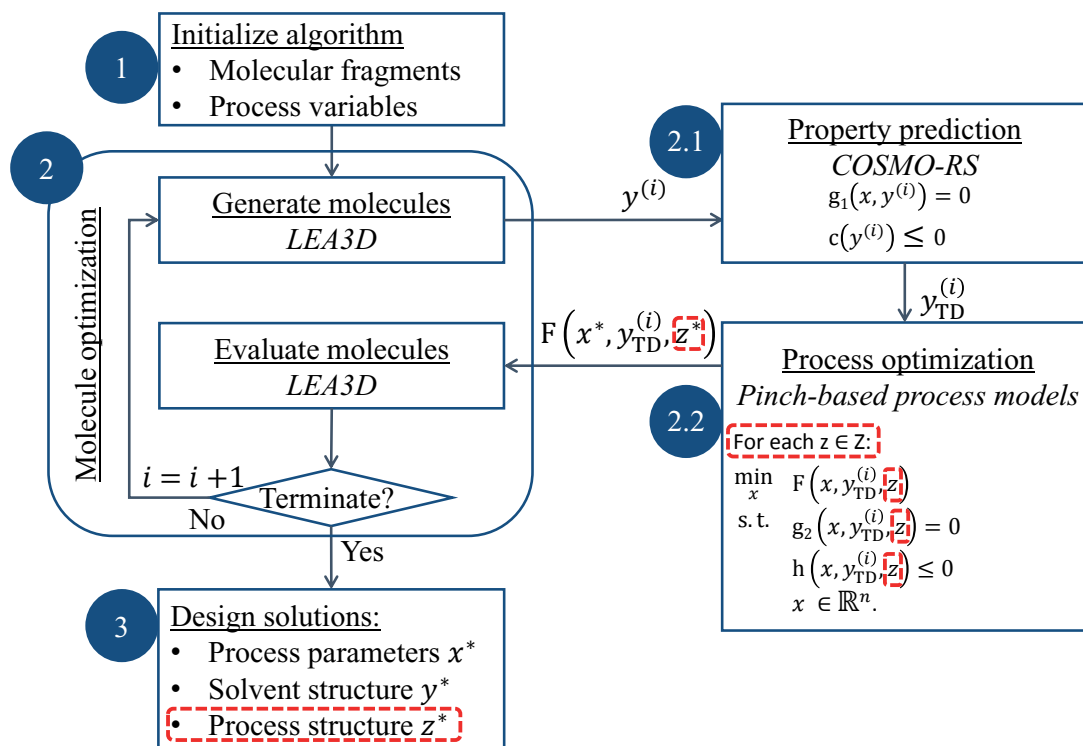


Figure 6.2: Integrated process and molecule design approach COSMO-CAMPD with extensions (dotted red boxes) from Chapter 5 to account for process flowsheet structure optimization in reaction-separation processes.

The full optimization approach for reaction-separation processes can be divided in 3 general steps:

1. The simulation is initialized (e.g., by specifying process variables and molecular fragments as degrees of freedom for the optimization).
2. Molecules y are optimized with COSMO-CAMPD according to the procedure proposed in Chapter 5. COSMO-CAMPD creates explicit molecular structures of a first generation i of molecules $y^{(i)}$. Molecules are evaluated based on the objective function $F(x, y, z)$ (Eq. (6.1)). The molecular design space is then explored by genetic operations on the molecular structure by LEA3D (Douguet et al., 2005) and a next set of molecules $y^{(i)}$ called generation $i = i + 1$ of

molecules is created until a specified termination criterion is reached, e.g., the number of generation $i = i_{\max}$.

- a) The molecular structures $y^{(i)}$ are input to property prediction with COSMO-RS (Klamt, 1995). In this chapter, all thermodynamic properties for pure components and mixtures are calculated by COSMOtherm15 (COSMOlogic, 2015d) using TZVP COSMO-RS parametrization ($y_{\text{TD}}^{(i)}$ in Fig. 6.2).
 - b) For all solvents $y_{\text{TD}}^{(i)}$ satisfying specified constraints $c(y)$ and $h(x, y, z)$ (Eq. (6.1)), each process variant z is optimized with respect to continuous process variables x by an NLP solver (here: *MATLAB fmincon*). The process flowsheet is modeled using pinch-based process models (Redepenning et al., 2016; Bausa et al., 1998) presented in Chapter 5. In addition, reactor models are implemented in this chapter for chemical reactions in vapor-liquid-liquid (VLLE) equilibrium and vapor-liquid-equilibrium (VLE) (for details see Appendix D). Overall, unit operations for single-/multiphase equilibrium reactions, distillation, heteroazeotropic distillation and extraction are evaluated.
3. When the termination criterion for the optimization is reached, the optimal solution $F(x^*, y^*, z^*)$ is obtained for the optimal molecule y^* in its optimal process variant z^* with optimized process parameters x^* . Additionally, all molecules y created during optimization are returned in all process variants z as a ranked list for possible post-processing, e.g., for refinement or experimental evaluation.

In the following, the integrated molecular and process design approach is applied to a case study of solvent and process design for the production of CO from CO₂ and H₂.

6.3 Case Study: Production of CO from CO₂ and H₂

A promising route for CO production from CO₂ and fluctuating H₂ supply is the use of formic acid derivatives (Formates and formamides) as chemical storage molecules (Jens et al., 2016). Jens et al. (2016) investigated several formic acid-based storage including dimethylformamide (DMF), diethylformamide (DEF), diisopropyl-formamide (DIPF), methylformate (MeF), ethylformate (EtF) and formic acid (FA). Among these storage molecules, DMF has been identified to be most efficient (Jens et al., 2016). The storage molecule MeF is predicted to be less efficient compared to DMF (Jens et al.,

2016). However, the additional reactant for MeF synthesis, methanol, is considered a promising solvent for CO₂ capture (Gatti et al., 2014). Thus, MeF as storage molecule could be used in an integrated CCU process which still indicates practical relevance for MeF (Jens et al., 2018). The overall CO synthesis process strongly depends on the employed solvents. Currently, solvents for the process are selected based on database screenings and selection heuristics (Jens et al., 2016). This heuristic procedure limits the design space and might results in suboptimal choices. In this chapter, systematic process and solvent design is thus performed for the two promising storage molecules MeF and DMF.

6.3.1 Process Specification

The process model is based on the process model developed by Jens et al. (2016) and contains three fundamental steps (cf. Fig. 6.3):

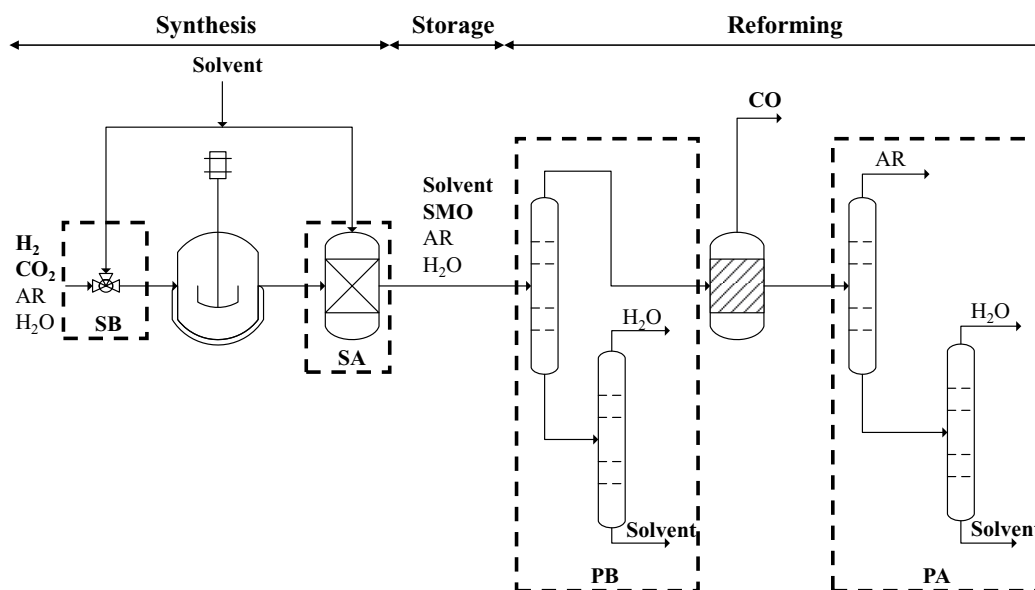


Figure 6.3: Process variants for the conversion of CO₂ to CO. The structural decisions (dashed lines) are 'solvent before (SB)/after (SA) synthesis' and 'purification before (PB)/after (PA) reforming'. Abbreviations: storage molecule (SMO), additional reactant (AR).

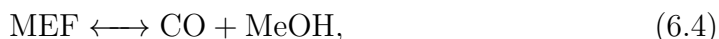
- I Homogeneously catalyzed **liquid phase synthesis** of the storage molecule SMO from CO₂ and H₂. The SMO is determined by the selection of an additional reactant (AR): Methanol (MeOH) leads to the formation of the MEF (Eq. (6.2)),

whereas Dimethylamine (DMA) leads to DMF (Eq. (6.3)). In both reactions, water is formed as by-product.



II **Storage** of the storage molecule MEF or DMF

III Continuous **reforming** of storage molecule SMO releasing CO and additional reactant AR for recycling to the synthesis reactor (Eqs. 6.4 and 6.5).



The integrated design of solvents and process in this chapter considers all reaction steps as well as the separation sequences for both sub-processes: the SMO synthesis (Step I) and the reforming to CO (Step III). Four structural design decisions are explicitly considered for the flowsheet (cf. Fig. 6.3):

In the synthesis step, a solvent is used to separate the storage molecule SMO from the aqueous phase which contains the homogeneous catalyst and can be recycled after the separation. Solvent can be added to the process either before (SB) or after (SA) the synthesis reactor (cf. Fig. 6.3). The solvent has several tasks, e.g., shifting the reaction equilibrium, extracting the storage molecule SMO after synthesis and enhance the purification. Obviously, there are trade-offs between these tasks: A high affinity between the solvent and the storage molecules SMO is generally preferred for reaction and extraction, but complicates the separation. The balance between these trade-offs differs in the individual processes flowsheets. Major advantage of SB process flowsheets is that the solvent can shift the reaction equilibrium to higher yields whereas for SA process flowsheets, the extraction performance is enhanced by multi-stage extraction.

In the reforming step, the purification strategy has a strong impact on the overall process. Here, the storage molecule SMO and additional reactant AR can either be purified before (PB) or after (PA) being fed to the reforming reactor (cf. Fig. 6.3). PA and PB process flowsheets differ in the main separation step: In PB flowsheets, storage molecule SMO and additional reactant AR are recovered from water and solvent. In PA flowsheets, only additional reactant AR needs to be recovered from water and solvent. Depending on the intermolecular interactions and resulting (non-

ideal) thermodynamics, one flowsheet variant might be favored over the other. E.g., for PB flowsheets, on the one hand, four components are present in the separation (cf. Fig. 6.3) which increases the possibilities for complex mixture topologies such as azeotropes that complicate the separation. On the other hand, for PB flowsheets, a smaller stream has to be vaporized for reforming leading to lower exergy demand in this process step. These complex trade-offs highlight the need for an integrated design approach requiring sound non-ideal thermodynamics and process-level assessment.

In total, there are four combinations of structural decisions, leading to four explicit flowsheets process structures z (Eq. (6.1), $Z = \{SB|PB, SA|PB, SB|PA, SA|PA\}$). In addition to structural decisions, two continuous process variables x (Eq. (6.1)) are considered in the synthesis reactor: The reactor pressure x_1 and the solvent to water ratio x_2 . The objective function $F(x, y, z)$ for the optimization problem (Eq. (6.1)) is the energy demand per mol CO measured in exergy (Jens et al., 2016). The exergy demand contains here the work for compression of CO_2 and H_2 to synthesis reaction pressure (E_{comp}), the minimum exergy demand for distillation columns ($E_{\text{boil}} + E_{\text{cond}}$) in the separation steps and the exergy for vaporizing the feed of the reforming reactor (E_{vap}):

$$F(x, y, z) = E_{\text{comp}} + E_{\text{vap}} + E_{\text{boil}} + E_{\text{cond}}. \quad (6.6)$$

Notably, the contribution E_{cond} appears with a negative sign in the minimum exergy demand Eq. (6.6), if cooling occurs above the reference temperature for exergy calculation (cf. Appendix D). I.e., E_{boil} and E_{cond} have opposite signs, if both heating and cooling occur above reference temperature. If heating above reference temperature and cooling below reference temperature is required, both E_{boil} and E_{cond} contribute to minimum exergy demand with same signs.

6.4 Reaction-Separation Process Models

The process models employed in this chapter depend on the flowsheets process structures $Z = \{SB|PB, SA|PB, SB|PA, SA|PA\}$. A general overview of the employed models is given in this section. Detailed process specifications and explicit flowsheets can be found in Appendix D.

6.4.1 Reactor Models

In the **synthesis step**, homogeneously catalyzed liquid-phase synthesis of the storage molecule SMO is performed (Jens et al., 2016). The synthesis reactor is modeled as equilibrium reactor and operates at 25 °C. The reactor pressure x_1 is optimized with respect to process exergy demand (Eq. (6.6)). Gibbs free enthalpies of reaction $\Delta G_{\text{ig,R}}$ are taken from Jens et al. (2016). Reactants H_2 and CO_2 are assumed to be completely removed from the reactor outlet by flashing to ambient pressure.

Retention of the catalyst and extraction of the storage molecule SMO is achieved by addition of a solvent depending on the flowsheet process structure (cf. Fig. 6.3). Two possible reactor types are considered: In SA process flowsheets, CO_2 , H_2 , additional reactant AR and water are fed to the synthesis reactor (Fig. 6.3), thus a vapor-liquid equilibrium (VLE) synthesis reactor is employed. After the synthesis, storage molecule SMO is extracted using an extraction column to recover the SMO and retain the catalyst in the aqueous phase. For SB process flowsheets, the solvent is added before the reaction leading to a vapor-liquid-liquid equilibrium (VLLE) with in situ extraction of the storage molecule SMO in the extraction phase and catalyst retention in the aqueous phase.

In this chapter, a multiphase reactor model is implemented that simultaneously solves the reaction and phase equilibria. All details for the reactor model are given in Appendix D. The quality of the reactor model was validated for the VLLE reactor in a comparison against Aspen Plus simulation using the built-in RCSTR equilibrium reactor model. For a validation set of 4731 reaction solvents from COSMO-RS database (converged solvents in database (COSMOlogic, 2015a) screening for MeF: 1962 and DMF: 2769), excellent agreement is achieved: A comparison of the reactor output concentrations of storage molecule SMO between Aspen Plus RCSTR and the multiphase reactor leads to a coefficient of determination of $R^2 = 0.975$. This quality of the multiphase reactor model is thus considered sufficiently accurate.

For the **reforming step**, full conversion of the storage molecule SMO is assumed at a sufficiently high temperature (Supronowicz et al., 2015). Selectivity is assumed to be 100 %. All CO is released after the reaction by cooling the reactor outlet.

6.4.2 Separation Models

As in previous chapters, purification by distillation or extraction unit operations are considered in this chapter. I.e., for the design of the sequence of separation units, pinch-based process models are employed for extraction (Redepenning et al., 2016)

and distillation (Bausa et al., 1998). The pinch-based process models assume an infinite amount of separation stages and sharp splits for the components (cf. Chapter 3). In the pinch-based process models for extraction and distillation, the minimum solvent demand and minimum energy demand are determined respectively. Based on minimum energy demand, the exergy demand for separation in this process (Eq. (6.6)) is determined by considering heating (cooling) demand of the reboiler (condenser) in the distillation columns and the respective temperatures. Other heating or cooling devices (e.g., preheating of the feed streams to boiling temperature) are assumed to be negligible. Calculation details can be found in Appendix D.

For **extraction columns** (SA process flowsheets), pure solvent is used for the extraction and the storage molecule is fully recovered from the reactor outlet. A liquid-liquid-equilibrium (LLE) with water is required for all solvents. Thus, all solvents in the integrated process and solvent design are tested for the existence of a LLE with water; furthermore, only extraction solvents are considered which are liquid at 25 °C (cf. Chapter 5).

The structure of the **distillation sequence** depends on the chosen process variant (PB or PA) and the topology of the mixture. The mixture topology (boiling points, regions of immiscibility, azeotropes) is determined fully automated using an algorithm by Fidkowski et al. (1993). Based on the mixture topology, structural decision rules are applied to specify the separation sequence and automatically applied by the design algorithm:

- For distillation, simple distillation columns are considered. The columns are modeled using the rectification body method (RBM, Bausa et al. (1998)). Sharp splits are carried out in all columns, drawing off pure components either at the top or at the bottom of distillation columns.
- Solvents forming homogeneous azeotropes cannot be recovered by simple distillation and are thus discarded. For solvents forming a heteroazeotrope with water, the solvent/water separation can be achieved in heteroazeotropic distillation. Pinch-based process models for heteroazeotropic distillation exist (Krämer et al. (2011a)) but are currently not available for large-scale automated solvent evaluation. Heteroazeotropic distillation in this chapter is thus approximated by combining a phase separator and two distillation columns. This assumption can lead to possible deviations of phase compositions (Krämer et al., 2011b) and should thus be validated, e.g., using rigorous simulations in consecutive validation steps (Skiborowski et al., 2015b).
- For PB variants, separation of additional reactant AR from storage molecule

SMO is performed (if possible) by simple distillation. If no heteroazeotrope between water and the solvent exists, separation of water and solvent is performed in two consecutive columns. If a heteroazeotrope between water and the solvent exists, the variant is considered feasible if water and solvent both boil lighter or heavier than the storage molecule SMO and the additional reactant AR. In this case, SMO and AR are removed in two consecutive columns before heteroazeotropic separation of water and solvent.

- For PA variants, if no heteroazeotrope between water and the solvent exists, water and solvent are removed subsequently before the reforming step. If a heteroazeotrope between water and the solvent exists, the additional reactant AR is removed before the heteroazeotropic separation of water and the solvent is performed.

The overall excellent agreement between pinch-based process models for extraction and the distillation and their respective rigorous model has been presented in previous chapters (Chapter 3, Chapter 5). Thus, the process models are considered sufficiently accurate in comparison to rigorous process models and are employed for COSMO-CAMPD.

6.4.3 Application of COSMO-CAMPD

For the integrated process and solvent design, COSMO-CAMPD is applied to identify the optimal solvent y^* with optimal process conditions x^* in the best possible process topology z^* (Eq. (6.1)). The procedure is started by initializing the process and solvent design setting the following specifications (Step 1): The storage molecule SMO for an optimization run is defined as well as the corresponding the additional reactant AR. Components present in the optimization are thus CO_2 , H_2 , AR, SMO, water and solvent. Temperature in the reactor is set to $T = 25^\circ\text{C}$ while pressure in the reactor is considered a continuous optimization variable x_1 (Eq. (6.1)). The pressure is initialized depending on the SMO (for DMF: $p = 1$ bar, for MeF $p = 100$ bar) and bounds are set accordingly (cf. Appendix D). The composition of the reactor feed consists of stoichiometric H_2 , CO_2 , and AR. For SB variants, solvent is added to the reactor while solvent to water ratio is considered a process variable x_2 . Pressure in the distillation columns is set to $p = 1$ bar. Temperature in the extraction is set to 25°C (For SA variants). Possible molecular fragments y (Eq. (6.1)) for solvents are set to contain methyl-, ethyl-, propyl-, butyl-, phenyl-, carbonyl-, carboxyl-, ether-, hydroxyl- and cyclohexyl-group fragments as well as alkynes and nitrogen or halogen containing molecules. A comprehensive list of employed molecular fragments y is given

in Appendix D.

Three COSMO-CAMPD optimization runs are performed for each storage molecule SMO, each with a termination criterion of $i_{\max} = 50$ generations and 40 solvents per generation (2040 molecules in total per generation). The optimization runs are performed on an Intel xeonCPUE5-1660v3@3.00GHz workstation using 8 parallel cores.

For the designed solvents, thermodynamic properties are calculated by COSMO-RS (Step 2.1). Pure component properties calculated are vapor pressure and enthalpy of vaporization. Mixture component properties are non-ideal vapor-liquid phase and liquid-liquid phase data. For all mixture data, temperature-dependent Non-Random Two Liquid (NRTL, Renon and Prausnitz (1968)) parameters are predicted for components present in the process with the procedure proposed in Chapter 3. All designed solvents are required to meet certain constraints ($h(x, y, z)$ and $c(y)$, Eq. (6.1)) and are discarded from the optimization else. The constraints are:

- Number of heavy atoms is constrained ($n_{\max} = 12$) and a miscibility gap with water and solvents is mandatory for either VLLE (SB variants) or extraction (SA variants). Thus, solvents are tested in an automated procedure for the existence of a binary liquid-liquid equilibrium (LLE) with water by the fast binary LLE calculation option in COSMOtherm (COSMOlogic, 2015d).
- Pure component properties are predicted and checked for limits on boiling points ($T = 25^{\circ}\text{C} < T_{\text{boil, solvents}}$) to guarantee a liquid-phase in the VLLE reactor (variant SB) as well as in the extraction column (variant SA).
- For solvents that meet the previous constraints, mixture properties are predicted with COSMO-RS with the procedure proposed in Chapter 3. Mixture topology for the multicomponent system CO_2 , H_2 , AR, SMO, water and the solvent is identified by determining azeotropes and boiling points with the procedure proposed by Fidkowski et al. (1993). Solvents with mixture topologies according to specifications in Section 6.4.2 are considered suitable solvents $y_{\text{TD}}^{(i)}$ for the process.

