

# **Systematic Solvent Screening and Design for Biomass Fractionation Processes**

## **Dissertation**

Zur Erlangung des akademischen Grades

## **Doktoringenieurin (Dr.-Ing.)**

von Laura König-Mattern, M.Sc.

geb. am 07. Oktober 1994 in Gardelegen

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik  
der Otto-von-Guericke-Universität Magdeburg

Gutachter/innen:

Prof. Dr.-Ing. habil. Kai Sundmacher

Prof. Dr. rer. nat. Katalin Barta Weissert

Prof. Dr.-Ing. habil. Mirjana Minceva

Promotionskolloquium am 16.12.2024

”

*Solvent selection is both an art and a science.*

---

Everyone in chemistry

“

Laura König-Mattern

„*Systematic Solvent Screening and Design for Biomass Fractionation Processes*“

Dissertation

Magdeburg, September 6, 2024

## Abstract

In view of depleting fossil-resources and climate change, bio-based feedstocks are envisioned to serve as a renewable carbon source in a circular economy. Biorefineries play a vital role in the circular economy, since they enable resource efficient utilisation of biomass, such as lignocellulose or microalgae. Biorefinery processes fractionate the feedstocks into molecular building blocks to produce pharmaceuticals, chemicals, food, feed, and fuels, thereby closing material loops.

In a fractionation process, these building blocks are extracted from the biomass and further separated, frequently utilising organic solvents. The use of organic solvents significantly impacts the product yields, the energy demand, greenhouse gas emissions, operational safety, and economic viability of a biorefinery process. Therefore, solvent selection is a molecular-level decision with far-reaching consequences for the overall process. Computational methods can significantly accelerate solvent selection and guide experiments toward the most promising candidates. However, there is a lack of efficient computational methods suitable for rational solvent selection and design for biomass fractionation, limiting the development of innovative biorefinery strategies.

The present thesis addresses this gap by introducing computational methods for solvent screening and design, which were experimentally validated on lignocellulose and microalgae. A high-throughput screening method was developed to evaluate a database containing more than 8000 potential solvents. The method applies computational models to predict important solvent properties. Based on these predictions, solvent candidates with undesired structural features, thermophysical and thermodynamic properties, as well as environmental, health, and safety properties can be eliminated. Thus, the search is narrowed, enabling targeted experimental tests. However, for highly constrained solvent selection problems, identifying a solvent within the database that meets the many selection criteria may become difficult. To allow for rational solvent selection beyond a pre-defined database, a computational solvent design method was developed. This method can be used to tailor the solvents' molecular structure toward the desired properties using a graph-based genetic algorithm. The close combination of computational methods and experiments enabled the identification effective solvents applicable for the fractionation of lignocellulosic and microalgal biomass.

For the fractionation of lignocellulosic biomass and further valorisation of the extracted

lignin, high lignin solubility is an important solvent selection criterion. Common fractionation approaches treat the biomass with acids, high temperatures and the selected solvent, resulting in undesired lignin condensation reactions which limit further lignin valorisation. Aldehyde-assisted fractionation, an innovative biorefinery approach, inhibits such condensation reactions by lignin stabilisation with aldehydes. In this way, not only the cellulose-rich pulp and the hemicellulose sugars, but also the lignin fraction can be effectively valorised. However, this approach employs the carcinogenic solvent 1,4-dioxane which should be replaced by more benign alternatives. The developed computational methods identified solvents with high, experimentally confirmed lignin solubilities ranging from 20 to 60 wt.% ( $T = 85\text{ }^{\circ}\text{C}$ ). Fractionation experiments showed, that lignin was effectively stabilised by an aldehyde in many of the identified solvents. Some of the tested solvents outperformed 1,4-dioxane in terms of hemicellulose sugar yield, and/or toxicity, with a slightly lower lignin monomer yield as a trade-off.

Microalgal biorefining faces several challenges, involving the use of toxic solvents and energy-intensive biomass drying. To circumvent the drying step, wet algal paste, still containing about 85 wt.% moisture, was investigated as a feedstock. The moisture is commonly considered as a barrier that further complicates biomass fractionation. The developed screening approach identified solvents applicable to the fractionation of wet biomass of the model alga *P. tricornutum*. By combining the computational solvent screening approach with experimental methods, a lab-scale biomass fractionation process for wet *P. tricornutum* biomass was developed. Breaking with the current view of treating water as a barrier, this approach exploits the presence of water to fractionate the biomass into lipids, carotenoids, carbohydrates, and proteins. The developed biorefinery approach does not require biomass drying and employs only two benign solvents to fractionate the biomass at ambient conditions.