For suitable solvents $y_{\text{TD}}^{(i)}$, process models are optimized for each possible process variant z (Step 2.2). Based on the topology of the flowsheet variant Z , flowsheets are automatically assembled according to specifications (Section 6.4.2). For each process model, the process parameters x are optimized with respect to exergy demand (Eq. (6.1), Step 3). The solvent structure is optimized iteratively with COSMO-CAMPD, until a maximum iteration ($i_{\max} = 50$ generations) is reached. As a result, the optimal solvent structure y^* with optimized process settings x^* and z^* is returned. Also, a comprehensive list is returned containing all solvents created dur-

ing the COSMO-CAMPD optimization that can be further processed by the design engineer.

6.5 Results for Integrated Process and Solvent Design

To benchmark the integrated process and solvent design approach, first, a database screening is performed for all solvents in the COSMO-RS database (COSMOlogic, 2015a). Thus, solvents y (Eq. (6.1)) are explicitly enumerated for all solvents (10,000) in the COSMO-RS database and assessed by process exergy demand (Eq. (6.6)). Two storage molecules are considered which leads to over 80,000 overall process and solvent combinations. Fig. 6.4 shows the results of the COSMO-RS database screening where each data point corresponds to a solvent with its optimal process flow sheet z^* and optimized parameters x^* . For MEF, 2,958 solvents lead to feasible processes and 2,403 for DMF. The solvent with the lowest exergy demand identified in the screening is 3-methyl-1,2-butadiene ($F(x^*, y, z^*) = 24.8 \text{ kJ mol}^{-1}_{\text{CO}}$) for DMF as a storage molecule and employs the SA|PA process variant. The solvent screening reveals process-inherent trade-offs: For the storage molecule MEF (Fig. 6.4A), a general trend exists towards low exergy demand for high MEF concentrations after synthesis indicating a strong influence of the reaction on process exergy demand. This strong dependency is due to high optimal reactor pressures for MEF synthesis ($x_1 \sim 100 \text{ bar}$) which compensate low equilibrium concentrations of MEF under ambient pressure. Solvents shifting the reaction equilibrium towards high MEF concentrations after synthesis thus allow for lower reactor pressures and decreased exergy demand for compression. In contrast, high DMF concentrations after synthesis are already obtained under low reactor pressures ($x_1 \sim 1 \text{ bar}$). Thus, for DMF no distinct trend exists between exergy demand and concentration after synthesis (Fig. 6.4B), which further emphasizes the need for an integrated process-based solvent evaluation.

To design the optimal solvent for the process, the integrated process and solvent design approach based on COSMO-CAMPD (cf. Fig. 6.2) is applied for both storage molecules MEF and DMF. Since the approach employs a stochastic algorithm, 3 optimization runs are performed for each storage molecule. In each run, 2040 solvents are generated in 51 generations with a computational demand of 2 days on 8 parallel cores. For both storage molecules, MEF and DMF, COSMO-CAMPD designs molecules that reduce the exergy demand compared to the best solvents from screening. In case of MEF, the improvement is comparably small (4.6 %, cross in Fig. 6.3A).

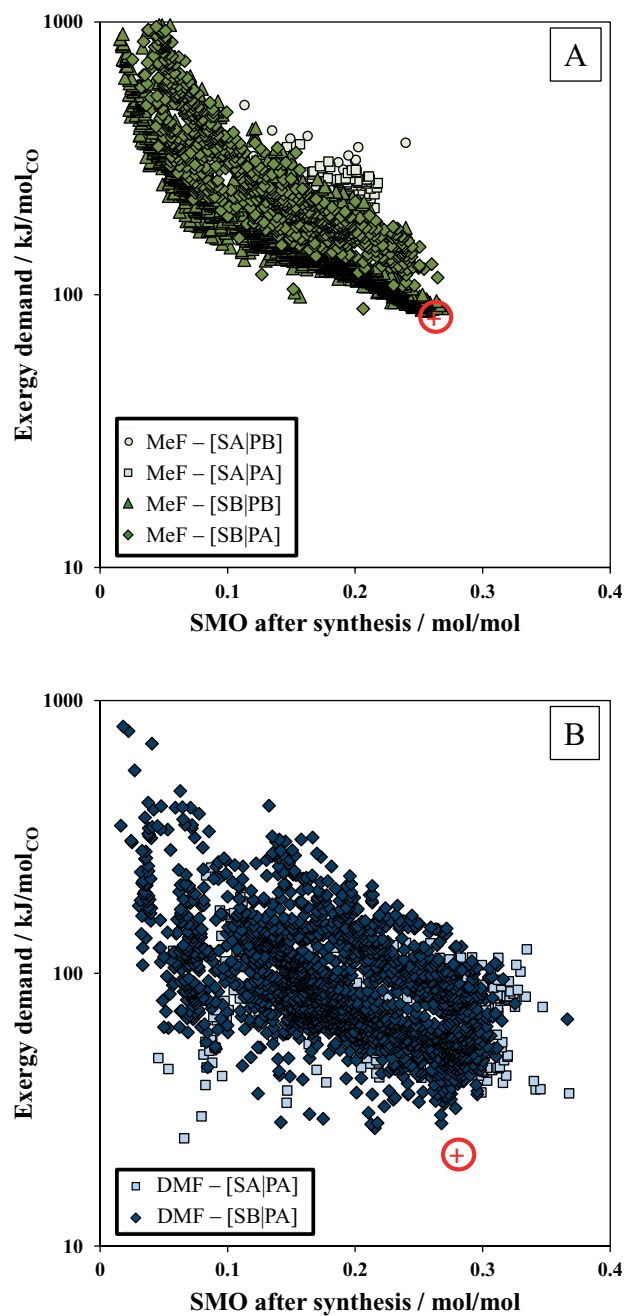


Figure 6.4: Exergy demand in process vs. storage molecule (SMO) concentration after synthesis reactor. Fig. 6.4A: MEF as SMO, Fig. 6.4B: DMF as SMO for all feasible solvents. The shape of symbols indicates for each solvent the optimal process flowsheet variant (SB|PB, SA|PB, SB|PA, SA|PA). The cross marks the best solvent designed by COSMO-CAMPD.

However, for DMF, the already lower exergy demand per mol CO is further reduced by 12.1 % (cross in Fig. 6.4). The overall best solvent for the process designed by COSMO-CAMPD is penta-1,2,4-triene for DMF as storage molecule in SB|PA process variant ($F(x^*, y^*, z^*) = 21.8 \text{ kJ mol}^{-1}_{\text{CO}}$). Comparing the best solvent designed by COSMO-CAMPD to the best solvent identified by the database screening highlights the need for an integrated design of processes and solvents: Although both solvents have similar molecular structures (acyclic hydrocarbons with 5 C atoms and only single and double bonds), the optimal process flowsheet differs (SB|PA vs. SA|PA). The molecule designed by COSMO-CAMPD shows a high affinity for DMF (SMO concentration after synthesis: $0.28 \text{ mol mol}^{-1}$) which favors the addition of solvent before the synthesis reactor (SB|PA variant). However, the SMO concentration after synthesis is not a sufficient indicator for the overall process performance due to complex trade-offs in molecular properties. The effect of these trade-offs can be visualized by a breakdown of contributions to the total process exergy demand (Eq. (6.6)) for a set of top solvents (Fig. 6.5).

The solvents in Fig. 6.5 are the solvent leading to the highest reaction conversion (1h,8h-hexadecafluorooctane, hfOCT), the best solvent identified in the screening (COSMObase) and the best solvent designed by COSMO-CAMPD. For each solvent, contributions to total process exergy are normalized to the overall process exergy demand $F(x, y, z)$ (Eq. (6.6)) for the respective solvent. The high equilibrium reaction conversion of hfOCT ($\xi_{\text{hfOCT}} = 0.93$, light blue in Fig. 6.5) indicates a high affinity of the solvent for the SMO. This high affinity is reflected in the favored process flowsheet variant (SB|PA) and high concentrations of the storage molecule SMO after synthesis ($x_{\text{SMO}} = 0.28 \text{ mol mol}^{-1}$). This high affinity comes at the price of a difficult separation with large exergy demand contributions of E_{vap} and E_{boil} (Fig. 6.5). In contrast to hfOCT, the best solvent identified in COSMObase (red in Fig. 6.5) shows low exergy for separation (E_{vap} and E_{boil}). However, the poor extraction quality of the solvent ($x_{\text{SMO}} = 0.07 \text{ mol mol}^{-1}$) leads to large solvent streams and increased cooling demand E_{cond} . The solvent designed by COSMO-CAMPD balances these effects: The solvent has a medium reaction conversion ($\xi_{\text{COSMO-CAMPD}} = 0.81$) which is reflected in the comparably high exergy demand for compression E_{comp} . However, the solvent allows for effective extraction ($x_{\text{SMO}} = 0.28 \text{ mol mol}^{-1}$) in a single stage (SB|PA process flowsheet) and low-exergy distillation. Thus, the solvent designed by COSMO-CAMPD efficiently performs reaction, extraction and separation leading to balanced exergy contributions and the lowest total process exergy demand. The complex process trade-offs further highlight the importance of molecular design based on process-level information in COSMO-CAMPD.

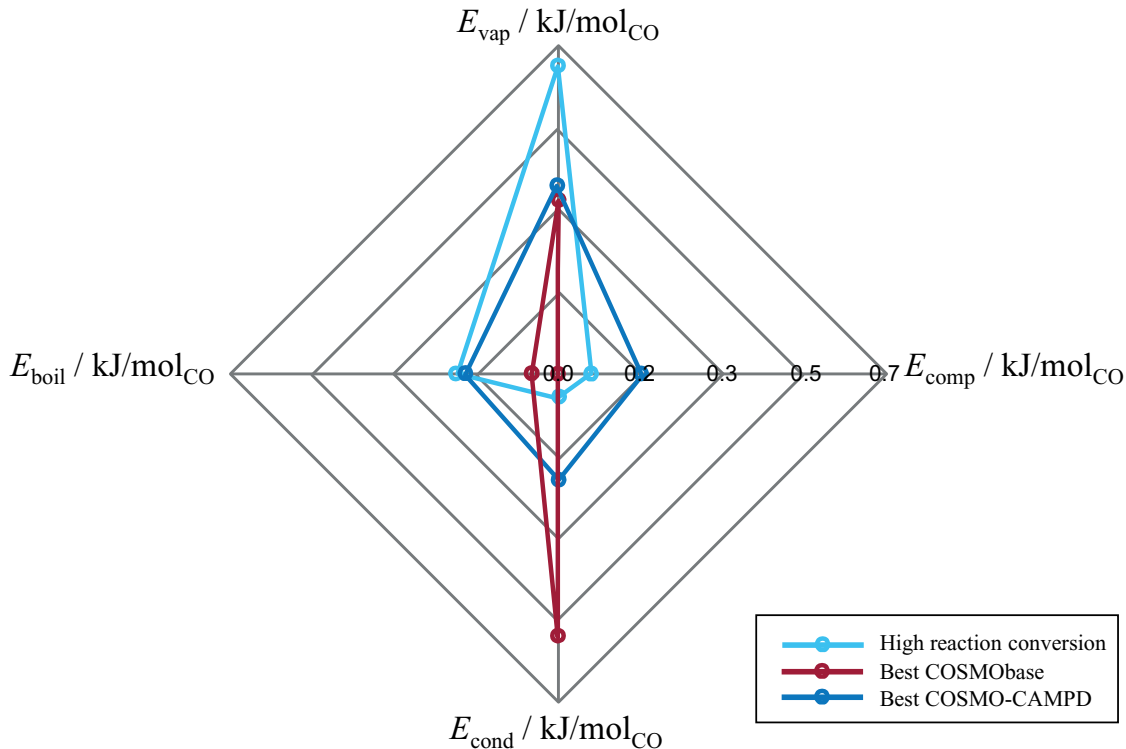


Figure 6.5: Breakdown of contributions to process exergy demand (Eq. (6.6)) for three solvents using DMF as storage molecule. Solvents are 1h,8h-hexadecafluorooctane (highest reaction conversion, light blue, SB|PA process flowsheet), best solvent identified in COSMObase (red, SA|PA process flowsheet) and best solvent designed by COSMO-CAMPD (dark blue, SB|PA process flowsheet). Contributions to process exergy demand are work for compression (E_{comp}), exergy demand for distillation columns ($E_{\text{boil}} + E_{\text{cond}}$) and exergy for vaporizing the feed to reforming reaction (E_{vap}). For each solvent, contributions to process exergy demand are normalized to the overall process exergy demand $F(x, y, z)$ (Eq. (6.6)) for the respective solvent.

6.6 Conclusions

In this chapter, an integrated molecular and process design approach is presented based on COSMO-RS for reaction-separation process. Molecules are evaluated by deterministic process flowsheet optimization using pinch-based process models. The process models consider single-/multiphase equilibrium reactions, distillation, heteroazeotropic distillation and extraction and thus allow to capture process-inherent trade-offs. The optimization approach is applied for solvent design in the production of CO from CO₂ and H₂ based on a sound process-level design target. The optimization considers different chemical storage molecules, process flowsheet variants and continuous process variables. The designed solvents improve the process performance by more than 12% compared to a massive database screening of over 80,000 combinations of solvents and structural process variants. Thus, COSMO-RS is efficiently integrated in molecular and process design considering separation, flowsheet selection and complex reaction process steps.

Summary, Conclusions and Future Perspective

7.1 Summary and Conclusions

In this thesis, quantum mechanics (QM)-based property prediction by COSMO-RS is systematically integrated into computer-aided molecular and process design. The integration is established stepwise by extending current CAMD/CAMPD approaches with limited-scope key elements towards large-scope CAMD/CAMPD approaches (Fig. 7.1):

The first step towards integration of COSMO-RS into CAMPD is presented in Chapter 3. In this chapter, a method is presented for efficient process-level evaluation of solvents using COSMO-RS property prediction and pinch-based process models for separation unit operations. The approach is not limited to a reduced number of components which is highlighted for a fully automated solvent screening of a large-scale database for the hybrid extraction-distillation of the platform chemical γ -valerolactone. Results show that more than 4600 solvents can be screened fast and efficiently with the presented screening approach. Novel promising solvents are identified with predicted better performance than the literature benchmark *n*-butyl acetate. Overall, the proposed solvent screening approach efficiently combines COSMO-RS property prediction with a comprehensive process-level assessment.

The second step towards COSMO-RS-based CAMPD is the integration of quantum mechanical information into CAMD in Chapter 4. For this purpose, COSMO-CAMD is presented as a framework for molecular design based on COSMO-RS. In COSMO-CAMD, optimization-based molecule design is achieved with the genetic algorithm LEA3D, which creates 3D molecular structure information as input for COSMO-RS. A hierarchical approach is developed employing two accuracy levels for quantum

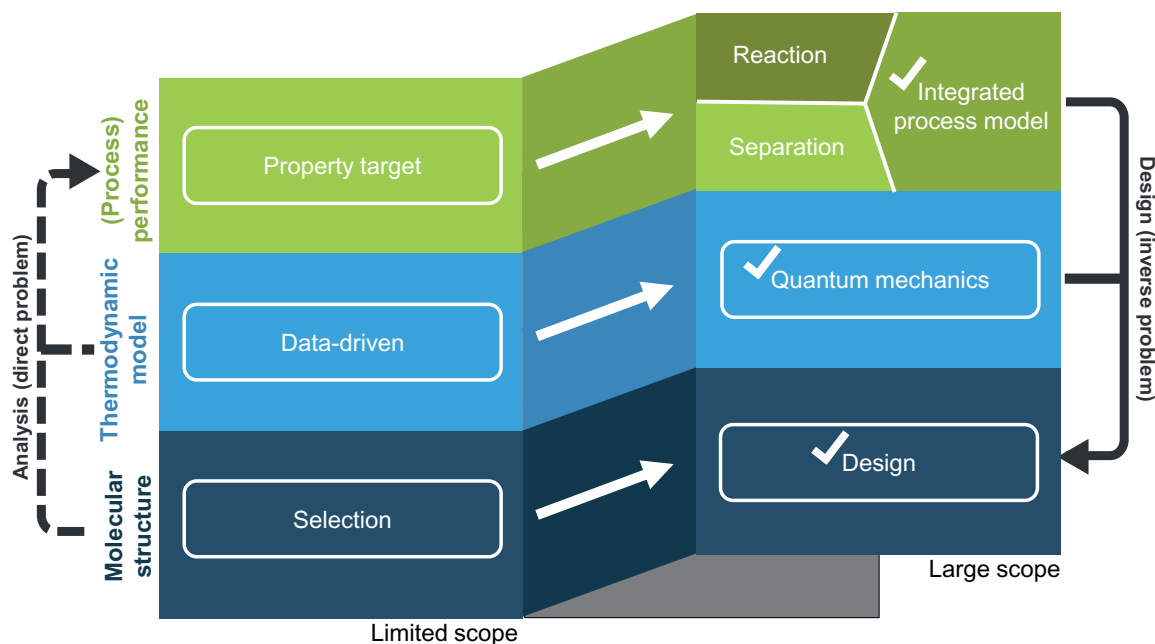


Figure 7.1: Contribution of this thesis in the broader CAMD/CAMPD context. Arrows indicate the target of this thesis, check-marks indicate the established large-scale CAMPD approach in this thesis.

mechanics. The resulting framework allows to explore a large search space (+1000 molecules within days) by optimization. In two case studies, the COSMO-CAMD framework is shown to successfully design novel promising solvents in liquid-liquid extraction of phenol and hydroxymethylfurfural from water. The case studies for phenol and HMF liquid-liquid extraction from water show that COSMO-CAMD successfully designs promising novel solvents, predicted to have better performance than molecules contained in extensive databases. Thus, COSMO-CAMD successfully integrates QM-based property predictions into CAMD and overcomes limitations of both, experimental and database screenings.

The application of COSMO-RS is fully extended towards CAMPD in Chapter 5 by presenting COSMO-CAMPD. COSMO-CAMPD builds upon COSMO-CAMD developed in Chapter 4. To overcome limitations of thermodynamic property-based targets, COSMO-CAMD is combined with pinch-based process models established in Chapter 3. The resulting COSMO-CAMPD framework is applied to a case study for solvent design in an extended process model for the hybrid extraction-distillation process from Chapter 3. The case study shows that COSMO-CAMPD successfully designs promising solvents that are not contained in extensive databases. A comparison of pinch-based process models calculations to rigorous process simulation in

Aspen Plus shows the high accuracy of the pinch-based process models. Predictions of COSMO-CAMPD are validated by experimental liquid-liquid measurements: the promising and commercially available designed solvents 2-methylfuran and toluene achieve 50 % reduction in process energy demand based on experimental LLE data. Thus, COSMO-CAMPD allows for the efficient and successful design of solvents leading to better process performance.

Complex process flowsheets such as the design of CO from H₂ and CO₂ require the consideration of solvent effects in reaction and separation steps. Thus, in Chapter 6, the COSMO-CAMPD framework developed in Chapter 5 is extended to design solvents in complex multiphase reactions and separation flowsheets. Solvents are evaluated by deterministic process flowsheet optimization using pinch-based process models for reaction and separation. The process models consider single-/multiphase equilibrium reactions, distillation, heteroazeotropic distillation and extraction and thus allow to capture process-inherent trade-offs. The optimization approach efficiently designs novel solvents for the production of CO from CO₂ and H₂ based on a sound process-level design target. The optimization considers different chemical storage molecules, process flowsheet variants and continuous process variables. The designed solvents improve the process performance by more than 12 % compared to a massive database screening of over 80,000 combinations of solvents and structural process variants.

Overall, the integration of COSMO-RS into CAMPD in this thesis allows for the design of tailor-made molecules in optimized processes free from limitations to experiment-based, data-driven property prediction and simplified performance targets. The proposed approaches in this thesis thus significantly enlarge the molecular design space while allowing for sound and efficient molecular evaluation on process-level. This thesis thus successfully extends the range of QM-based property prediction with COSMO-RS towards integrated molecular and process design.

7.2 Future Perspective

The integrated molecule and process design approach in this thesis focuses on a process-evaluation of molecules. However, the need for sustainable solvent design has been highlighted (Anastas and Zimmerman, 2016) based on green-chemistry principles (Gilbertson et al., 2015), toxicity limitations (Kostal et al., 2015) as well as a large spectrum of environmental, safety, and health aspects (Whelton et al., 2015; Zimmerman and Anastas, 2015; Voutchkova et al., 2011). Accordingly, environmental assessment of processes in an early development stage has gained increasing interest

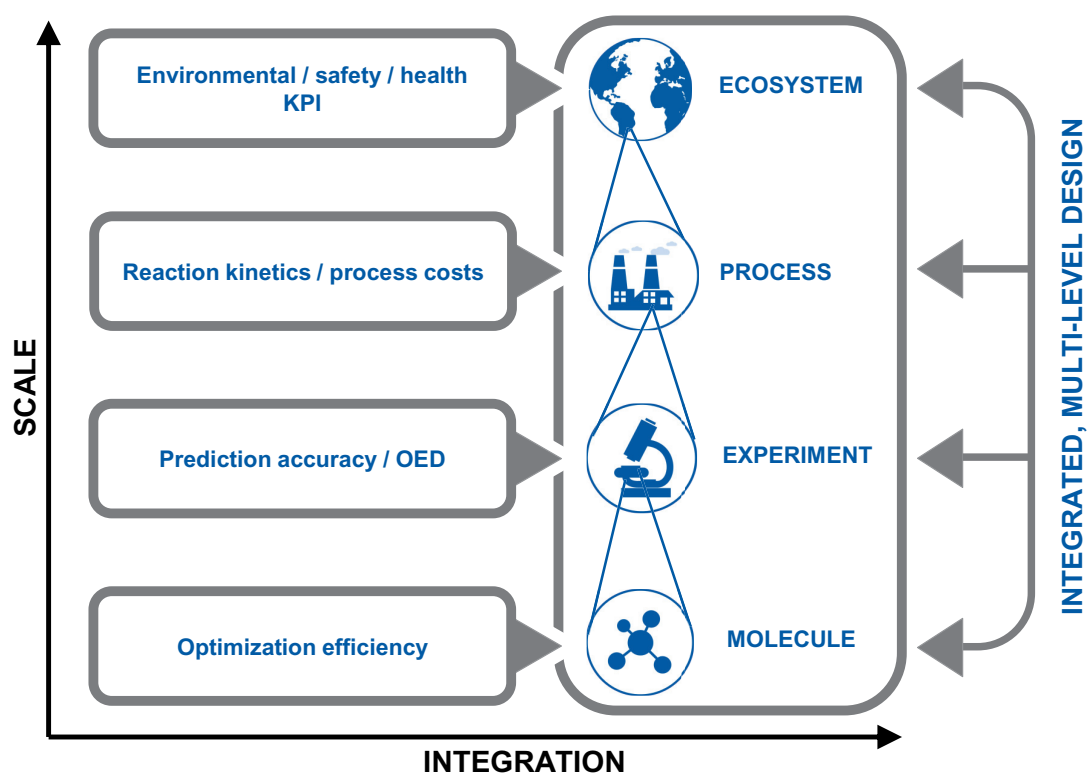


Figure 7.2: Panorama of further research topics classified by level of integration and scale. A detailed description of the highlighted topics is given in the text.

in recent years (Broeren et al., 2017). Conversely, integrated molecule and process design is currently limited to focus on techno-economics, i.e., novel solvent for a process are not selected based on environmental criteria. Highly desirable would thus be a method for early stage process and molecule design considering both economic and environmental performance criteria.

Here, further development of COSMO-CAMPD could bridge the gap between environmental and economic targets. Future research should thus focus on integrating COSMO-CAMPD in a predictive framework that allows for molecular and process design beyond common categories of economic process performance. Instead, large-scale ecosystem implications of process and molecule design should be considered, e.g., based on life-cycle assessment (LCA). Such a predictive framework could be envisioned in a multi-level, integrated design approach spanning levels from molecule to the ecosystem (Fig. 7.2). The multi-level design approach should encompass molecule design, experimental data validation, process optimization and ecosystem consideration (Fig. 7.2). For each level, future research targets can be identified. Eventually, the individual contributions to each level should be integrated in a multi-level design framework. In the following, a perspective on single- and multi-level research targets is given.

Molecule-Level Perspective: COSMO-CAMPD currently relies on heuristic optimization which comes at the risk of limited convergence speed and the lack of knowledge about global optimal solutions (cf. Chapter 2). Here, future research should improve the performance of molecular optimization in COSMO-CAMPD, i.e., speed up the optimization procedure and establish a measure for quality of the optimization results. Current approaches to improve computational speed of QM-based CAMD/CAMPD combine QM-based and data-driven prediction methods. E.g., group-contribution (GC) methods are used to pre-evaluate molecules to identify targets for full QM optimization (Lehmann and Maranas, 2004; Struebing et al., 2013). However, these methods depends on the available data for GC parametrization (cf. Section 2.3.1). Here, data for GC parametrization could be provided by large-scale database screenings presented in this thesis (cf. Chapter 3), e.g., by data regression of GC parameters to screening results. Thus, future research should aim at integrating database screening information into molecular design to improve the optimization efficiency of COSMO-CAMPD. To address the challenge of finding global optimal solutions to CAMD/CAMPD problems, deterministic optimization methods can be employed such as branch-and-bound (BnB) algorithms (cf. Section 2.2.2). Commonly, deterministic optimization methods depend on the relaxation of integer variables that describe the molecular structure. Thereby, current deterministic optimization ap-

proaches rely on simple property prediction methods such as first order GC methods where full structural information of the molecules is neglected (Sahinidis et al., 2003). In contrast, molecules in COSMO-CAMPD are input to iterative DFT/COSMO calculations which require 3D molecular geometry information. Thus, the relaxation of molecular structures is not trivial in COSMO-CAMPD. First, the number of integer variables describing the molecular structure in COSMO-CAMPD is tremendous compared to, e.g., first-order GC description leading to a significant increase of optimization variables. Second, for each value of these relaxed optimization variables, deterministic optimization algorithms would require full iterative QM calculations which comes at high computational costs. An alternative to optimization based on relaxation are derivative-free optimization algorithms which still allow for near-global solutions (Rios and Sahinidis, 2013). Derivative-free optimization algorithms could thus be a possible way to increase optimization performance in COSMO-CAMPD and should be further investigated.

Experiment-Level Perspective: For multi-scale molecular and process design problems, reliability becomes an issue that requires optimization under uncertainty (Sahinidis, 2004; Zhou et al., 2015a). The key for a reliable process-level assessment of novel molecules and processes are experimentally validated input data for process simulations. To increase reliability and applicability of process model predictions, so-called self-optimized process design concepts have been proposed with integrated feedback from experiments (Lapkin et al., 2017; Houben and Lapkin, 2015). However, mostly empirical models such as response-surface models are currently employed in such approaches (Jacob et al., 2017). Conversely, a direct feedback to processes from experimental measurements by optimal experimental design (OED) can significantly improve process modeling quality (Franceschini and Macchietto, 2008). Currently, OED is mostly limited to improvement of a single experiment that yield one kind of physical property (Franceschini and Macchietto, 2007). A system-wide feedback from OED in a closed-loop COSMO-CAMPD optimization procedure would thus be highly desirable. Here, future research should focus on improving accuracy and reliability of COSMO-CAMPD predictions by learning from optimally designed experiments. E.g., an integrated framework for selection and refinement of thermodynamic models could be envisioned. The framework should use optimal experimental design (OED) to gather sensitivities of process models for predictive thermodynamic data. Based on these sensitivities, optimal experiments for maximum improvement in predictive data need to be identified. Furthermore, an automated experimental measurement procedure could be established for improving quality of thermodynamic prediction by an adaptive selection of appropriate thermodynamic prediction. E.g., coupling

COSMO-RS with equation of state models leads to improved quality of property prediction (Kaminski et al., 2017). Re-parametrized models can then integrate novel data from higher-level theory and/or experiments to improve COSMO-CAMPD prediction quality. Overall, a combination of COSMO-CAMPD with OED reduces experimental effort and offers guaranteed accuracy and validated predictions.

Process-Level Perspective: Currently, process-level models in COSMO-CAMPD provide input for the predictive assessment of energy demand in molecule and process design. The COSMO-CAMPD approach is based on pinch-based process models that are inherently limited to the assumption of infinite separation stages and equilibrium reactions. These models cannot account for overall process costs and the ranking of designed molecules might be biased. In order to enable a process evaluation directly based on process cost, two major tasks need to be tackled: First, costs for reaction steps need to be determined. Reaction steps scale with the size of employed reactors that in turn depend on reaction kinetics. Similarly, complex reaction networks can lead to undesired selectivity and yield, which often requires elaborate reactor designs. Recently, predictive reaction models using quantum chemistry have been successfully applied to solvent selection problems in chemical reaction engineering (Austin et al., 2018; Zhou et al., 2015a; Struebing et al., 2013). Thus, future research should extend COSMO-CAMPD to account for reaction costs based on reaction kinetics from QM property prediction. Second, costs for reaction steps need to be integrated in a cost-based assessment of the overall process. Here, COSMO-CAMPD needs to be combined with optimization-based evaluation of the minimum total annual costs by rigorous equilibrium tray models. E.g., a hierarchical approach can be envisioned that designs molecules employing pinch-based process models and assessing top designed molecules based on rigorously determined process cost. First investigations show that integration of rigorous Aspen Plus process calculations in an automated procedure is possible with good initialization by pinch-based process models (Mayer, 2017). However, computational costs are comparably large to pinch-based models and the reliability of the approach is questionable due to tedious convergence procedures. A possible solution is the employment of rigorous optimization by solving a process superstructure in an automatic solution strategy as proposed by Skiborowski et al. (2015a).

Ecosphere-Level Perspective: Currently, COSMO-CAMPD design is based on process-level design targets while neglecting environmental aspects of designed molecules or processes. Environmental assessment of molecules however requires comprehensive LCA data which is usually not available or limited in early design phases. Thus, reliable prediction methods need to be established that translate molecular properties

to environmental key performance indicators (KPI), e.g., global warming potential, safety and health. Recent work has shown that environmental KPI such as toxicity can be predicted with in silico methods. An example is the prediction of acute aqueous toxicity LC50 by simple descriptors of molecules in COSMO-RS (Ghanem et al., 2017). Additionally, predictive Ecological Structure Activity Relationship (ECOSAR) models exist that estimate aquatic short-term (acute) and long-term (chronic) toxicity (Heger et al., 2012). Similarly, toxicokinetic and toxicodynamic modeling predicts human toxicity of chemicals (Cedergreen et al., 2017). In LCA, early approaches exist that systematically aim at evaluating environmental impacts using automated process evaluations (Kim and Overcash, 2003). Methods to predict environmental KPI should thus be extensively investigated and integrated in COSMO-CAMPD.

Integrated Multi-Level Design Approach Perspective: Finally, all of the above single-level research perspectives should be combined in a predictive multi-level molecular and process design surpassing common categories of process performance assessment. Novel designed processes and molecules should be assessed in a system-wide approach using all scales involved: from molecular design using integrated experimental validation to process optimization based on economic and environmental design targets. Such an approach thus allows for designing economically viable *and* environmentally benign molecules and processes, e.g., reaction-separation processes using solvents according to green chemistry principles (Gilbertson et al., 2015). The presented COSMO-CAMPD framework in this thesis forms the foundation of such an endeavor.

APPENDIX A

Process-Level Solvent Evaluation

Property prediction using COSMO-RS

All thermodynamic properties in this thesis are predicted by COSMO-RS (Klamt et al. (2010)). COSMO files are used from the 1501-BP-TZVP databank and COSMO-RS calculations are performed using COSMOthermX Version 15 (COSMOlogic, 2015d).

Pure component properties used in this thesis are vapor pressure, $p_{\text{sat},X}$, and molar enthalpy of vaporization, $\Delta h_{\text{vap},X}$, for each solvent candidate X . The vapor pressure is calculated using the Antoine equation

$$p_{\text{sat},X} = \exp \left(A_X - \frac{B_X}{C_X + T} \right). \quad (\text{A.1})$$

The solvent specific coefficients, A_X , B_X and C_X , are calculated by COSMO-RS using built-in calculation routines in COSMOthermX (COSMOlogic, 2015d).

The molar enthalpy of vaporization, $\Delta h_{\text{vap},X}$, is estimated by combining the Antoine equation, Eq. (A.1), with the Clausius-Clapeyron equation

$$\frac{d \ln (p_{\text{sat},X}/p_{\text{sat},X}^0)}{dT} = \frac{\Delta h_{\text{vap},X}}{RT^2}, \quad (\text{A.2})$$

which leads to

$$\Delta h_{\text{vap},X} = \frac{B_X}{(T + C_X)^2} RT^2. \quad (\text{A.3})$$

Non-ideal liquid mixture properties are expressed using activity coefficients $\gamma_i(\mathbf{z}, T)$ that are calculated by COSMO-RS. With $\gamma_i(\mathbf{z}, T)$, vapor-liquid equilibria (VLE) and liquid-liquid equilibria (LLE) are calculated.

In this thesis, activity coefficients $\gamma_i(\mathbf{z}, T)$ are expressed by the non-random-two-liquid NRTL model (Renon and Prausnitz, 1968).

The extraction shortcut process model uses isothermal binary NRTL parameters to calculate $\gamma_i(\mathbf{z})$ in a mixture of components i and j . Isothermal binary NRTL parameters τ_{ij} , τ_{ji} , α_{ij} and α_{ji} are defined by

$$\ln \gamma_i(\mathbf{z}) = z_j^2 \left[\tau_{ji} \left(\frac{G_{ji}}{z_i + z_j G_{ji}} \right)^2 + \frac{\tau_{ij} G_{ij}}{(z_j + z_i G_{ij})^2} \right], \quad (\text{A.4})$$

$$\ln \gamma_j(\mathbf{z}) = z_i^2 \left[\tau_{ij} \left(\frac{G_{ij}}{z_j + z_i G_{ij}} \right)^2 + \frac{\tau_{ji} G_{ji}}{(z_i + z_j G_{ji})^2} \right], \quad (\text{A.5})$$

$$\ln G_{ij} = -\alpha_{ij} \tau_{ij}, \quad (\text{A.6})$$

$$\ln G_{ji} = -\alpha_{ji} \tau_{ji}. \quad (\text{A.7})$$

Generally, $\alpha_{ij} = \alpha_{ji} = \alpha$ is assumed.

In this thesis, an equally-spaced concentration grid is used for $\mathbf{z} \in [0, 1]$ with $n_{\mathbf{z}} = 11$ evaluations of $\gamma(\mathbf{z})$ to determine τ_{ij} , τ_{ji} , α_{ij} and α_{ji} .

The distillation shortcut process model uses temperature-dependent binary NRTL parameters to calculate $\gamma_i(\mathbf{z}, T)$. In this thesis, temperature-dependent binary NRTL parameters are calculated using regression parameters a_{ij} , b_{ij} , c_{ij} , d_{ij} , e_{ij} , f_{ij} which are defined by

$$\tau_{ij} = a_{ij} + \frac{b_{ij}}{T} + c_{ij} \ln T + d_{ij} T, \quad (\text{A.8})$$

$$\alpha_{ij} = e_{ij} + f_{ij} T. \quad (\text{A.9})$$

To determine the parameters a_{ij} , b_{ij} , c_{ij} , d_{ij} , e_{ij} , f_{ij} in this thesis, an equally-spaced temperature grid is used for $T \in [298.15 \text{ K}, 473.15 \text{ K}]$ with $n_T = 10$ temperature grid points. Overall, $n_{\mathbf{z}} = 11$ evaluations $\gamma_i(\mathbf{z}, T)$ on each of the $n_T = 10$ temperature levels are performed, i.e., 110 evaluations for each component.

Conventional screening criteria

The phase distribution coefficient P is defined as

$$P = \frac{z_{\text{GVL,E}}}{z_{\text{GVL,R}}} = \frac{\gamma_{\text{GVL,R}}}{\gamma_{\text{GVL,E}}}. \quad (\text{A.10})$$

Here, E denotes the extract phase and R the raffinate phase of a binary LLE between a solvent candidate X and water. P is evaluated at $T = 298.15$ K and infinite dilution of the solute GVL is assumed.

The relative volatility α (Pretel et al. (1994)) is calculated according to

$$\alpha = \frac{\gamma_{\text{water},X}^{\infty} \cdot p_{\text{sat,water}}}{\gamma_{\text{GVL},X}^{\infty} \cdot p_{\text{sat,GVL}}}, \quad (\text{A.11})$$

where α is evaluated at $T = 298.15$ K with infinite dilution of water in solvent candidate X for $\gamma_{\text{water},X}^{\infty}$ and infinite dilution of GVL in solvent candidate X for $\gamma_{\text{GVL},X}^{\infty}$.

Extended Screening results Massive Screening

The results of the solvent screening based on minimum energy demand $Q_{\min}/\text{MJ kmol}^{-1}$ is presented in Tab. A.1 for all solvents with a lower minimum energy demand $Q_{\min}$ than the benchmark molecule n-butyl acetate. Abbreviations: PPS = propenylester, PES = propylester.

Table A.1: Extended list of the screening results

Rank	Solvent	$Q_{\min}$	Rank	Solvent	$Q_{\min}$
1	1,5-hexadiyne	3.70	79	3-butenoidacidethylester	8.64
2	1,6-heptadiyne	3.94	80	benzene	8.67
3	1,3-hexadien-5-yne	3.99	81	cis-1,3,5-hexatriene	8.69
4	1,7-octadiyne	4.12	82	2,3-dihdropyran	8.71
5	1-penten-4-yne	4.14	83	3-methyl-2-butenoidacidmethylester	8.73
6	2-methyl-1-buten-3-yne	4.28	84	methylemethacrylate	8.73
7	trans-3-penten-1-yne	4.33	85	4-pentenal	8.74
8	ethoxyacetylene	4.46	86	2-propenoidacidethylester	8.75
9	furan	4.64	87	hexenylformiat trans3	8.79
10	3-methyl-3-penten-1-yne	4.75	88	ethylpyruvate	8.80

Table continues on next page

Appendix A Process-Level Solvent Evaluation

Rank	Solvent	$Q_{\min}$	Rank	Solvent	$Q_{\min}$
11	pentynes	4.83	89	n-crotylacetate	8.81
12	3-methyl-1-butyne	4.88	90	diethylcarbonate	8.83
13	divinylether	4.99	91	trans-1,3,5-hexatriene	8.85
14	2-vinylfuran	5.06	92	styrene	8.85
15	3-methylfuran	5.26	93	tert-butylformate	8.87
16	1,8-nonadiyne	5.33	94	n-propylformate	8.93
17	ethynylbenzene	5.40	95	tetramethylorthocarbonate	8.94
18	3,3-dimethyl-1-butyne	5.58	96	2-methyl-1-penten-3-ol	8.94
19	4-methylene-2-oxetanone	5.58	97	3,4-dihydro-2-methoxy-2h-pyran	8.96
20	4-methyl-1-pentyne	5.70	98	1,3,5,7-cyclooctatetraene	8.98
21	hexyne	5.71	99	2,4-dimethylfuran	8.98
22	2-methoxyfuran	5.75	100	1,1-diallylethanol	8.98
23	vinylformate	5.86	101	2-propenoicacidPES	9.02
24	1-methoxy-1,3-butadiene	6.17	102	propanoicacid,anhydride	9.02
25	5-methyl-1-hexyne	6.20	103	salicylaldehide	9.06
26	ethyl2-propynoate	6.31	104	2-methoxyphenol	9.06
27	2-methylfuran	6.36	105	vinylloxirane	9.09
28	1-heptyne	6.52	106	ethylmethylcarbonate	9.09
29	allylvinylether	6.57	107	methylacrylate	9.11
30	2,3-pentanedione	6.65	108	3,3-dimethyl-2-oxobutanoicacid	9.17
31	vinylacetate	6.75	109	methylisovalerate	9.20
32	acetylperoxide	6.90	110	2-propenoicacid-1-methylethylester	9.20
33	allylformate	6.95	111	methylbutyrate	9.27
34	2,3-hexanedione	6.98	112	2-(ethenyloxy)-2-methylpropane	9.29
35	3-ethyl-1-pentyn-3-ol	7.08	113	2-methylpropanoicacidmethylester	9.30
36	3,4-hexanedione	7.15	114	2-(ethenyloxy)-propane	9.31
37	3,4-dimethyl-1-pentyn-3-ol	7.18	115	propionaldehydediethylacetal	9.33
38	2,3-dihydro-1,4-dioxin	7.28	116	diethylperoxide	9.34
39	4-methyl-2,3-pentanedione	7.31	117	hexanal	9.37
40	1,2-cyclobutanedione	7.34	118	methyl-trans-2-butenate	9.37
41	2,4-hexadiyne	7.36	119	3-cyclohexene-1-carboxaldehide	9.41
42	propanoicacidethylester	7.36	120	methylvalerate	9.42
43	benzofuran	7.47	121	methylglyoxal	9.45
44	1,3-benzodioxole	7.50	122	1-hexen-3-ol	9.45
45	allylacrylate	7.65	123	peroxyaceticacid	9.46