Overall, the combination of computational solvent selection methods and experimental work allowed to replace harmful solvents from current fractionation approaches and paved the way for developing innovative biorefinery processes.

## Zusammenfassung

Angesichts des Klimawandels und der schwindenden fossilen Ressourcen wird angestrebt, Biomasse als nachwachsende Kohlenstoffquelle in einer Kreislaufwirtschaft zu nutzen. Bio-raffinerien spielen dabei eine zentrale Rolle, da sie eine ressourceneffiziente Valorisierung von Biomasse, wie beispielsweise Lignocellulose oder Mikroalgen, ermöglichen. In Bio-raffinerieprozessen wird Biomasse in ihre molekularen Grundbausteine fraktioniert. Diese Grundbausteine dienen der Produktion von Arzneimitteln, Chemikalien, Lebens- und Futtermitteln, sowie von Kraftstoffen und können Stoffkreisläufe innerhalb der Kreislaufwirtschaft schließen.

Die Fraktionierung umfasst die Extraktion dieser Grundbausteine aus der Biomasse und weitere Separationsschritte, wobei häufig organische Lösungsmittel zum Einsatz kommen. Die Verwendung von Lösungsmitteln hat einen erheblichen Einfluss auf die Produktausbeute, den Energiebedarf, die Treibhausgasemissionen, die Betriebssicherheit, und die Wirtschaftlichkeit eines Bio-raffinerieprozesses. Daher stellt die Auswahl der Lösungsmittel eine Entscheidung auf molekularer Ebene dar, die weitreichende Auswirkungen auf den Gesamtprozess hat. Computergestützte Methoden können die Lösungsmittelauswahl erheblich beschleunigen und Experimente auf die vielversprechendsten Kandidaten beschränken. Obwohl die systematische Lösungsmittelauswahl von hoher Bedeutung für Bio-raffinerieprozesse ist, wurde das Potential computergestützter Methoden zu diesem Zweck bisher kaum ausgeschöpft.

In der vorliegenden Arbeit wurden computerbasierte Screening- und Designmethoden entwickelt, die eine systematische Lösungsmittelauswahl für die Biomassefraktionierung ermöglichen. Die entwickelte Methodik wurde experimentell für Lignocellulose und Mikroalgenbiomasse validiert. Im Rahmen des Lösungsmittel-Screenings wird eine Datenbank mit mehr als 8000 potenziellen Lösungsmitteln systematisch nach passenden Lösungsmitteln durchsucht. Diese Methode verwendet computergestützte Modelle, um die wichtigsten Lösungsmiteleigenschaften zu berechnen. Auf Grundlage dieser Berechnungen können Lösungsmittelkandidaten mit erwünschten strukturellen Merkmalen, thermophysikalischen und thermodynamischen Eigenschaften, sowie unkritischen Umwelt-, Gesundheits- und Sicherheitseigenschaften identifiziert werden. Um eine rationale Lösungsmittelauswahl über die vordefinierte Datenbank hinaus zu ermöglichen, wurde eine rechnergestützte Methode

für das Design von Lösungsmitteln mit gewünschten Zieleigenschaften entwickelt. Diese Methode abstrahiert die molekulare Struktur der Lösungsmittel als Graphen, der mithilfe eines genetischen Algorithmus an die Zieleigenschaft angepasst wird. Die entwickelte Screening- und Design-Methodik konnte in enger Kombination mit experimenteller Arbeit effektive Lösungsmittel für die Fraktionierung von Lignocellulose und Mikroalgenbiomasse identifizieren.