Table continues on next page

Rank	Solvent	$Q_{\min}$	Rank	Solvent	$Q_{\min}$
46	cyclopropylvinylether	7.67	124	cyclopentadiene	9.47
47	isopropenylacetate	7.76	125	3,5-dimethyl-1-hexen-3-ol	9.50
48	3-methoxy-propyne	7.86	126	3-methylpentanal	9.51
49	5-methyl-2,3-hexanedione	7.93	127	4,5-octanedione	9.54
50	octyne	7.95	128	cycloheptatriene	9.54
51	1,5-hexadien-3-yne	7.95	129	2-methyl-2-propenoicacid-2-PPS	9.56
52	1-hexyn-3-ol	7.96	130	2-methylphenol	9.63
53	3-methyl-1-pentyn-3-ol	8.00	131	trans-ethylcrotonate	9.67
54	1-penten-3-yne	8.02	132	(4r)-4-methylhexanal	9.67
55	butylformate	8.05	133	2-hexen-4-yne	9.67
56	propanoicacid-2-PPS	8.07	134	heptanal	9.68
57	2,3-heptanedione	8.08	135	(4s)-4-methylhexanal	9.68
58	formicacidpentylester	8.09	136	2-PPSbutanoicacid	9.71
59	2-ethylfuran	8.11	137	methyl2methylbutyrate	9.72
60	1,5-hexadien-3-ol	8.15	138	methyl-(s)-(+)-2-methylbutanoate	9.73
61	isoamylformate	8.17	139	ethylmethacrylate	9.75
62	(z)-3-hexenal	8.21	140	methylcyclopropylcarboxylate	9.75
63	butadione	8.22	141	isopropylformate	9.76
64	diallylether	8.24	142	1,6-heptadien-4-ol	9.81
65	anisole	8.24	143	1-penten-3-ol	9.82
66	isobutylformate	8.32	144	2,5-dimethylfuran	9.83
67	hexenylformiate cis3	8.34	145	formicacidhexylester	9.87
68	2,3-dihydrofuran	8.34	146	methyltiglate	9.87
69	2-(methoxymethyl)-furan	8.34	147	3-butyn-2-one	9.91
70	1-hexyn-5-one	8.41	148	ethylbutyrate	9.95
71	aceticanhydride	8.42	149	5-methyl-5-hexen-2-one	9.98
72	4-methyl-1-penten-3-ol	8.47	150	ethylpropionate	9.99
73	ethoxy-ethene	8.47	151	2,4-hexadienoicacidmethylester	10.03
74	aceticacid-2-PPS	8.51	152	2-methyl-pentanal	10.04
75	formicacid-1-methylPES	8.52	153	cis-5-octenal	10.04
76	vinylbutyrate	8.54	154	3-methyl-2-oxobutanoicacid	10.09
77	sec-butenylacetate	8.55	155	3-methylhexanal	10.09
78	4-heptenal	8.64			

End of table

Rigorous validation of the screening results

Rigorous process model calculations using Aspen Plus V8.4 are performed in an automated procedure to validate the results from shortcut process models for the top 50 solvents. The minimum energy demand $Q_{\min}$ is calculated by Aspen Plus using the models 'Extract' and 'Radfrac'. In Aspen Plus, the flowsheet presented in Chapter 3 is initialized with results from the shortcut models and converged to minimum solvent flow and minimum reboiler heat duty.

Minimum energy demand $Q_{\min}$ is determined to achieve 99.9999 % recovery of GVL in the extract stream E and 99.9999 % purity and recovery of the solute A in the bottom stream B. Infinite numbers of stages, that are assumed in the process shortcut, are approximated by using $n = 100$ stages in the extraction column and $n = 200$ stages (Feed stage at $n = 100$) in the distillation column. Convergence of the distillation column is achieved using 'standard', 'nonideal' or 'azeotropic' solver. Standard flash convergence algorithm is used. The results of the validation of the minimum energy demand $Q_{\min}$ is presented in Table A.2. Non-converged Aspen Plus calculations in the automated validation procedure are indicated by 'n.c.'.

Table A.2: Extended list of the screening results

Rank	Solvent	$Q_{\min}$ MJ kmol ⁻¹	$Q_{\min, \text{Aspen}}$ MJ kmol ⁻¹	$R_{\min, \text{Aspen}}$ kmol kmol ⁻¹
1	1,5-hexadiyne	3.70	3.97	0.05
2	1,6-heptadiyne	3.94	4.40	0.11
3	1,3-hexadien-5-yne	3.99	3.99	0.03
4	1,7-octadiyne	4.12	4.71	0.14
5	1-penten-4-yne	4.14	n.c.	-
6	2-methyl-1-buten-3-yne	4.28	n.c.	-
7	trans-3-penten-1-yne	4.33	3.76	$2.5 \cdot 10^{-4}$
8	ethoxyacetylene	4.46	4.15	0.03
9	furan	4.64	3.82	$1.6 \cdot 10^{-4}$
10	3-methyl-3-penten-1-yne	4.75	5.14	0.03
11	pentyne	4.83	4.26	0.01
12	3-methyl-1-butyne	4.88	4.49	0.01
13	divinylether	4.99	4.00	0.01
14	2-vinylfuran	5.06	5.31	0.09
15	3-methylfuran	5.26	5.00	0.03
16	1,8-nonadiyne	5.33	6.11	0.40

Table continues on next page

Rank	Solvent	$Q_{\min}$ MJ kmol ⁻¹	$Q_{\min, \text{Aspen}}$ MJ kmol ⁻¹	$R_{\min, \text{Aspen}}$ kmol kmol ⁻¹
17	ethynylbenzene	5.40	5.97	0.40
18	3,3-dimethyl-1-butyne	5.58	5.10	0.02
19	4-methylene-2-oxetanone	5.58	6.08	0.13
20	4-methyl-1-pentyne	5.70	5.12	0.02
21	hexyne	5.71	5.30	0.02
22	2-methoxyfuran	5.75	6.46	0.27
23	vinylformate	5.86	5.90	0.11
24	1-methoxy-1,3-butadiene	6.17	5.64	0.03
25	5-methyl-1-hexyne	6.20	6.28	0.05
26	ethyl 2-propynoate	6.31	7.03	0.28
27	2-methylfuran	6.36	5.47	0.03
28	1-heptyne	6.52	6.62	0.06
29	allylvinylether	6.57	5.74	0.02
30	2,3-pentanedione	6.65	6.75	0.04
31	vinylacetate	6.75	6.24	0.02
32	acetylperoxide	6.90	7.88	0.33
33	allylformate	6.95	6.98	0.02
34	2,3-hexanedione	6.98	7.30	0.05
35	3-ethyl-1-pentyn-3-ol	7.08	8.21	0.10
36	3,4-hexanedione	7.15	7.21	0.02
37	3,4-dimethyl-1-pentyn-3-ol	7.18	8.31	0.10
38	2,3-dihydro-1,4-dioxin	7.28	7.66	0.04
39	4-methyl-2,3-pentanedione	7.31	7.48	0.07
40	1,2-cyclobutanedione	7.34	8.13	0.20
41	2,4-hexadiyne	7.36	7.30	0.07
42	propanoic acid ethylester	7.36	7.09	0.04
43	benzofuran	7.47	8.49	0.67
44	1,3-benzodioxole	7.50	8.41	0.81
45	allylacrylate	7.65	8.39	0.14
46	cyclopropylvinylether	7.67	6.60	0.03
47	isopropenylacetate	7.76	7.92	0.05
48	3-methoxy-propyne	7.86	9.13	0.16
49	5-methyl-2,3-hexanedione	7.93	8.71	0.177
50	octyne	7.95	8.72	0.134

End of table

Overall, the calculated minimum energy demand by the screening approach and the ASPEN simulations are in very good agreement (Mean average percentage error = +9%, Spearman's rank correlation coefficient 0.93 for validated top 50 solvents). The Aspen Plus simulations show that two components, trans-3-penten-1-yne and furan, ranked # 7 and # 9 by the screening require even less energy demand than 1,5-hexadiyne according to the Aspen Plus simulations. This effect is due to different convergence settings in the flowsheets: Aspen Plus calculates $R_{min} < 10^{-3}$ (cf. Table A.2, # 7 and # 9) whereas the shortcut process models determines only $R_{min} > 10^{-3}$. Additionally, a finite number of stages in Aspen Plus is assumed whereas infinite number of stages is assumed in the process shortcut models.

Statistical measures

The statistical measures used in this thesis are briefly described in the following.

The accuracy of a linear regression model $\hat{y}(x)$ is expressed by the coefficient of determination R^2 (Eq. (A.12)). $R^2 = 1$ corresponds to a perfect linear correlation. $R \approx 0$ indicates that there is no linear correlation.

$$R^2 = \frac{\sum_i^N (\hat{y}_i - \bar{y})^2}{\sum_i^N (y_i - \bar{y})^2} \quad (\text{A.12})$$

where $\hat{y}_i = \hat{y}(x_i)$ is the predicted value and y_i the observed one.

The root-mean-square error can be also used to measure the size of the discrepancy between predicted values and the observed ones. In this case, the $RMSE$ is calculated by

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (\hat{y}_i - y_i)^2} \quad (\text{A.13})$$

Strength of a correlation: The Pearson correlation coefficient $r_{a,b}$ is defined by

$$r_{a,b} = \frac{\sum_{i=1}^N (a_i - \bar{a}) (b_i - \bar{b})}{\sqrt{\sum_{i=1}^N (a_i - \bar{a})^2} \cdot \sqrt{\sum_{i=1}^N (b_i - \bar{b})^2}}. \quad (\text{A.14})$$

In Eq. A.14, r is the Pearson correlation coefficient, $a = \{a_1, \dots, a_i, \dots, a_N\}$ and

$b = \{b_1, \dots, b_i, \dots, b_N\}$ are data sets of possibly correlated sample values a_i and b_i and $\bar{a}$ and $\bar{b}$ are mean values of a and b respectively. The Pearson correlation coefficient $r_{a,b}$ can take on values from $r_{a,b} = -1$ to $r_{a,b} = 1$. Pearson correlation coefficients of $r_{a,b} = -1$ and $r_{a,b} = 1$ correspond to an ideal correlation of the sample values a_i and b_i . In contrast, $r_{a,b} = 0$ indicates no correlation of a_i and b_i .

Similarly, the Spearman's or rank correlation coefficient $r_{x,y}$ determines if a correlation between two rankings x and y is monotonous. For $r_{x,y} = 1$ the relation between ranks is strictly monotonously increasing, i.e. $x_2 < x_1 \Leftrightarrow y_2 < y_1$, and for $r_{x,y} = -1$ strictly monotonously decreasing. For $r_{x,y} \approx 0$ no monotony is found. The Spearman's coefficient can be approximated for rankings with size N by

$$r_{x,y} = 1 - \frac{6}{N(N^2 - 1)} \sum_i^N (\text{rank}(x_i) - \text{rank}(y_i))^2. \quad (\text{A.15})$$

APPENDIX B

Computer-Aided Molecular Design

In this thesis, the following parameters and specifications for the genetic algorithm LEA3D are set:

Table B.1: Parameters for optimization with LEA3D

Parameter	Value
Mutation probability	70 %
suppress fragment	25 %
add fragment	25 %
replace fragment	25 %
permute/move/scramble fragment	25 %
Crossover probability	30 %
Elitism-Strategy	true

The figures B.1 and B.2 show the molecular fragments for the genetic algorithm LEA3D that are included in the initial molecular fragment libraries. The X marks a connector in the molecular fragment. Each X can be connected to an X from another molecular fragment in the genetic algorithm to build a new molecular structure. Unconnected connectors X are automatically replaced by hydrogen atoms. The structural feasibility of the molecular structures is ensured by LEA3D.

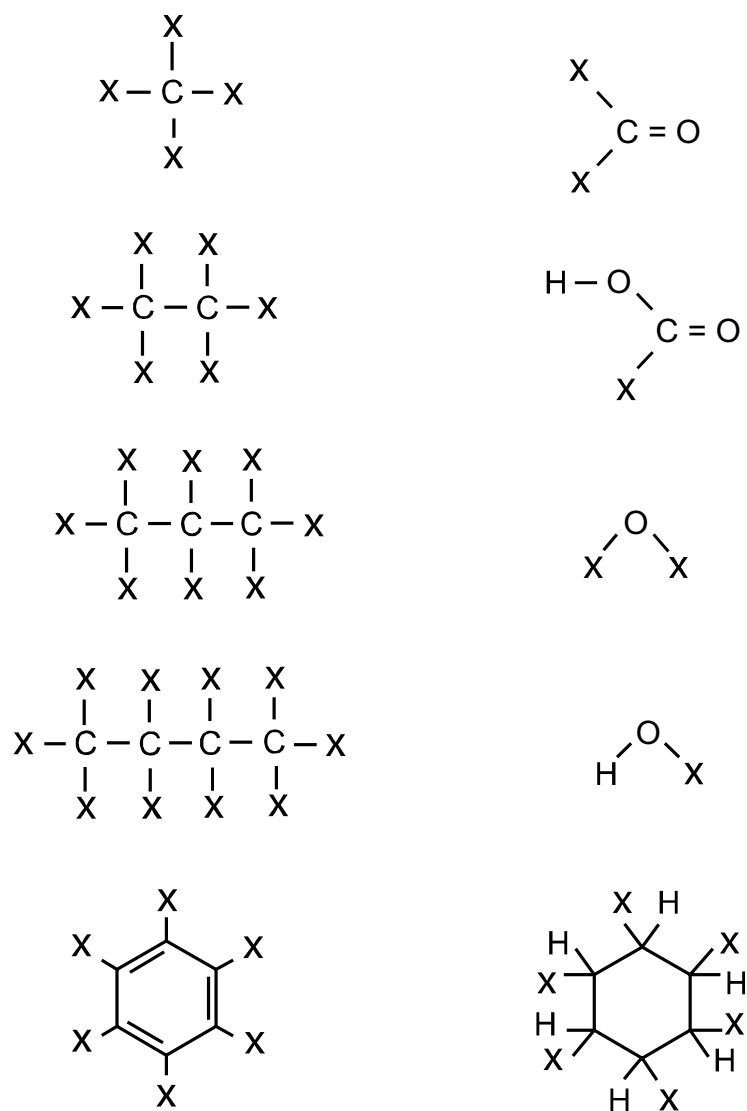


Figure B.1: Molecular fragment library for case study 1: Extraction of phenol from water.

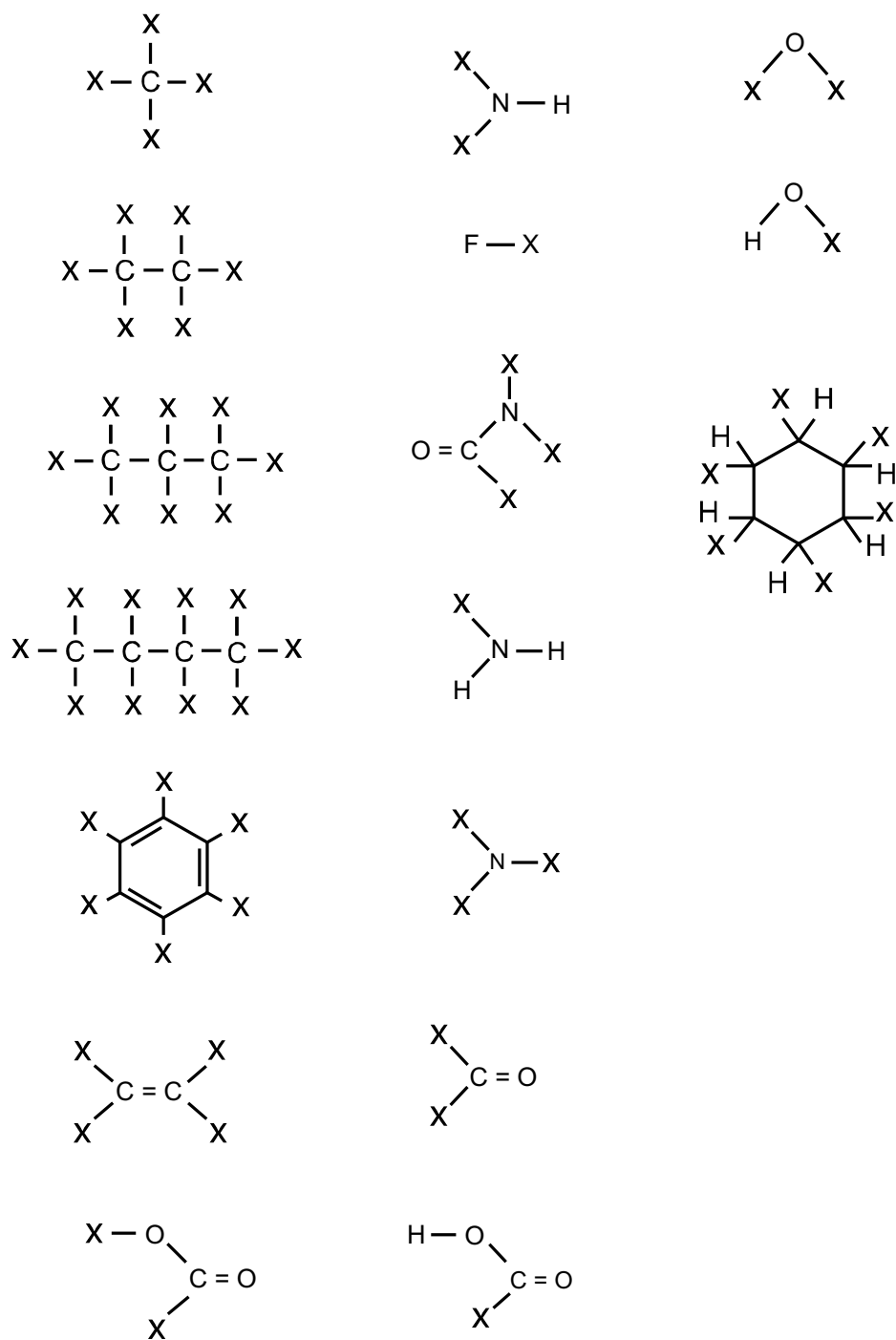


Figure B.2: Molecular fragment library for case study 2: Extraction of HMF from water.

APPENDIX C

Computer-Aided Molecular and Process Design

Process assumptions and specifications for the purification of GVL

All parameters are predicted using COSMO-RS, i.e., Antoine coefficients and temperature-dependent NRTL coefficients (cf. Appendix A). The specifications for the hybrid extraction-distillation process are taken from Chapter 3. The temperatures of the extraction column is considered as a degree of freedom. Tables C.1 and C.2 summarize the specifications.

Block specifications:

Table C.1: Block specifications for the hybrid extraction-distillation process.

Block	Specification
Extraction column	Raffinate is free of GVL Temperature T_{Extr} optimized with respect to the minimal energy demand
Distillation column	All GVL in the feed is recovered as a pure stream at the bottom All entering and leaving streams are liquid Pressure: 1 bar
Phase split	Temperature $T = 25\text{ }^{\circ}\text{C}$

Stream specifications:

Table C.2: Stream compositions for the hybrid extraction-distillation process.

Stream	Specification
Feed (F)	95% water, 0.05% GVL
Fresh solvent (S_{rep})	Same composition as organic phase from phase split

Details on the COSMO-CAMPD optimization

Calculations have been performed using 8 parallelized cores. Table C.3 lists statistics for all optimization runs. It should be emphasized that the time demand strongly correlates with the number of COSMOconf calculations.

The comprehensive fragment library employed in the optimization runs for GVL purification is given in Fig. C.1.

Table C.3: Statistics for the COSMO-CAMD optimization for GVL. The first four rows add up to the total amount of 2,040 molecules. If a COSMO file is available, it is taken from the database, otherwise it is created by COSMOconf. If a fragmentation failure occurs or the molecule contains more than 12 bold atoms (others than H) or undesired functional groups, no COSMO file is created.