Bei der Fraktionierung von Lignocellulose und der Valorisierung des extrahierten Lignins spielt die Ligninlöslichkeit des gewählten Lösungsmittels eine große Rolle. Lignin neigt während der Fraktionierung, die häufig bei hohen Temperaturen und in saurer Umgebung stattfindet, zu unerwünschter Kondensierung, welche die Nutzbarkeit des extrahierten Lignins stark einschränkt. Ein innovativer Bioraffinerieansatz ist die aldehydgestützte Fraktionierung, die eine Ligninkondensierung durch Stabilisierung mittels Aldehyden verhindert. Somit wird neben der Cellulose-Fraktion und den Hemicellulose-Zuckern auch Lignin für ein breites Produktspektrum nutzbar. In diesem Ansatz wird jedoch das karzinogene Lösungsmittel 1,4-Dioxan eingesetzt, welches mithilfe der in dieser Arbeit entwickelten, rechnergestützten Methoden ersetzt werden soll. Diese Methoden identifizierten Lösungsmittel mit hohen Ligninlöslichkeiten zwischen 20 und 60 Gew.% ( $T = 85 \text{ °C}$ ), welche experimentell bestätigt wurden. Fraktionierungsexperimente zeigten, dass Lignin in einer Vielzahl der identifizierten Lösungsmittel effektiv durch ein Aldehyd stabilisiert werden konnte. Einige der getesteten Lösungsmittel übertrafen 1,4-Dioxan in Bezug auf die Ausbeute von Hemicellulose-Zuckern bei geringerer Toxizität, wobei die Ausbeute der Ligninmonomere etwas niedriger ausfiel.

Besondere Herausforderungen bei der Fraktionierung von Mikroalgenbiomasse stellen die Verwendung toxischer Lösungsmittel und die energieintensive Trocknung der Biomasse dar. Um den Trocknungsschritt zu umgehen, soll feuchte Algenpaste, die noch etwa 85 Gew.% Feuchtigkeit enthält, als Rohstoff eingesetzt werden. Die Feuchtigkeit wird im Allgemeinen als Barriere betrachtet, die die Fraktionierung der Biomasse erschwert. Durch die Kombination des Lösungsmittel-Screenings mit experimentellen Methoden konnte für die Modellalge *P. tricornutum* ein neuer Ansatz zur Biomassefraktionierung entwickelt werden. Im Gegensatz zur gängigen Auffassung nutzt dieser Ansatz das Vorhandensein von Wasser gezielt aus, um die Biomasse in Lipide, Carotinoide, Kohlenhydrate und Proteine zu fraktionieren. Der entwickelte Fraktionierungsprozess erfordert keine Trocknung der Biomasse und verwendet lediglich zwei ungiftige Lösungsmittel zur Fraktionierung der Biomasse bei Umgebungsbedingungen.

Insgesamt konnten durch die Kombination von computergestützten Lösungsmittelauswahlverfahren und experimentellen Arbeiten schädliche Lösungsmittel aus den derzeitigen Fraktionierungsansätzen ersetzt und neue Bioraffinerieprozesse entwickelt werden.

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>List of symbols</b>	<b>xiii</b>
<b>1 Introduction and motivation</b>	<b>1</b>
1.1 Planetary boundaries and the chemical industry . . . . .	1
1.2 The need for a circular bioeconomy . . . . .	2
1.3 Biorefineries as drivers of a circular bioeconomy . . . . .	4
1.4 Solvent selection for biorefinery processes: State of the art and open questions	5
1.5 Aims and outline of this work . . . . .	7
<b>2 Lignocellulose processing</b>	<b>9</b>
2.1 Biomass composition and molecular structure of lignocellulose . . . . .	9
2.1.1 Cellulose . . . . .	10
2.1.2 Hemicellulose . . . . .	11
2.1.3 Lignin . . . . .	12
2.2 State-of-the-art lignocellulose processing . . . . .	13
2.2.1 Pyrolysis . . . . .	13
2.2.2 Kraft pulping . . . . .	14
2.2.3 Organosolv processing . . . . .	14
2.2.4 Lignin-first biorefineries . . . . .	15
<b>3 Microalgal processing</b>	<b>19</b>
3.1 Biomass composition and molecular structure of the model alga <i>P. tricornutum</i>	20
3.1.1 Pigments . . . . .	20
3.1.2 Carbohydrates . . . . .	21
3.1.3 Lipids . . . . .	22
3.1.4 Proteins . . . . .	22
3.2 State-of-the-art microalgal processing . . . . .	22