	GVL run 1	GVL run 2	GVL run 3
COSMOconf	237	70	152
Database	1,502	1,817	1,636
Fragmentation failure	51	5	6
Not built	250	148	246
Time (d:h:m:s)	1:17:19:17	0:14:40:10	1:07:02:05

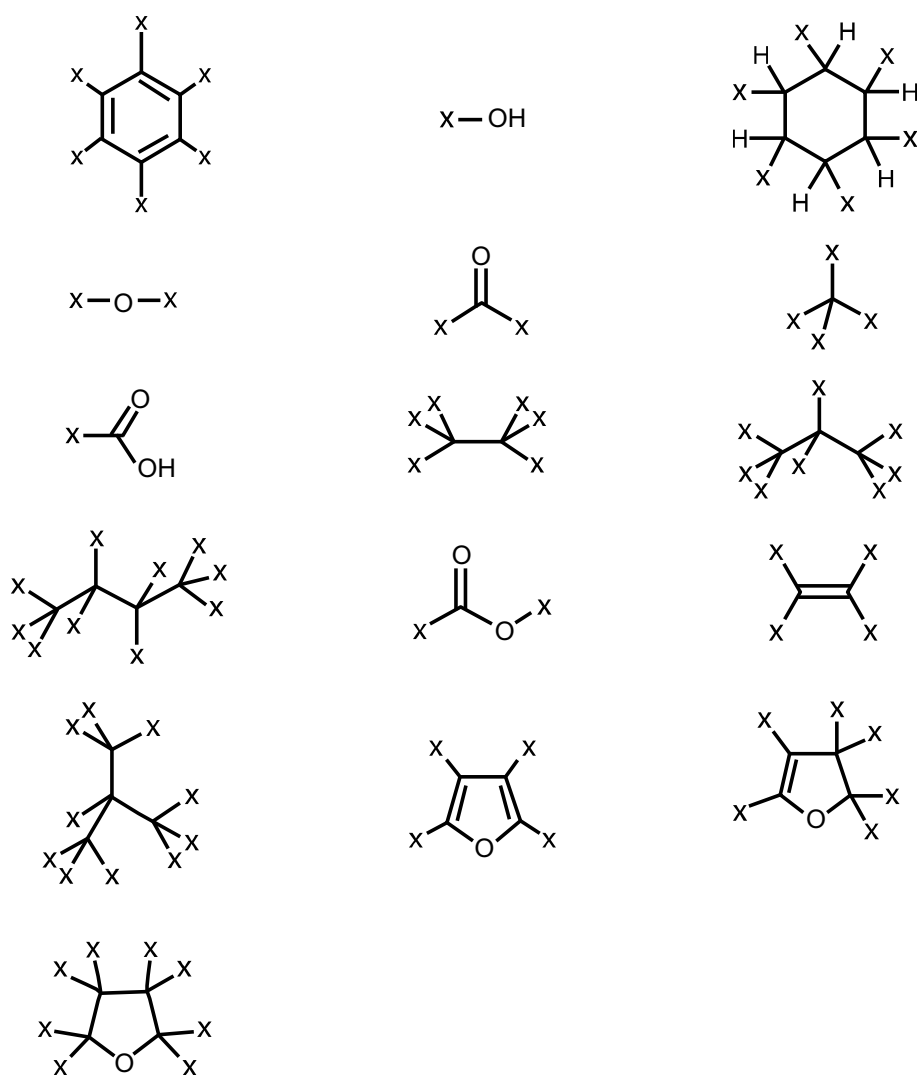


Figure C.1: Molecular fragment library for GVL purification.

Rigorous process simulations

Rigorous process simulations are performed using Aspen Plus version 8.4. COSMO-RS predicted parameters for Antoine's equation and the NRTL model are used. The heat of vaporization is expressed by a polynomial of third order which is fitted to

the results of the Clausius Clapeyron equation (cf. Appendix A). All other pure compound data, which is not available by COSMO-RS prediction such as heat capacities, are assumed to be constant generic values (e.g., heat capacity of water for all solvents) to obtain systematic errors between different solvents. The impact of different values for heat capacities on the distillation result was tested and shown to be negligible.

In Aspen simulations, the *Extract* model is used for extraction and the *RadFrac* model for distillation. The same specifications for the blocks are used as in the pinch-based process models. Assumptions for approximating infinite trays and sharp splits can be found in Table C.4.

Table C.4: Assumptions for the rigorous simulation of the hybrid extraction-distillation process.

Block	Specification
Extraction column	99.99% of GVL is recovered in the distillate. 100 stages are used.
Distillation column	99.99% of GVL is recovered at the bottom. The bottom stream is 99.99% pure. 50 trays are used, the feed stage is number 25.

Experimental Methods

The measurement setup uses a liquid-handling autosampler, which prepares LLE with a total volume of 1 mL in a temperature-controlled tray and takes samples from the upper and the lower phase after equilibration. The samples are directly injected into a gas chromatograph without dilution nor addition of a standard. A detailed description of the experimental procedure can be found in Dechambre et al. (2014a).

Table C.5: List of chemicals

	provider	specification	purity (GC, A %)
γ -valerolactone	Sigma-Aldrich	ReagentPlus [®]	99
water	Merck	SupraSolv [®]	
toluene	VWR	spectronorm	99.8
2-methylfuran	Sigma-Aldrich		99
n-butyl acetate	Merck	for analysis	99.5

Table C.6: Gas chromatography parameters

parameter	value
column	FS-CW20M-CB
carrier gas	He
inlet temperature in K	523.15
split	25 : 1
p in bar	0.3
T1 in K	373.15
hold in min	4
Tramp in K min ⁻¹	50
T2 in K	453.15
hold in min	9
detector	TCD/FID
detector temperature in K	523.15523.15

Results of the LLE measurements

A comparison of COSMO-RS predictions on TZVP level and TZVPD-FINE level for the investigated solvents n-butyl acetate and toluene are shown in Fig. C.2 and Fig. C.3.

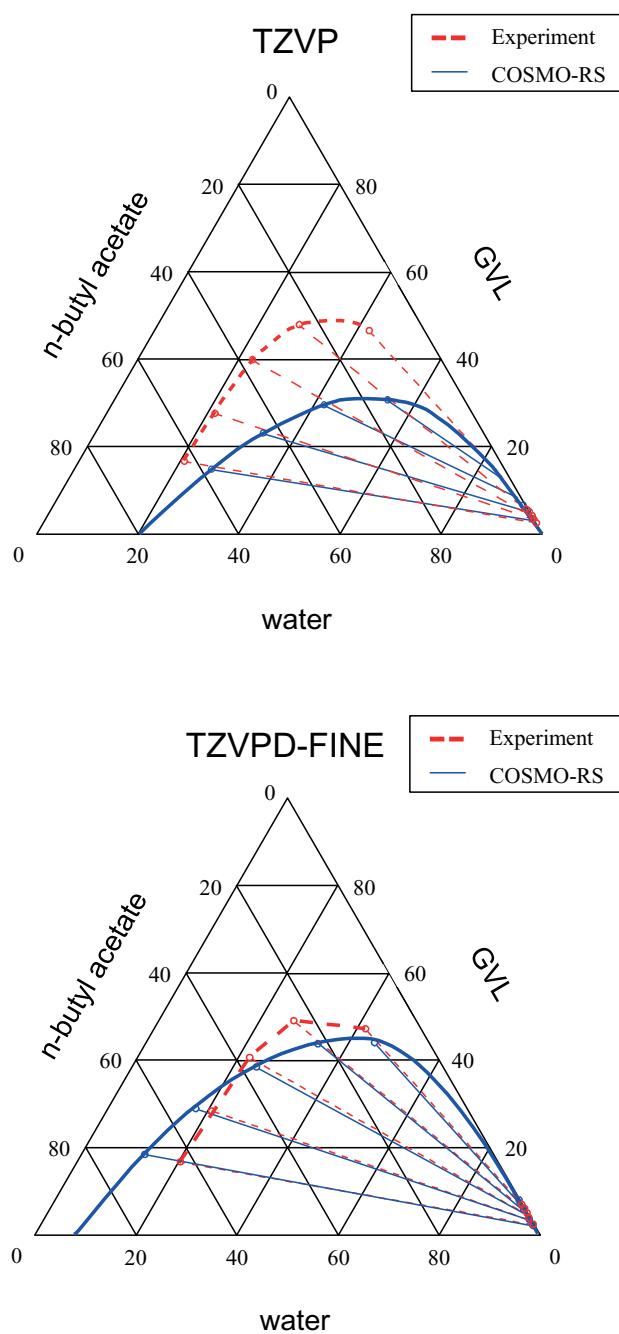


Figure C.2: Experimental validation for solvent n-butyl acetate with COSMO-RS predictions (solid blue) vs. experimental data (dashed red) for $T = 25\text{ }^{\circ}\text{C}$. Top: TZVP, Bottom: TZVPD-FINE.

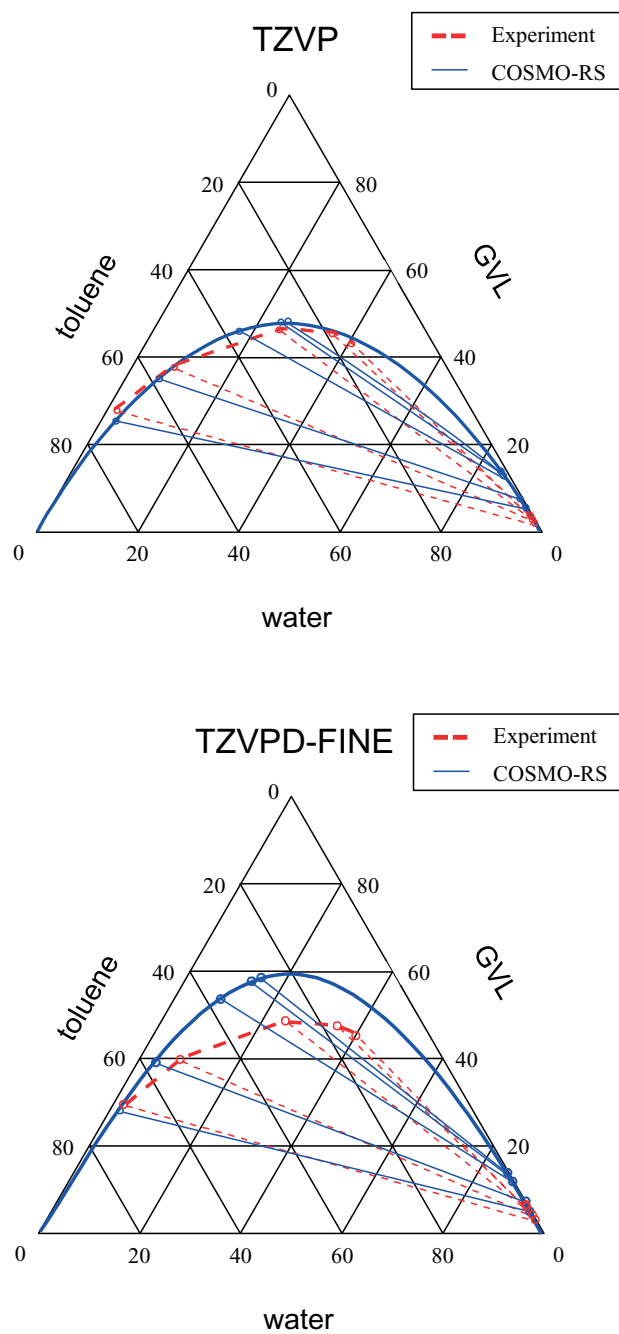


Figure C.3: Experimental validation for solvent toluene with COSMO-RS predictions (solid blue) vs. experimental data (dashed red) for $T = 25^\circ\text{C}$. Top: TZVP, Bottom: TZVPD-FINE.

The results from experimental liquid-liquid measurements used in this work are given in in Tables C.7 - C.9.

Table C.7: LLE measurements for the system 2-methylfuran(2MF)-GVL-water(H₂O) with molefractions and measurement uncertainties at 25 °C and 40 °C.

T/°C	Phase	z_{2MF}	z_{GVL}	z_{H_2O}	$u(z_{2MF})$	$u(z_{GVL})$	$u(z_{H_2O})$
25	ORG	0.77260	0.21355	0.01385	0.02867	0.00486	0.00259
		0.50915	0.44014	0.05071	0.03421	0.01410	0.00707
		0.27455	0.48941	0.23604	0.00982	0.00787	0.01018
		0.13597	0.45028	0.41376	0.00998	0.01284	0.02810
	WAT	0.09880	0.42575	0.47545	0.00465	0.00751	0.01713
		0.00120	0.01210	0.98670	0.00009	0.00054	0.00055
		0.00141	0.02108	0.97751	0.00010	0.00220	0.00215
		0.00174	0.03621	0.96205	0.00007	0.00164	0.00158
		0.00253	0.04641	0.95106	0.00004	0.00269	0.00256
		0.00319	0.05222	0.94459	0.00029	0.00286	0.00271
40	ORG	0.74571	0.24152	0.01277	0.01459	0.00360	0.00303
		0.45824	0.47047	0.07128	0.00961	0.00754	0.00894
		0.22304	0.45244	0.32452	0.00670	0.00723	0.01149
		0.10539	0.37558	0.51903	0.00826	0.01336	0.03381
	WAT	0.00146	0.01008	0.98846	0.00011	0.00061	0.00061
		0.00173	0.02005	0.97822	0.00009	0.00125	0.00122
		0.00238	0.02944	0.96818	0.00011	0.00139	0.00135
		0.00332	0.03482	0.96186	0.00005	0.00197	0.00189

Table C.8: LLE measurements for the system toluene(TOL)-GVL-water(H₂O) with molefractions and measurement uncertainties at 25 °C and 40 °C.

T/°C	Phase	z_{TOL}	z_{GVL}	$z_{\text{H}_2\text{O}}$	$u(z_{\text{TOL}})$	$u(z_{\text{GVL}})$	$u(z_{\text{H}_2\text{O}})$
25	ORG	0.80569	0.18196	0.01235	0.01009	0.01273	0.00023
		0.64384	0.30627	0.04989	0.00774	0.01256	0.00083
		0.36174	0.46608	0.17218	0.00389	0.01170	0.00269
		0.23200	0.50632	0.26168	0.00254	0.01220	0.00311
		0.17887	0.49810	0.32303	0.00145	0.00879	0.00333
		0.91565	0.08085	0.00350	0.01287	0.01410	0.00011
	WAT	0.88371	0.11169	0.00460	0.01229	0.01381	0.00226
		0.83528	0.15471	0.01001	0.01072	0.01299	0.00024
		0.00199	0.02230	0.97570	0.00143	0.00089	0.00164
		0.00118	0.03876	0.96005	0.00058	0.00081	0.00096
		0.00220	0.04998	0.94781	0.00062	0.00101	0.00112
		0.00181	0.06064	0.93754	0.00005	0.00177	0.00166
		0.00244	0.07091	0.92665	0.00021	0.00131	0.00122
		0.00169	0.00281	0.99551	0.00242	0.00028	0.00242
		0.00125	0.01574	0.98302	0.00082	0.00438	0.00438
		0.00056	0.01743	0.98202	0.00018	0.00058	0.00060
40	ORG	0.82014	0.16417	0.01569	0.01180	0.01454	0.00114
		0.64324	0.29089	0.06587	0.00745	0.01196	0.00175
		0.36013	0.43588	0.20398	0.00698	0.01978	0.00848
		0.22163	0.47274	0.30563	0.00344	0.01617	0.00751
		0.17329	0.46342	0.36329	0.00251	0.01536	0.00689
		0.90145	0.08360	0.01495	0.01215	0.01361	0.00230
		0.84803	0.13114	0.02083	0.01137	0.01322	0.00210
		0.77588	0.20757	0.01655	0.01217	0.01439	0.00441
	WAT	0.00084	0.02076	0.97840	0.00050	0.00062	0.00078
		0.00090	0.03314	0.96596	0.00023	0.00093	0.00093
		0.00181	0.04610	0.95209	0.00008	0.00185	0.00176
		0.00269	0.05982	0.93749	0.00014	0.00128	0.00120
		0.00309	0.06965	0.92726	0.00016	0.00321	0.00298
		0.00127	0.00359	0.99514	0.00033	0.00002	0.00033
		0.00119	0.01725	0.98156	0.00021	0.00033	0.00039
		0.00053	0.02194	0.97753	0.00009	0.00048	0.00048

Table C.9: LLE measurements for the system n-butyl acetate(nBAC)-GVL-water(H_2O) with molefractions and measurement uncertainties at 25 °C and 40 °C.

T/°C	Phase	z_{nBAC}	z_{GVL}	$z_{\text{H}_2\text{O}}$	$u(z_{\text{nBAC}})$	$u(z_{\text{GVL}})$	$u(z_{\text{H}_2\text{O}})$
25	ORG	0.61559	0.17204	0.21238	0.01141	0.01814	0.00631
		0.49997	0.28895	0.21108	0.01301	0.02673	0.00701
		0.36119	0.41873	0.22008	0.00665	0.01961	0.00551
		0.22874	0.50671	0.26456	0.00512	0.02478	0.00714
		0.09713	0.49307	0.40980	0.00223	0.02480	0.01141
	WAT	0.00129	0.01909	0.97961	0.00009	0.00045	0.00045
		0.00140	0.02896	0.96964	0.00002	0.00103	0.00100
		0.00202	0.03485	0.96313	0.00007	0.00212	0.00204
		0.00288	0.04399	0.95313	0.00008	0.00168	0.00160
		0.00523	0.06494	0.92983	0.00009	0.00155	0.00144
40	ORG	0.61339	0.18763	0.19898	0.01350	0.02204	0.00589
		0.48482	0.29811	0.21707	0.01531	0.03257	0.00847
		0.36444	0.38247	0.25310	0.00978	0.02829	0.00842
		0.22653	0.47608	0.29738	0.00451	0.02160	0.00747
		0.08731	0.46160	0.45109	0.00308	0.03728	0.01935
	WAT	0.00120	0.01815	0.98065	0.00005	0.00072	0.00071
		0.00145	0.02795	0.97061	0.00004	0.00099	0.00096
		0.00177	0.03774	0.96050	0.00003	0.00158	0.00151
		0.00307	0.04359	0.95334	0.00008	0.00252	0.00240
		0.00580	0.06588	0.92832	0.00015	0.00361	0.00335

Integrated Design of Reaction-Separation Processes

Process model equations

Multiphase reactor model

The multiphase reactor developed in this thesis is modeled in *MATLAB* using calculation routines implemented in the RBM (Bausa and Marquardt, 2000) for phase equilibrium calculations. This section summarizes the equations that have been implemented in the multiphase reactor model.

A schematic representation of a three-phase reactor with vapor phase V and two liquid phases LI and LII is depicted in Fig. D.1. Note: The notation in this appendix has been adjusted to the common notation of liquid phase compositions (x) and vapor phase compositions (y).

The mole balance of species i for a known reaction equation with stoichiometric coefficients ν_i reads

$$n_i^V + n_i^{LI} + n_i^{LII} = n_i^F + \nu_i \xi, \quad (\text{D.1})$$

where ξ represents the extent of reaction.

The reactor is assumed to operate at a constant temperature T and pressure p . Introducing mole fractions in the respective phases (Eq. (D.2)), equilibrium conditions can be added for each component. For a vapor-liquid equilibrium (VLE), the relation between the liquid mole fraction x_i^{LI} and the gaseous one y_i can be expressed with the

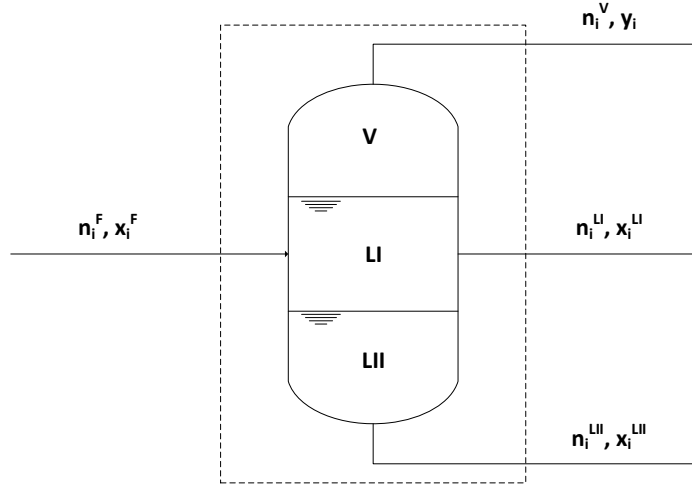


Figure D.1: Schematic representation of the VLLE reactor.

absolute pressure p , the activity coefficient of species i in phase LI γ_i^{LI} and its vapor pressure p_i^V (Eq. (D.3))¹

$$y_i = \frac{n_i^V}{\sum_i n_i^V}; \quad x_i^{LI} = \frac{n_i^{LI}}{\sum_i n_i^{LI}}; \quad x_i^{LII} = \frac{n_i^{LII}}{\sum_i n_i^{LII}}, \quad (\text{D.2})$$

$$y_i p = x_i^{LI} \gamma_i^{LI} p_i^V, \quad (\text{D.3})$$

$$x_i^{LI} \gamma_i^{LI} = x_i^{LII} \gamma_i^{LII}, \quad (\text{D.4})$$

where γ_i for a specific temperature and composition of the respective phase can be calculated using the NRTL-model (Renon and Prausnitz (1968)) and p_i^V with Antoine's equation (cf. Appendix A).

Finally, the condition for chemical equilibrium is added for the phase, where the reaction takes place. The specific equation for a liquid-phase (in phase LI) reaction reads

¹Pointing factors and fugacity coefficients will not be considered here. In case of liquid-liquid equilibrium (LLE), Eq. (D.4) has to hold.

$$\prod_i (x_i^{LI} \gamma_i^{LI})^{\nu_i} = \left[\prod_i \left(\frac{p^0}{p_i^V} \right)^{\nu_i} \right] e^{\frac{-\Delta G_R^{ig}}{RT}} \quad (\text{D.5})$$

where ΔG_R^{ig} denotes the Gibbs free enthalpy of reaction in the ideal gas reference state at temperature T and pressure p^0 , which can be obtained from tabulated experimental data or calculated with QM methods.

A suitable approach to solve the model is to guess an initial extent of reaction and solve the consecutive three-phase flash problem with given feed, temperature and pressure². If the chemical equilibrium is not reached, the extent of reaction is iterated until this condition is fulfilled. For a given conversion of a target compound, the multiphase reactor model reduces to a phase-split problem. The multiphase reactor model can be extended to more than three phases by adding additional phase equilibrium constraints.

The agreement between the results obtained with the multiphase reactor model and those from an Aspen Plus simulations (V8.4) using the built-in *RCSTR* model (with specified reaction equations and reaction equilibrium data from Jens et al. (2016)) is given in Fig.D.2. Two three-phase reactions are evaluated. The respective reaction equations read



with (R1) evaluated at 100 bar and (R2) evaluated at 1 bar.

For the molar concentration of the storage molecule (SM) in the extract phase ($x_{\text{Extr}}^{\text{SM}}$) of the multiphase reactor, a coefficient of determination of $R^2 = 0.975$ and a root-mean-square error of $RMSE = 0.0068$ are obtained in comparison to Aspen Plus simulations, indicating a very good agreement between the models. Fig.D.2 shows that results for $x_{\text{Extr}}^{\text{SM}}$ between the models increasingly deviate for larger molar concentration of the storage molecule $x_{\text{Extr}}^{\text{SM}}$. The reason for this deviation are differences in calculated multiphase equilibrium compositions between solutions from Aspen Plus and the RBM, which uses a calculation procedure established by Bausa and Marquardt (2000). Thus, deviations for $x_{\text{Extr}}^{\text{SM}}$ between Aspen Plus and the multiphase reactor model in this thesis are currently inherent to the available implementation of the RBM and should be addressed in future work.

²A solver for three-phase flash problems is available within the implementation of the RBM.

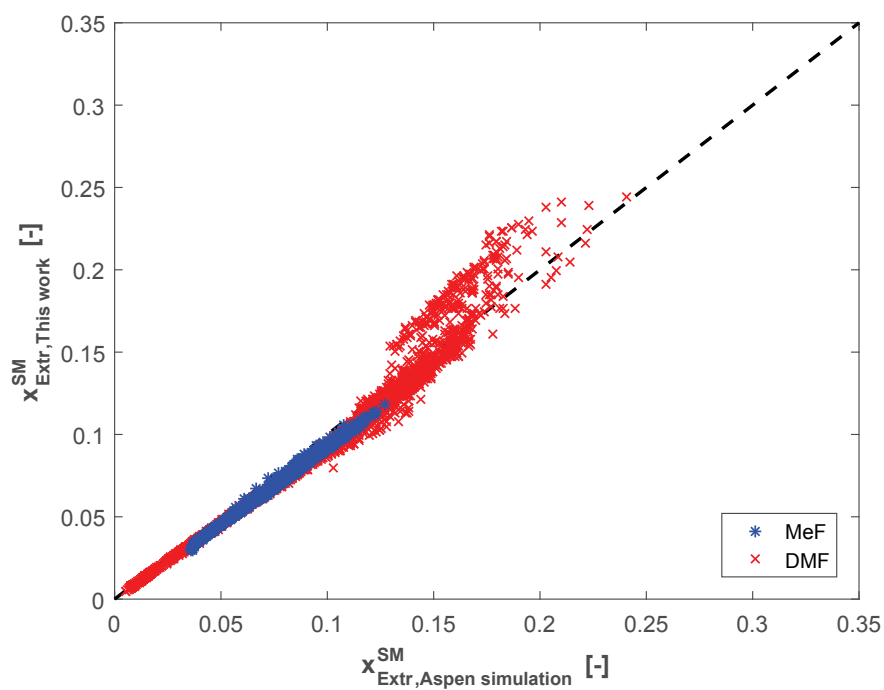


Figure D.2: Comparison of the multiphase reactor model to commercial process simulations. 4,731 reaction systems are evaluated (MeF (R1): 1,962 and DMF (R2): 2,769). R1: blue asterisks, R2: red marks.

Compressor model

The compressor in this work is modeled in the limiting case of ideal gas compression with infinite multistage compressor stages and intermediate cooling. These assumptions lead to the equation for the work W required for isothermal compression from pressure p_1 to p_2 of a gaseous stream n at temperature T :

$$W = nRT \ln\left(\frac{p_2}{p_1}\right). \quad (\text{D.6})$$

CO Process: flow sheets and specifications

An overview of all possible process structures is given in Fig.D.3. In the following, assumptions for process simulations are listed.

Exergy demand: The exergy demand E is estimated by the sum of the exergy demand for compression E_{comp} , the exergy demand for vaporizing the reforming feed E_{vap} and the netto exergy demand of the distillation columns (sum of boiling E_{boil} and condensing E_{cond}) (Eq. (D.7)). All other exergy streams are assumed to be small (heating and cooling) or are constant for all processes (exergy of material streams).

$$E = E_{\text{comp}} + E_{\text{vap}} + E_{\text{boil}} + E_{\text{cond}} \quad (\text{D.7})$$

where the exergy of a work stream W is

$$E_W = W \quad (\text{D.8})$$

and the exergy of a heat streams Q at temperature T and environment temperature 298.15 K is

$$E_Q = Q \left(1 - \frac{298.15\text{K}}{T}\right) \quad (\text{D.9})$$

Input flow: H_2 , CO_2 , the AR and the solvent are fed to the process in equal molar amounts. The amount of water added is optimized. A phase split into an aqueous phase (catalyst phase) and an organic phase (product or extract phase) must be ensured.

Synthesis reactor: The reactor operates at 25 °C. The pressure is optimized with respect to the exergy demand. Gibbs free enthalpies of reaction ΔG_R^{ig} are taken from Jens et al. (2016) (24.900 J mol⁻¹ for the synthesis of MeF and 9.920 J mol⁻¹ for the synthesis of DMF). It is assumed that H₂ and CO₂ are completely removed from the reactor outlet by flashing to ambient pressure.

Extraction column: The temperature of the extraction column is set to 25 °C. Pure solvent is used. The storage molecule is completely extracted from the reactor outlet.

Reforming reactor: It is assumed that a full conversion of the storage molecule can be achieved at a sufficiently high temperature. The selectivity is assumed to be 100% (Supronowicz et al., 2015). All CO is released in 100% purity by cooling the reactor outlet.

Distillation columns: All entering and leaving streams are assumed to be liquid. Sharp splits are applied, such that a component is either recovered completely in the distillate or at the bottom. Two components forming an heteroazeotrope can be separated by using two columns and a decanter. If a homogeneous azeotrope appears between two or more components present in one column, the distillation is considered infeasible.

Process structure: The overall process structure depends on the chosen process variant (PB or PA) and the order of the boiling points. Simple heuristics are used for the design of the reaction-separation sequence described in Chapter 6. Fig. D.3 shows the resulting process flowsheet structure.

Details on the COSMO-CAMPD optimization

Calculations have been performed using 8 parallelized cores. Tables D.1 and D.2 list statistics for all optimization runs. It should be emphasized that the time demand strongly correlates with the number of COSMOconf calculations.

The comprehensive fragment library employed in the optimization runs for the reaction-separation process case study are given in Fig. D.4.

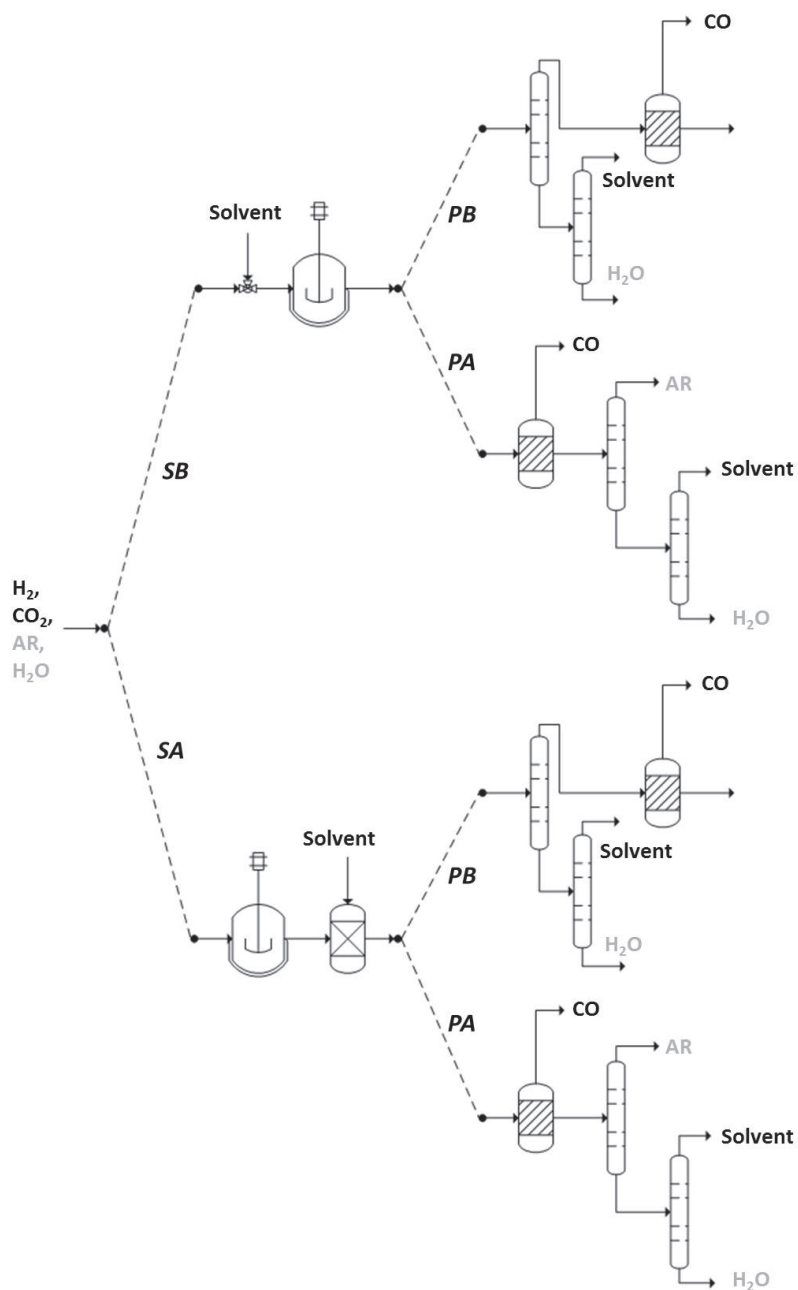


Figure D.3: Process flowsheet structure for the conversion of CO_2 to CO (Jens et al. (2016)). Abbreviations: Solvent before (SB)/after (SA) synthesis, purification before (PB)/after (PA) reforming, additional reactant (AR).

Table D.1: Statistics for the COSMO-CAMD optimization for DMF. The first four rows add up to the total amount of 2,040 molecules. If a COSMO file is available, it is taken from the database, otherwise it is created by COSMOconf. If a fragmentation failure occurs or the molecule contains more than 12 bold atoms (others than H) or undesired functional groups, no COSMO file is created.

	DMF run 1	DMF run 2	DMF run 3
COSMOconf	624	602	308
Database	960	1,174	1,482
Fragmentation failure	59	12	7
Not built	397	252	243
Time (d:h:m:s)	2:22:58:36	2:19:04:24	2:00:26:43

Table D.2: Statistics for the COSMO-CAMD optimization for MeF. The first four rows add up to the total amount of 2,040 molecules. If a COSMO file is available, it is taken from the database, otherwise it is created by COSMOconf. If a fragmentation failure occurs or the molecule contains more than 12 bold atoms (others than H) or undesired functional groups, no COSMO file is created.

	MeF run 1	MeF run 2	MeF run 3
COSMOconf	519	471	452
Database	1,144	1,094	1,199
Fragmentation failure	3	2	2
Not built	374	473	387
Time (d:h:m:s)	2:07:12:10	2:10:27:15	3:04:09:29

Bibliography

- ACS Green Chemistry Institute (2016). ACS GCI pharmaceutical roundtable solvent selection guide, <http://www.acs.org/content/acs/en/greenchemistry/industry-business/pharmaceutical.html>.
- Adjiman, C. S., Galindo, A., and Jackson, G. (2014). Molecules Matter: The Expanding Envelope of Process Design. In *M. R. Eden (Ed.), Proceedings of the 8th International Conference on Foundations of Computer-Aided Process Design*, volume 34 of *Computer-Aided Chemical Engineering*, pages 55–64. Elsevier, Amsterdam.
- Alonso, D. M., Gallo, J. M. R., Mellmer, M. A., Wettstein, S. G., and Dumesic, J. A. (2013a). Direct conversion of cellulose to levulinic acid and gamma-valerolactone using solid acid catalysts. *Catalysis Science & Technology*, 3(4):927–931.
- Alonso, D. M., Wettstein, S. G., and Dumesic, J. A. (2013b). Gamma-valerolactone, a sustainable platform molecule derived from lignocellulosic biomass. *Green Chemistry*, 15(3):584–595.
- Anastas, P. T. and Zimmerman, J. B. (2016). Safer by design. *Green Chemistry*, 18(16):4324–4324.
- Ashcroft, C. P., Dunn, P. J., Hayler, J. D., and Wells, A. S. (2015). Survey of Solvent Usage in Papers Published in Organic Process Research & Development 1997–2012. *Organic Process Research & Development*, 19(7):740–747.
- Austin, N. D., Sahinidis, N. V., Konstantinov, I. A., and Trahan, D. W. (2018). COSMO-based computer-aided molecular/mixture design: A focus on reaction solvents. *AIChE Journal*, 64(1):104–122.
- Austin, N. D., Sahinidis, N. V., and Trahan, D. W. (2016a). A COSMO-based approach to computer-aided mixture design. *Chemical Engineering Science*, 159:93–105.

- Austin, N. D., Sahinidis, N. V., and Trahan, D. W. (2016b). Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chemical Engineering Research and Design*, 116:2–26.
- Bai, P., Jeon, M. Y., Ren, L., Knight, C., Deem, M. W., Tsapatsis, M., and Siepmann, J. I. (2015). Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling. *Nature Communications*, 6:5912.
- Balasubramanian, K. (1985). Applications of combinatorics and graph theory to spectroscopy and quantum chemistry. *Chemical Reviews*, 85(6):599–618.
- Bardow, A., Steur, K., and Gross, J. (2010). Continuous-Molecular Targeting for Integrated Solvent and Process Design. *Industrial & Engineering Chemistry Research*, 49(6):2834–2840.
- Barnicki, S. D. and Fair, J. R. (1990). Separation system synthesis: A knowledge-based approach. 1. Liquid mixture separations. *Industrial & Engineering Chemistry Research*, 29(3):421–432.
- Barnicki, S. D. and Fair, J. R. (1992). Separation system synthesis: A knowledge-based approach. 2. Gas/vapor mixtures. *Industrial & Engineering Chemistry Research*, 31(7):1679–1694.
- Bausa, J. and Marquardt, W. (2000). Quick and reliable phase stability test in VLLE flash calculations by homotopy continuation. *Computers & Chemical Engineering*, 24(11):2447–2456.
- Bausa, J., Watzdorf, R. v., and Marquardt, W. (1998). Shortcut methods for nonideal multicomponent distillation: I. Simple columns. *AIChE Journal*, 44(10):2181–2198.
- Behr, A., Ebbinghaus, P., and Naendrup, F. (2004). Process concepts for the transition metal catalyzed syntheses of formic acid and dimethylformamide based on carbon dioxide. *Chemical Engineering & Technology*, 27(5):495–501.
- Bemis, G. W. and Murcko, M. A. (1996). The properties of known drugs. 1. Molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893.
- Blumenthal, L. C., Jens, C. M., Ulbrich, J., Schwering, F., Langrehr, V., Turek, T., Kunz, U., Leonhard, K., and Palkovits, R. (2016). Systematic Identification of Solvents Optimal for the Extraction of 5-Hydroxymethylfurfural from Aqueous Reactive Solutions. *ACS Sustainable Chemistry & Engineering*, 4(1):228–235.

- Bommareddy, S., Chemmangattuvalappil, N. G., Solvason, C. C., and Eden, M. R. (2010). Simultaneous solution of process and molecular design problems using an algebraic approach. *Computers & Chemical Engineering*, 34(9):1481–1486.
- Bonet, J., Plesu, V., Bonet-Ruiz, A.-E., Tuluc, A., Iancu, P., Tohaneanu, M. C., and Llorens, J. (2015). Fast solvent screening for counter-current liquid–liquid extraction columns. *Clean Technologies and Environmental Policy*, 17(5):1227–1238.
- Braun, E., Zurhelle, A. F., Thijssen, W., Schnell, S. K., Lin, L.-C., Kim, J., Thompson, J. A., and Smit, B. (2016). High-throughput computational screening of nanoporous adsorbents for CO₂ capture from natural gas. *Molecular Systems Design & Engineering*, 1(2):175–188.
- Brittain, J. M., de la Mare, Peter B. D., and Newman, P. A. (1981). Electrophilic substitution with rearrangement. 9. Dienones derived from brominations of o-, m-, and p-cresol. *Journal of the Chemical Society, Perkin Transactions*, 2:32–41.
- Broeren, M. L., Zijp, M. C., Waaijers-van der Loop, S. L., Heugens, E. H., Posthuma, L., Worrell, E., and Shen, L. (2017). Environmental assessment of bio-based chemicals in early-stage development: A review of methods and indicators. *Biofuels, Bioproducts and Biorefining*, 11(4):701–718.
- Brüggemann, S. and Marquardt, W. (2004). Shortcut methods for nonideal multicomponent distillation: 3. Extractive distillation columns. *AIChE Journal*, 50(6):1129–1149.
- Brüggemann, S. and Marquardt, W. (2011). Conceptual design of distillation processes for mixtures with distillation boundaries. 1. Computational assessment of split feasibility. *AIChE Journal*, 57(6):1526–1539.
- Bubnov, Y. N., Tsyban, A. V., and Mikhailov, B. M. (1980). A Convenient Synthesis of 1,4-Pentenyne (Allylacetylenes) via Allylboranes. *Synthesis*, 11:904–905.
- Burger, J., Papaioannou, V., Gopinath, S., Jackson, G., Galindo, A., and Adjiman, C. S. (2015). A hierarchical method to integrated solvent and process design of physical CO₂ absorption using the SAFT- γ Mie approach. *AIChE Journal*, 61(10):3249–3269.
- Burghoff, B., Goetheer, Earl L. V., and Haan, Andre B. de (2008). COSMO-RS-Based Extractant Screening for Phenol Extraction As Model System. *Industrial & Engineering Chemistry Research*, 47(12):4263–4269.

- Cancès, E., Defranceschi, M., Kutzelnigg, W., Le Bris, C., and Maday, Y. (2003). Computational quantum chemistry: A primer. In *Special Volume, Computational Chemistry*, volume 10 of *Handbook of Numerical Analysis*, pages 3–270. Elsevier.
- Capello, C., Fischer, U., and Hungerbühler, K. (2007). What is a green solvent? A comprehensive framework for the environmental assessment of solvents. *Green Chemistry*, 9(9):927–934.
- Cedergreen, N., Dalhoff, K., Li, D., Gottardi, M., and Kretschmann, A. C. (2017). Can toxicokinetic and toxicodynamic modeling be used to understand and predict synergistic interactions between chemicals? *Environmental science & technology*, 51(24):14379–14389.
- Chapman, W. G., Gubbins, K. E., Jackson, G., and Radosz, M. (1989). SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilibria*, 52:31–38.
- Chavali, S., Lin, B., Miller, D. C., and Camarda, K. V. (2004). Environmentally-benign transition metal catalyst design using optimization techniques. *Computers & Chemical Engineering*, 28(5):605–611.
- Chemangattuvalappil, N. G. and Eden, M. R. (2013). A Novel Methodology for Property-Based Molecular Design Using Multiple Topological Indices. *Industrial & Engineering Chemistry Research*, 52(22):7090–7103.
- Chemangattuvalappil, N. G., Solvason, C. C., Bommarreddy, S., and Eden, M. R. (2010a). Combined property clustering and GC+ techniques for process and product design. *Computers & Chemical Engineering*, 34(5):582–591.
- Chemangattuvalappil, N. G., Solvason, C. C., Bommarreddy, S., and Eden, M. R. (2010b). Reverse problem formulation approach to molecular design using property operators based on signature descriptors. *Computers & Chemical Engineering*, 34(12):2062–2071.
- Chen, B., Lei, Z., Li, Q., and Li, C. (2005). Application of CAMD in separating hydrocarbons by extractive distillation. *AIChE Journal*, 51(12):3114–3121.
- Churi, N. and Achenie, L. E. K. (1996). Novel Mathematical Programming Model for Computer Aided Molecular Design. *Industrial & Engineering Chemistry Research*, 35(10):3788–3794.

- Conte, E., Gani, R., and Ng, K. M. (2011). Design of formulated products: A systematic methodology. *AIChE Journal*, 57(9):2431–2449.
- COSMOlogic (2015a). COSMObase 1501. *COSMOlogic, Leverkusen, Germany*.
- COSMOlogic (2015b). COSMOconf 4.0. *COSMOlogic, Leverkusen, Germany*.
- COSMOlogic (2015c). COSMOquick 1.5. *COSMOlogic, Leverkusen, Germany*.
- COSMOlogic (2015d). COSMOtherm C30_1501. *COSMOlogic, Leverkusen, Germany*.
- Cremaschi, S. (2015). A perspective on process synthesis: Challenges and prospects. *Computers & Chemical Engineering*, 81:130–137.
- Dechambre, D., Pauls, C., Greiner, L., Leonhard, K., and Bardow, A. (2014a). Towards automated characterisation of liquid–liquid equilibria. *Fluid Phase Equilibria*, 362:328–334.
- Dechambre, D., Wolff, L., Pauls, C., and Bardow, A. (2014b). Optimal experimental design for the characterization of liquid–liquid equilibria. *Industrial & Engineering Chemistry Research*, 53(50):19620–19627.
- Deglmann, P., Hungenberg, K.-D., and Vale, H. M. (2017). Dependence of Propagation Rate Coefficients in Radical Polymerization on Solution Properties. *Macromolecular Reaction Engineering*, 11(1).
- Deglmann, P., Schäfer, A., and Lennartz, C. (2015). Application of quantum calculations in the chemical industry-An overview. *International Journal of Quantum Chemistry*, 115(3):107–136.
- Delidovich, I., Leonhard, K., and Palkovits, R. (2014). Cellulose and hemicellulose valorisation: an integrated challenge of catalysis and reaction engineering. *Energy & Environmental Science*, 7(9):2803–2830.
- Diorazio, L. J., Hose, D. R. J., and Adlington, N. K. (2016). Toward a More Holistic Framework for Solvent Selection. *Organic Process Research & Development*, 20(4):760–773.
- Douguet, D., Munier-Lehmann, H., Labesse, G., and Pochet, S. (2005). LEA3D: a computer-aided ligand design for structure-based drug design. *Journal of Medicinal Chemistry*, 48(7):2457–2468.

- Douguet, D., Thoreau, E., and Grassy, G. (2000). A genetic algorithm for the automated generation of small organic molecules: Drug design using an evolutionary algorithm. *Journal of Computer-aided Molecular Design*, 14(5):449–466.
- Duran, M. A. and Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36(3):307–339.
- Durvedi, A. P. and Achenie, L. E. (1996). Designing environmentally safe refrigerants using mathematical programming. *Chemical Engineering Science*, 51(15):3727–3739.
- Eckert, F. and Klamt, A. (2002). Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal*, 48(2):369–385.
- Eden, M., Jørgensen, S., Gani, R., and El-Halwagi, M. (2004). A novel framework for simultaneous separation process and product design. *Chemical Engineering and Processing: Process Intensification*, 43(5):595–608.
- Eljack, F. T., Eden, M. R., Kazantzi, V., Qin, X., and El-Halwagi, M. M. (2007). Simultaneous process and molecular design—A property based approach. *AIChE Journal*, 53(5):1232–1239.
- Eslick, J. C., Ye, Q., Park, J., Topp, E. M., Spencer, P., and Camarda, K. V. (2009). A computational molecular design framework for crosslinked polymer networks. *Computers & Chemical Engineering*, 33(5):954–963.
- Everett, J. L. and Kon, G. A. R. (1950). 610. the preparation of some cytotoxic epoxides. *Journal of the Chemical Society (Resumed)*, 1(1):3131–3135.
- Fang, J., Zhao, R., Su, W., Li, C., Liu, J., and Li, B. (2016). A molecular design method based on the COSMO-SAC model for solvent selection in ionic liquid extractive distillation. *AIChE Journal*, 62(8):2853–2869.
- Feng, L.-C., Chou, C.-H., Tang, M., and Chen, Y.-P. (1998). Vapor–liquid equilibria of binary mixtures 2-butanol + butyl acetate, hexane + butyl acetate, and cyclohexane + 2-butanol at 101.3 kpa. *Journal of Chemical & Engineering Data*, 43(4):658–661.
- Ferro, V. R., de Riva, J., Sanchez, D., Ruiz, E., and Palomar, J. (2015). Conceptual design of unit operations to separate aromatic hydrocarbons from naphtha using ionic liquids. COSMO-based process simulations with multi-component “real” mixture feed. *Chemical Engineering Research and Design*, 94:632–647.

- Ferro, V. R., Ruiz, E., Tobajas, M., and Palomar, J. F. (2012). Integration of COSMO-based methodologies into commercial process simulators: Separation and purification of reuterin. *AIChE Journal*, 58(11):3404–3415.
- Fidkowski, Z. T., Malone, M. F., and Doherty, M. F. (1993). Computing azeotropes in multicomponent mixtures. *Computers & Chemical Engineering*, 17(12):1141–1155.
- Fink, T., Bruggesser, H., and Reymond, J.-L. (2005). Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angewandte Chemie (International ed. in English)*, 44(10):1504–1508.
- Folić, M., Adjiman, C. S., and Pistikopoulos, E. N. (2007). Design of solvents for optimal reaction rate constants. *AIChE Journal*, 53(5):1240–1256.
- Folić, M., Adjiman, C. S., and Pistikopoulos, E. N. (2008). Computer-Aided Solvent Design for Reactions: Maximizing Product Formation. *Industrial & Engineering Chemistry Research*, 47(15):5190–5202.
- Fouskakis, D. and Draper, D. (2002). Stochastic optimization: A review. *International Statistical Review*, 70(3):315–349.
- Franceschini, G. and Macchietto, S. (2007). Validation of a model for biodiesel production through model-based experiment design. *Industrial & Engineering Chemistry Research*, 46(1):220–232.
- Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19):4846–4872.
- Fredenslund, A., Jones, R. L., and Prausnitz, J. M. (1975). Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21(6):1086–1099.
- Freire, M. G., Santos, L. M., Marrucho, I. M., and Coutinho, J. A. (2007). Evaluation of COSMO-RS for the prediction of LLE and VLE of alcohols+ionic liquids. *Fluid Phase Equilibria*, 255(2):167–178.
- Gani, R. (2004a). Chemical product design: Challenges and opportunities. *Computers & Chemical Engineering*, 28(12):2441–2457.
- Gani, R. (2004b). Computer-Aided Methods and Tools for Chemical Product Design. *Chemical Engineering Research and Design*, 82(11):1494–1504.

- Gani, R. and Brignole, E. A. (1983). Molecular Design of Solvents for Liquid Extraction based on UNIFAC. *Fluid Phase Equilibria*, 13:331–340.
- Gani, R., Nielsen, B., and Fredenslund, A. (1991). A group contribution approach to computer-aided molecular design. *AIChE Journal*, 37(9):1318–1332.
- Garcia-Chavez, L. Y., Hermans, A. J., Schuur, B., and de Haan, André B. (2012). COSMO-RS assisted solvent screening for liquid–liquid extraction of mono ethylene glycol from aqueous streams. *Separation and Purification Technology*, 97:2–10.
- Gatti, M., Martelli, E., Marechal, F., and Consonni, S. (2014). Review, modeling, heat integration, and improved schemes of rectisol®-based processes for CO₂ capture. *Applied Thermal Engineering*, 70(2):1123–1140.
- Gebreslassie, B. H. and Diwekar, U. M. (2015). Efficient ant colony optimization for computer aided molecular design: Case study solvent selection problem. *Computers & Chemical Engineering*, 78:1–9.
- Ghanem, O. B., Mutalib, M. I. A., Leveque, J.-M., and El-Harbawi, M. (2017). Development of QSAR model to predict the ecotoxicity of *Vibrio fischeri* using COSMO-RS descriptors. *Chemosphere*, 170:242–250.
- Gilbertson, L. M., Zimmerman, J. B., Plata, D. L., Hutchison, J. E., and Anastas, P. T. (2015). Designing nanomaterials to maximize performance and minimize undesirable implications guided by the principles of green chemistry. *Chemical Society Reviews*, 44(16):5758–5777.
- Gmehling, J. (2003). Potential of thermodynamic tools (group contribution methods, factual data banks) for the development of chemical processes. *Fluid Phase Equilibria*, 210(2):161–173.
- Gmehling, J. (2009). Present status and potential of group contribution methods for process development. *The Journal of Chemical Thermodynamics*, 41(6):731–747.
- Gmehling, J., Constantinescu, D., and Schmid, B. (2015). Group Contribution Methods for Phase Equilibrium Calculations. *Annual Review of Chemical and Biomolecular Engineering*, 6:267–292.
- Gmehling, J. and Schedemann, A. (2014). Selection of Solvents or Solvent Mixtures for Liquid–Liquid Extraction Using Predictive Thermodynamic Models or Access to the Dortmund Data Bank. *Industrial & Engineering Chemistry Research*, 53(45):17794–17805.

- Gmehling, J., Wittig, R., Lohmann, J., and Joh, R. (2002). A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Industrial & Engineering Chemistry Research*, 41(6):1678–1688.
- Gontrani, L., Ramondo, F., and Caminiti, R. (2006). Furan and thiophene in liquid phase: An X-ray and molecular dynamics study. *Chemical Physics Letters*, 422(1-3):256–261.
- Gopinath, S., Jackson, G., Galindo, A., and Adjiman, C. S. (2016). Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. *AIChE Journal*, 62(9):3484–3504.
- Harper, P. M. and Gani, R. (2000). A multi-step and multi-level approach for computer aided molecular design. *Computers & Chemical Engineering*, 24(2-7):677–683.
- Harper, P. M., Gani, R., Kolar, P., and Ishikawa, T. (1999). Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilibria*, 158-160:337–347.
- Heger, S., Bluhm, K., Agler, M. T., Maletz, S., Schäffer, A., Seiler, T.-B., Angenent, L. T., and Hollert, H. (2012). Biotests for hazard assessment of biofuel fermentation. *Energy & Environmental Science*, 5(12):9778–9788.
- Heilbron, I. M., Jones, E. R. H., Smith, P., and Weedon, B. C. L. (1946). 16. Researches on acetylenic compounds. IV. The hydration of some acetylenylcarbinols derived from $\alpha\beta$ -unsaturated aldehydes. *Journal of the Chemical Society (Re-sumed)*, 1(1):54–58.
- Henderson, R. K., Jiménez-González, C., Constable, D. J. C., Alston, S. R., Inglis, G. G. A., Fisher, G., Sherwood, J., Binks, S. P., and Curzons, A. D. (2011). Expanding GSK’s solvent selection guide – embedding sustainability into solvent selection starting at medicinal chemistry. *Green Chemistry*, 13(4):854–862.
- Hohenberg, P. and Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871.
- Holland, J. H. (1973). Genetic Algorithms and the Optimal Allocation of Trials. *SIAM Journal on Computing*, 2(2):88–105.
- Hostrup, M., Harper, P. M., and Gani, R. (1999). Design of environmentally benign processes: integration of solvent design and separation process synthesis. *Computers & Chemical Engineering*, 23(10):1395–1414.

- Houben, C. and Lapkin, A. A. (2015). Automatic discovery and optimization of chemical processes. *Current Opinion in Chemical Engineering*, 9:1–7.
- Huck, J. M., Lin, L.-C., Berger, A. H., Shahrak, M. N., Martin, R. L., Bhowan, A. S., Haranczyk, M., Reuter, K., and Smit, B. (2014). Evaluating different classes of porous materials for carbon capture. *Energy & Environmental Science*, 7(12):4132–4146.
- Jacob, P.-M., Yamin, P., Perez-Storey, C., Hopgood, M., and Lapkin, A. A. (2017). Towards automation of chemical process route selection based on data mining. *Green Chemistry*, 19(1):140–152.
- Jakob, A., Grensemann, H., Lohmann, J., and Gmehling, J. (2006). Further Development of Modified UNIFAC (Dortmund): Revision and Extension 5. *Industrial & Engineering Chemistry Research*, 45(23):7924–7933.
- Jaksland, C. A., Gani, R., and Lien, K. M. (1995). Separation process design and synthesis based on thermodynamic insights. *Chemical Engineering Science*, 50(3):511–530.
- Jasperson, L. V., McDougal, R. J., Diky, V., Paulechka, E., Chirico, R. D., Kroenlein, K., Iisa, K., and Dutta, A. (2017). Liquid–Liquid Equilibrium Measurements for Model Systems Related to Catalytic Fast Pyrolysis of Biomass. *Journal of Chemical & Engineering Data*, 62(1):243–252.
- Jens, C. M., Müller, L., Leonhard, K., and Bardow, A. (2018). To integrate or not to integrate – techno-economic and life cycle assessment of CO₂ capture and conversion to methyl formate using methanol. *Submitted to ACS Sustainable Chemistry & Engineering*.
- Jens, C. M., Nowakowski, K., Scheffczyk, J., Leonhard, K., and Bardow, A. (2016). CO from CO₂ and fluctuating renewable energy via formic-acid derivatives. *Green Chemistry*, 18(20):5621–5629.
- Joback, K. G. and Stephanopoulos, G. (1995). Searching Spaces of Discrete Solutions: The Design of Molecules Possessing Desired Physical Properties. *Advances in Chemical Engineering*, 21:257–311.
- Jonuzaj, S., Akula, P. T., Kleniati, P.-M., and Adjiman, C. S. (2016). The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study. *AIChE Journal*, 62(5):1616–1633.

- Joos, L., Lejaeghere, K., Huck, J. M., van Speybroeck, V., and Smit, B. (2015). Carbon capture turned upside down: High-temperature adsorption & low-temperature desorption (HALD). *Energy & Environmental Science*, 8(8):2480–2491.
- Kaminski, S., Kirgios, E., Bardow, A., and Leonhard, K. (2017). Improved Property Predictions by Combination of Predictive Models. *Industrial & Engineering Chemistry Research*, 56(11):3098–3106.
- Karunanithi, A. T., Achenie, L. E., and Gani, R. (2006). A computer-aided molecular design framework for crystallization solvent design. *Chemical Engineering Science*, 61(4):1247–1260.
- Karunanithi, A. T., Achenie, L. E. K., and Gani, R. (2005). A New Decomposition-Based Computer-Aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures. *Industrial & Engineering Chemistry Research*, 44(13):4785–4797.
- Kiezyk, P. R. and Mackay, D. (1973). The screening and selection of solvents for the extraction of phenol from water. *The Canadian Journal of Chemical Engineering*, 51(6):741–745.
- Kim, K.-J. and Diwekar, U. M. (2002). Hammersley stochastic annealing: efficiency improvement for combinatorial optimization under uncertainty. *IIE Transactions*, 34(9):761–777.
- Kim, S. and Overcash, M. (2003). Energy in chemical manufacturing processes: Gate-to-gate information for life cycle assessment. *Journal of Chemical Technology & Biotechnology*, 78(9):995–1005.
- Klamt, A. (1995). Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry*, 99(7):2224–2235.
- Klamt, A., Eckert, F., and Arlt, W. (2010). COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annual Review of Chemical and Biomolecular Engineering*, 1:101–122.
- Klamt, A., Eckert, F., and Hornig, M. (2001). COSMO-RS: A novel view to physiological solvation and partition questions. *Journal of Computer-aided Molecular Design*, 15(4):355–365.

- Klamt, A., Eckert, F., and Hornig, M. (2005). COSMOfrag: a novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry. *Journal of Chemical Information and Modeling*, 45(5):1169–1177.
- Klamt, A., Jonas, V., Bürger, T., and Lohrenz, John C. W. (1998). Refinement and Parametrization of COSMO-RS. *The Journal of Physical Chemistry A*, 102(26):5074–5085.
- Kohn, W., Becke, A. D., and Parr, R. G. (1996). Density functional theory of electronic structure. *The Journal of Physical Chemistry*, 100(31):12974–12980.
- Kohn, W. and Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138.
- Kontogeorgis, G. M., Michelsen, M. L., Folas, G. K., Derawi, S., von Solms, N., and Stenby, E. H. (2006). Ten Years with the CPA (Cubic-Plus-Association) Equation of State. 1. Pure Compounds and Self-Associating Systems. *Industrial & Engineering Chemistry Research*, 45(14):4855–4868.
- Kossack, S., Krämer, K., Gani, R., and Marquardt, W. (2008). A systematic synthesis framework for extractive distillation processes. *Chemical Engineering Research and Design*, 86(7):781–792.
- Kostal, J., Voutchkova-Kostal, A., Anastas, P. T., and Zimmerman, J. B. (2015). Identifying and designing chemicals with minimal acute aquatic toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 112(20):6289–6294.
- Krämer, K., Harwardt, A., Bronneberg, R., and Marquardt, W. (2011a). Separation of butanol from acetone–butanol–ethanol fermentation by a hybrid extraction–distillation process. *Computers & Chemical Engineering*, 35(5):949–963.
- Krämer, K., Harwardt, A., Skiborowski, M., Mitra, S., and Marquardt, W. (2011b). Shortcut-based design of multicomponent heteroazeotropic distillation. *Chemical Engineering Research and Design*, 89(8):1168–1189.
- Kröger, L. C., Kopp, W. A., and Leonhard, K. (2017). Prediction of Chain Propagation Rate Constants of Polymerization Reactions in Aqueous NIPAM/BIS and VCL/BIS Systems. *The Journal of Physical Chemistry. B*, 121(13):2887–2895.
- Lampe, M., Stavrou, M., Bucker, H. M., Gross, J., and Bardow, A. (2014). Simultaneous Optimization of Working Fluid and Process for Organic Rankine Cycles Using PC-SAFT. *Industrial & Engineering Chemistry Research*, 53(21):8821–8830.

- Lampe, M., Stavrou, M., Schilling, J., Sauer, E., Gross, J., and Bardow, A. (2015). Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. *Computers & Chemical Engineering*, 81:278–287.
- Lapkin, A. A., Heer, P. K., Jacob, P.-M., Hutchby, M., Cunningham, W., Bull, S. D., and Davidson, M. G. (2017). Automation of route identification and optimisation based on data-mining and chemical intuition. *Faraday discussions*, 202:483–496.
- Lapkin, A. A., Peters, M., Greiner, L., Chemat, S., Leonhard, K., Liauw, M. A., and Leitner, W. (2010). Screening of new solvents for artemisinin extraction process using ab initio methodology. *Green Chemistry*, 12(2):241–251.
- Larriba, M., de Riva, J., Navarro, P., Moreno, D., Delgado-Mellado, N., García, J., Ferro, V. R., Rodríguez, F., and Palomar, J. (2018). COSMO-based/Aspen Plus process simulation of the aromatic extraction from pyrolysis gasoline using the [4empy][NTf₂] + [emim][DCA] ionic liquid mixture. *Separation and Purification Technology*, 190:211–227.
- Le Nhien, C., van Long, N. D., Kim, S., and Lee, M. (2016). Design and Assessment of Hybrid Purification Processes through a Systematic Solvent Screening for the Production of Levulinic Acid from Lignocellulosic Biomass. *Industrial & Engineering Chemistry Research*, 55(18):5180–5189.
- Lee, J. W., Brüggemann, S., and Marquardt, W. (2003). Shortcut method for kinetically controlled reactive distillation systems. *AIChE Journal*, 49(6):1471–1487.
- Lehmann, A. and Maranas, C. D. (2004). Molecular Design Using Quantum Chemical Calculations for Property Estimation. *Industrial & Engineering Chemistry Research*, 43(13):3419–3432.
- Leonhard, K. and Deiters, U. K. (2000). Monte carlo simulations of neon and argon using ab initio potentials. *Molecular Physics*, 98(20):1603–1616.
- Lin, B., Chavali, S., Camarda, K., and Miller, D. C. (2005). Computer-aided molecular design using Tabu search. *Computers & Chemical Engineering*, 29(2):337–347.
- Linke, P., Papadopoulos, A. I., and Seferlis, P. (2015). Systematic methods for working fluid selection and the design, integration and control of Organic Rankine Cycles – a review. *Energies*, 8(6):4755–4801.

- Lyu, Z., Zhou, T., Chen, L., Ye, Y., Sundmacher, K., and Qi, Z. (2014). Simulation based ionic liquid screening for benzene–cyclohexane extractive separation. *Chemical Engineering Science*, 113:45–53.
- Mac Dowell, N., Llovel, F., Adjiman, C. S., Jackson, G., and Galindo, A. (2010). Modeling the Fluid Phase Behavior of Carbon Dioxide in Aqueous Solutions of Monoethanolamine Using Transferable Parameters with the SAFT-VR Approach. *Industrial & Engineering Chemistry Research*, 49(4):1883–1899.
- Marcoulaki, E. C. and Kokossis, A. C. (1998). Molecular design synthesis using stochastic optimisation as a tool for scoping and screening. *Computers & Chemical Engineering*, 22:S11–S18.
- Marquardt, W., Kossack, S., and Krämer, K. (2008). A Framework for the Systematic Design of Hybrid Separation Processes. *Chinese Journal of Chemical Engineering*, 16(3):333–342.
- Marrero, J. and Gani, R. (2001). Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183-184:183–208.
- Marrero, J. and Gani, R. (2002). Group-Contribution-Based Estimation of Octanol/Water Partition Coefficient and Aqueous Solubility. *Industrial & Engineering Chemistry Research*, 41(25):6623–6633.
- Mayer, F. (2017). Economic process evaluation for integrated process and solvent design using shortcut process models and COSMO-RS. *Bachelorthesis, Institute of Technical Thermodynamics, RWTH Aachen University*.
- McBride, K., Gaide, T., Vorholt, A., Behr, A., and Sundmacher, K. (2016). Thermomorphic solvent selection for homogeneous catalyst recovery based on COSMO-RS. *Chemical Engineering and Processing: Process Intensification*, 99:97–106.
- McLeese, S. E., Eslick, J. C., Hoffmann, N. J., Scurto, A. M., and Camarda, K. V. (2010). Design of ionic liquids via computational molecular design. *Computers & Chemical Engineering*, 34(9):1476–1480.
- Moity, L., Durand, M., Benazzouz, A., Pierlot, C., Molinier, V., and Aubry, J.-M. (2012). Panorama of sustainable solvents using the COSMO-RS approach. *Green Chemistry*, 14(4):1132–1145.

- Moity, L., Molinier, V., Benazzouz, A., Joossen, B., Gerbaud, V., and Aubry, J.-M. (2016). A “top-down” in silico approach for designing ad hoc bio-based solvents: Application to glycerol-derived solvents of nitrocellulose. *Green Chemistry*, 18(11):3239–3249.
- Murat Sen, S., Henao, C. A., Braden, D. J., Dumesic, J. A., and Maravelias, C. T. (2012). Catalytic conversion of lignocellulosic biomass to fuels: Process development and technoeconomic evaluation. *Chemical Engineering Science*, 67(1):57–67.
- Ng, L. Y., Chong, F. K., and Chemmangattuvalappil, N. G. (2015). Challenges and opportunities in computer-aided molecular design. *Computers & Chemical Engineering*, 81:115–129.
- Odele, O. and Macchietto, S. (1993). Computer aided molecular design: a novel method for optimal solvent selection. *Fluid Phase Equilibria*, 82:47–54.
- Okabe, M., Tamagawa, H., and Tada, M. (2006). Synthesis of (3-Furanyl) Methyl Derivative. *Synthetic Communications*, 13(5):373–378.
- Ourique, J. E. and Silva Telles, A. (1998). Computer-aided molecular design with simulated annealing and molecular graphs. *Computers & Chemical Engineering*, 22:S615–S618.
- Papadopoulos, A. I., Badr, S., Chremos, A., Forte, E., Zarogiannis, T., Seferlis, P., Papadokostantakis, S., Galindo, A., Jackson, G., and Adjiman, C. S. (2016). Computer-aided molecular design and selection of CO₂ capture solvents based on thermodynamics, reactivity and sustainability. *Molecular Systems Design & Engineering*, 1(3):313–334.
- Papadopoulos, A. I. and Linke, P. (2005). A Unified Framework for Integrated Process and Molecular Design. *Chemical Engineering Research and Design*, 83(6):674–678.
- Papadopoulos, A. I. and Linke, P. (2006a). Efficient integration of optimal solvent and process design using molecular clustering. *Chemical Engineering Science*, 61(19):6316–6336.
- Papadopoulos, A. I. and Linke, P. (2006b). Multiobjective molecular design for integrated process-solvent systems synthesis. *AIChE Journal*, 52(3):1057–1070.
- Pavurala, N. and Achenie, L. E. (2013). A mechanistic approach for modeling oral drug delivery. *Computers & Chemical Engineering*, 57:196–206.

- Peng, D.-Y. and Robinson, D. B. (1976). A New Two-Constant Equation of State. *Industrial & Engineering Chemistry Fundamentals*, 15(1):59–64.
- Pereira, F. E., Keskes, E., Galindo, A., Jackson, G., and Adjiman, C. S. (2011). Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane. *Computers & Chemical Engineering*, 35(3):474–491.
- Phillips, K. A., Wambaugh, J. F., Grulke, C. M., Dionisio, K. L., and Isaacs, K. K. (2017). High-throughput screening of chemicals as functional substitutes using structure-based classification models. *Green Chemistry*, 19(4):1063–1074.
- Prat, D., Hayler, J., and Wells, A. (2014). A survey of solvent selection guides. *Green Chemistry*, 16(10):4546–4551.
- Preißinger, M., Schwöbel, J. A., Klamt, A., and Brüggemann, D. (2017). Multi-criteria evaluation of several million working fluids for waste heat recovery by means of Organic Rankine Cycle in passenger cars and heavy-duty trucks. *Applied Energy*, 206:887–899.
- Pretel, E. J., López, P. A., Bottini, S. B., and Brignole, E. A. (1994). Computer-aided molecular design of solvents for separation processes. *AIChE Journal*, 40(8):1349–1360.
- Rappoport, Z. (1967). *CRC handbook of tables for organic compound identification*. CRC Press, Boca Raton, Florida, 3rd edition.
- Recker, S., Skiborowski, M., Redepenning, C., and Marquardt, W. (2015). A unifying framework for optimization-based design of integrated reaction–separation processes. *Computers & Chemical Engineering*, 81:260–271.
- Redepenning, C. and Marquardt, W. (2017). Pinch-based shortcut method for the conceptual design of adiabatic absorption columns. *AIChE Journal*, 63(4):1213–1225.
- Redepenning, C., Recker, S., and Marquardt, W. (2016). Pinch-based shortcut method for the conceptual design of isothermal extraction columns. *AIChE Journal*, 63(4):1236–1245.
- Renon, H. and Prausnitz, J. M. (1968). Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal*, 14(1):135–144.

- Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293.
- Román-Leshkov, Y., Barrett, C. J., Liu, Z. Y., and Dumesic, J. A. (2007). Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature*, 447(7147):982–985.
- Sahinidis, N. V. (2004). Optimization under uncertainty: State-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6-7):971–983.
- Sahinidis, N. V., Tawarmalani, M., and Yu, M. (2003). Design of alternative refrigerants via global optimization. *AIChE Journal*, 49(7):1761–1775.
- Salleh, Z., Wazeer, I., Mulyono, S., El-blidi, L., Hashim, M. A., and Hadj-Kali, M. K. (2017). Efficient removal of benzene from cyclohexane-benzene mixtures using deep eutectic solvents – COSMO-RS screening and experimental validation. *The Journal of Chemical Thermodynamics*, 104:33–44.
- Samudra, A. P. and Sahinidis, N. V. (2013). Optimization-based framework for computer-aided molecular design. *AIChE Journal*, 59(10):3686–3701.
- Satola, B. J., Rarey, J., and Ramjugernath, D. (2017). Quality of Component- and Group-Interaction-Based Regression of Binary Vapor–Liquid Equilibrium Data. *Industrial & Engineering Chemistry Research*, 56(37):10500–10505.
- Schilling, J., Lampe, M., Gross, J., and Bardow, A. (2017a). 1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT. *Chemical Engineering Science*, 159:217–230.
- Schilling, J., Tillmanns, D., Lampe, M., Hopp, M., Gross, J., and Bardow, A. (2017b). From molecules to dollars: Integrating molecular design into thermo-economic process design using consistent thermodynamic modeling. *Molecular Systems Design & Engineering*, 2:301–320.
- Schwöbel, J. A. H., Preißinger, M., Brüggemann, D., and Klamt, A. (2017). High-Throughput Screening of Working Fluids for the Organic Rankine Cycle (ORC) Based on Conductor-like Screening Model for Realistic Solvation (COSMO-RS) and Thermodynamic Process Simulations. *Industrial & Engineering Chemistry Research*, 56(3):788–798.

- Sendek, A. D., Yang, Q., Cubuk, E. D., Duerloo, K.-A. N., Cui, Y., and Reed, E. J. (2017). Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.*, 10(1):306–320.
- Sheldon, T. J., Folić, M., and Adjiman, C. S. (2006). Solvent Design Using a Quantum Mechanical Continuum Solvation Model. *Industrial & Engineering Chemistry Research*, 45(3):1128–1140.
- Sigma-Aldrich (2017). <https://www.sigmaaldrich.com/>.
- Sinha, M., Achenie, L. E., and Ostrovsky, G. M. (1999). Environmentally benign solvent design by global optimization. *Computers & Chemical Engineering*, 23(10):1381–1394.
- Sinha, M., Achenie, L. E. K., and Gani, R. (2003). Blanket Wash Solvent Blend Design Using Interval Analysis. *Industrial & Engineering Chemistry Research*, 42(3):516–527.
- Skiborowski, M., Harwardt, A., and Marquardt, W. (2013). Conceptual Design of Distillation-Based Hybrid Separation Processes. *Annual Review of Chemical and Biomolecular Engineering*, 4(1):45–68.
- Skiborowski, M., Harwardt, A., and Marquardt, W. (2015a). Efficient optimization-based design for the separation of heterogeneous azeotropic mixtures. *Computers & Chemical Engineering*, 72:34–51.
- Skiborowski, M., Rautenberg, M., and Marquardt, W. (2015b). A Hybrid Evolutionary–Deterministic Optimization Approach for Conceptual Design. *Industrial & Engineering Chemistry Research*, 54(41):10054–10072.
- Smith, B. D. and Brinkley, W. K. (1960). General short-cut equation for equilibrium stage processes. *AIChE Journal*, 6(3):446–450.
- Soave, G. (1972). Equilibrium constants from a modified redlich-kwong equation of state. *Chemical Engineering Science*, 27(6):1197–1203.
- Sondheimer, F., Ben-Efraim, D. A., and Gaoni, Y. (1961). Unsaturated Macrocyclic Compounds. XVIII. The Prototropic Rearrangement of Linear 1,5-Diynes to Conjugated Polyyenes. *Journal of the American Chemical Society*, 83(7):1682–1685.
- Song, Z., Zhou, T., Qi, Z., and Sundmacher, K. (2017). Systematic Method for Screening Ionic Liquids as Extraction Solvents Exemplified by an Extractive Desulfurization Process. *ACS Sustainable Chemistry & Engineering*, 5(4):3382–3389.

- Song, Z., Zhou, T., Zhang, J., Cheng, H., Chen, L., and Qi, Z. (2015). Screening of ionic liquids for solvent-sensitive extraction – with deep desulfurization as an example. *Chemical Engineering Science*, 129:69–77.
- Spearman, C. (1987). The proof and measurement of association between two things. by C. Spearman, 1904. *The American journal of psychology*, 100(3-4):441–471.
- Spieß, A. C., Eberhard, W., Peters, M., Eckstein, M. F., Greiner, L., and Büchs, J. (2008). Prediction of partition coefficients using COSMO-RS: Solvent screening for maximum conversion in biocatalytic two-phase reaction systems. *Chemical Engineering and Processing: Process Intensification*, 47(6):1034–1041.
- Spuhl, O. and Arlt, W. (2004). COSMO-RS Predictions in Chemical Engineering – A Study of the Applicability to Binary VLE. *Industrial & Engineering Chemistry Research*, 43(4):852–861.
- Stanescu, I. and Achenie, L. E. (2006). A theoretical study of solvent effects on Kolbe–Schmitt reaction kinetics. *Chemical Engineering Science*, 61(18):6199–6212.
- Stavrou, M., Lampe, M., Bardow, A., and Gross, J. (2014). Continuous Molecular Targeting–Computer-Aided Molecular Design (CoMT–CAMD) for Simultaneous Process and Solvent Design for CO₂ Capture. *Industrial & Engineering Chemistry Research*, 53(46):18029–18041.
- Stewart, James J. P. (2012). MOPAC program package, <http://openmopac.net/>.
- Struebing, H., Ganase, Z., Karamertzanis, P. G., Sioumkrou, E., Haycock, P., Piccione, P. M., Armstrong, A., Galindo, A., and Adjiman, C. S. (2013). Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nature Chemistry*, 5(11):952–957.
- Supronowicz, W., Ignatyev, I. A., Lolli, G., Wolf, A., Zhao, L., and Mleczko, L. (2015). Formic acid: A future bridge between the power and chemical industries. *Green Chemistry*, 17(5):2904–2911.
- Tobiszewski, M., Tsakovski, S., Simeonov, V., Namieśnik, J., and Pena-Pereira, F. (2015). A solvent selection guide based on chemometrics and multicriteria decision analysis. *Green Chemistry*, 17(10):4773–4785.
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3093.

- Tomasi, J. and Persico, M. (1994). Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chemical Reviews*, 94(7):2027–2094.
- Ulonska, K., Skiborowski, M., Mitsos, A., and Viell, J. (2016). Early-stage evaluation of biorefinery processing pathways using process network flux analysis. *AIChE Journal*, 62(9):3096–3108.
- Underwood, A. J. V. (1949). Fractional Distillation of Multicomponent Mixtures. *Industrial & Engineering Chemistry*, 41(12):2844–2847.
- van Putten, R.-J., van der Waal, Jan C, de Jong, E., Rasrendra, C. B., Heeres, H. J., and de Vries, Johannes G (2013). Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chemical reviews*, 113(3):1499–1597.
- Venkatasubramanian, V., Chan, K., and Caruthers, J. M. (1995). Evolutionary Design of Molecules with Desired Properties Using the Genetic Algorithm. *Journal of Chemical Information and Modeling*, 35(2):188–195.
- Voutchkova, A. M., Kostal, J., Steinfeld, J. B., Emerson, J. W., Brooks, B. W., Anastas, P., and Zimmerman, J. B. (2011). Towards rational molecular design: Derivation of property guidelines for reduced acute aquatic toxicity. *Green Chemistry*, 13(9):2373–2379.
- Wang, J. and Lakerveld, R. (2017). Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT. *AIChE Journal*, 83(6):730.
- Weininger, D. (1990). SMILES. 3. Graphical depiction of chemical structures. *Journal of Chemical Information and Modeling*, 30(3):237–243.
- Weis, D. C. and Visco, D. P. (2010). Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. *Computers & Chemical Engineering*, 34(7):1018–1029.
- Weiss, H., Deglmann, P., In ’t Veld, Pieter J, Cetinkaya, M., and Schreiner, E. (2016). Multiscale Materials Modeling in an Industrial Environment. *Annual Review of Chemical and Biomolecular Engineering*, 7:65–86.
- Whelton, A. J., McMillan, L., Connell, M., Kelley, K. M., Gill, J. P., White, K. D., Gupta, R., Dey, R., and Novy, C. (2015). Residential tap water contamination following the freedom industries chemical spill: perceptions, water quality, and health impacts. *Environmental Science & Technology*, 49(2):813–823.

- Wilson, G. M. and Deal, C. H. (1962). Activity coefficients and molecular structure. Activity coefficients in changing environments-solutions of groups. *Industrial & Engineering Chemistry Fundamentals*, 1(1):20–23.
- Won, W., Motagamwala, A. H., Dumesic, J. A., and Maravelias, C. T. (2017). A co-solvent hydrolysis strategy for the production of biofuels: Process synthesis and technoeconomic analysis. *Reaction Chemistry & Engineering*, 2(3):397–405.
- Wu, H. S. and Sandler, S. I. (1991). Use of ab initio quantum mechanics calculations in group contribution methods. 1. theory and the basis for group identifications. *Industrial & Engineering Chemistry Research*, 30(5):881–889.
- Xu, W. and Diwekar, U. M. (2005). Improved Genetic Algorithms for Deterministic Optimization and Optimization under Uncertainty. II. Solvent Selection under Uncertainty. *Industrial & Engineering Chemistry Research*, 44(18):7138–7146.
- Zaitseva, A., Pokki, J.-P., Le, H. Q., Alopaeus, V., and Sixta, H. (2016). Vapor–Liquid Equilibria, Excess Enthalpy, and Density of Aqueous γ -Valerolactone Solutions. *Journal of Chemical & Engineering Data*, 61(2):881–890.
- Zhang, L., Cignitti, S., and Gani, R. (2015). Generic mathematical programming formulation and solution for computer-aided molecular design. *Computers & Chemical Engineering*, 78:79–84.
- Zhou, T., Chen, L., Ye, Y., Chen, L., Qi, Z., Freund, H., and Sundmacher, K. (2012). An Overview of Mutual Solubility of Ionic Liquids and Water Predicted by COSMO-RS. *Industrial & Engineering Chemistry Research*, 51(17):6256–6264.
- Zhou, T., Lyu, Z., Qi, Z., and Sundmacher, K. (2015a). Robust design of optimal solvents for chemical reactions-A combined experimental and computational strategy. *Chemical Engineering Science*, 137:613–625.
- Zhou, T., McBride, K., Zhang, X., Qi, Z., and Sundmacher, K. (2015b). Integrated solvent and process design exemplified for a Diels-Alder reaction. *AIChE Journal*, 61(1):147–158.
- Zhou, T., Qi, Z., and Sundmacher, K. (2014). Model-based method for the screening of solvents for chemical reactions. *Chemical Engineering Science*, 115:177–185.
- Zhou, T., Wang, J., McBride, K., and Sundmacher, K. (2016). Optimal design of solvents for extractive reaction processes. *AIChE Journal*, 62(9):3238–3249.

- Zhou, T., Zhou, Y., and Sundmacher, K. (2017). A hybrid stochastic–deterministic optimization approach for integrated solvent and process design. *Chemical Engineering Science*, 159:207–216.
- Zimmerman, J. B. and Anastas, P. T. (2015). Toward substitution with no regrets. *Science*, 347(6227):1198–1199.

Aachener Beiträge zur Technischen Thermodynamik

ABTT 1

Philip Voll

Automated Optimization-Based Synthesis of Distributed Energy Supply Systems

1. Auflage 2014

ISBN 978-3-86130-474-6

ABTT 2

Johannes Jung

Comparative Life Cycle Assessment of Industrial Multi-Product Processes

1. Auflage 2014

ISBN 978-3-86130-471-5

ABTT 3

Franz Lanzerath

Modellgestützte Entwicklung von Adsorptionswärmepumpen

1. Auflage 2014

ISBN 978-3-86130-472-2

ABTT 4

Thorsten Brands

Einfluss der Gemischzusammensetzung auf die Verbrennung im Diesel- und GCI-Motor

1. Auflage 2014

ISBN 978-3-95886-006-3

ABTT 5

Dominique Dechambre

Efficient Measurement of Liquid-Liquid Equilibria using Automation and Optimal Experimental Design

1. Auflage 2016

ISBN 978-3-95886-077-3

ABTT 6

Niklas von der Aßen

From Life-Cycle Assessment towards life-Cycle Design of Carbon Dioxide Capture and Utilization

1. Auflage 2016

ISBN 978-3-95886-080-3

ABTT 7

Matthias Lampe

Integrated Process and Organic Rankine Cycle Working Fluid Design in the Continuous-Molecular Targeting Framework

1. Auflage 2016

ISBN 978-3-95886-086-5

ABTT 8

Thomas Hülser

Optische Untersuchung der Zündvorgänge und deren Auswirkung auf die Verbrennung in PKW-Motoren

1. Auflage 2016

ISBN 978-3-95886-090-2

Aachener Beiträge zur Technischen Thermodynamik

ABTT 9

Malte Döntgen

Reaction Models from Reactive Molecular Dynamics and High-Level Kinetics Predictions

1. Auflage 2016

ISBN 978-3-95886-156-5

ABTT 10

Heike Schreiber

Experiments and Validated Models for Adsorption Thermal Energy Storage in Industrial and Residential Application

1. Auflage 2017

ISBN 978-3-95886-178-7

ABTT 11

André Dirk Sternberg

System-Wide Perspective for Life Cycle Assessment of CO₂-based C1-Chemicals

1. Auflage 2017

ISBN 978-3-95886-193-0

ABTT 12

Uwe Bau

From Dynamic Simulation to Optimal Design and Control of Adsorption Energy Systems

1. Auflage 2018

ISBN 978-3-95886-216-6

ABTT 13

Christian Jens

Modellbasiertes Design von Produkt, Lösungsmittel und Prozess für die Ameisensäure-synthese aus CO₂ und H₂

1. Auflage 2018

ISBN 978-3-95886-231-9

ABTT 14

Jan David Scheffczyk

Integrated Computer-Aided Design of Molecules and Processes using COSMO-RS

1. Auflage 2018

ISBN 978-3-95886-236-4