3.2.1	Cultivation . . . . .	22
3.2.2	Harvest . . . . .	23
3.2.3	Drying . . . . .	24
3.2.4	Cell disruption . . . . .	24
3.2.5	Biofuel production from microalgae . . . . .	24
3.2.6	Separation methods in microalgal biorefineries . . . . .	25
<b>4</b>	<b>Computer-aided solvent screening and design: state of the art</b>	<b>29</b>
4.1	The inverse problem as an optimisation problem . . . . .	30
4.2	Methods for chemical space exploration . . . . .	31
4.2.1	Screening methods . . . . .	31
4.2.2	Molecular design methods . . . . .	32
4.3	Property predictions . . . . .	34
4.3.1	Thermodynamic models . . . . .	34
4.3.2	EHS property models . . . . .	39
4.4	Molecular representation . . . . .	40
<b>5</b>	<b>Development of computer-aided solvent screening and design methods for biorefinery processes</b>	<b>43</b>
5.1	Solvent selection in biorefineries as a CAMSD optimisation problem . . . . .	43
5.2	Solvent screening method for biorefinery processes . . . . .	44
5.2.1	Database . . . . .	46
5.2.2	Representative biomolecules . . . . .	46
5.2.3	Thermodynamic property predictions . . . . .	47
5.2.4	Prediction of EHS properties . . . . .	49
5.2.5	Computational details . . . . .	51
5.2.6	Advantages and limitations . . . . .	52
5.3	Solvent design method: The genetic algorithm <i>PSEvolve</i> . . . . .	53
5.3.1	Structure-altering operations . . . . .	55
5.3.2	Computational details . . . . .	59
5.3.3	Advantages and limitations . . . . .	59
5.4	Summary . . . . .	60
<b>6</b>	<b>Solvent screening and design for lignocellulose biorefineries</b>	<b>61</b>
6.1	Solvent screening . . . . .	62
6.1.1	Optimisation problem and screening procedure . . . . .	63
6.1.2	Representative lignocellulose molecules . . . . .	64
6.1.3	Screening results and solvent ranking . . . . .	66



6.2	Solvent design for lignin dissolution in lignin-first biorefineries and lignin up- grading . . . . .	72
6.2.1	Optimisation problem and solvent design algorithm . . . . .	73
6.2.2	Lignin solubility predictions using a graph neural network . . . . .	74
6.2.3	Solvent design for lignin dissolution . . . . .	77
6.2.4	Solvent design for aldehyde-assisted fractionation . . . . .	80
6.3	Experimental validation and investigating underlying structural patterns for lignin solubility . . . . .	82
6.3.1	Solvent selection . . . . .	82
6.3.2	Lignin solubility measurements . . . . .	82
6.3.3	Propionaldehyde-assisted pretreatment . . . . .	86
6.4	Experimental methods . . . . .	90
6.4.1	Materials . . . . .	90
6.4.2	Lignin solubility measurements by gas chromatography . . . . .	91
6.4.3	Lignin solubility measurements by evaporation . . . . .	92
6.4.4	Propionaldehyde-assisted pretreatment . . . . .	92
6.5	Summary and conclusions . . . . .	93
<b>7</b>	<b>Solvent screening for the fractionation of wet microalgal biomass</b>	<b>97</b>
7.1	Biomass composition and potential biorefinery products of <i>P. tricornutum</i> . . . . .	98
7.2	Solvent screening . . . . .	99
7.2.1	Mechanism of extracting wet microalgal biomass and solvent selection criteria . . . . .	100
7.2.2	Optimisation problem and screening procedure . . . . .	102
7.2.3	Representative microalgal molecules and reference solvent system . . . . .	103
7.2.4	Screening results . . . . .	106
7.3	Experimental validation . . . . .	112
7.3.1	Lipid extraction from wet <i>P. tricornutum</i> biomass . . . . .	113
7.3.2	Carotenoid and chlorophyll extraction from wet microalgal biomass . . . . .	114
7.3.3	Lipid and pigment extraction from dried microalgal biomass . . . . .	116
7.4	Lab-scale fractionation process for wet <i>P. tricornutum</i> biomass . . . . .	117
7.4.1	Integrated extraction and separation of lipids and carotenoids from wet <i>P. tricornutum</i> biomass . . . . .	118
7.4.2	Lab-scale 2-butanol/water-based biomass fractionation process for wet <i>P. tricornutum</i> biomass . . . . .	120
7.5	Experimental methods . . . . .	123
7.5.1	Cultivation of <i>P. tricornutum</i> . . . . .	123
7.5.2	Determination of the moisture content . . . . .	123

7.5.3	Lyophilisation . . . . .	124
7.5.4	Extraction procedure for wet and dry biomass . . . . .	124
7.5.5	Lipid analysis . . . . .	124
7.5.6	Pigment analysis . . . . .	125
7.5.7	Protein analysis . . . . .	126
7.5.8	Carbohydrate analysis . . . . .	126
7.5.9	Yield calculation . . . . .	127
7.5.10	Ash content . . . . .	127
7.5.11	Quantification of microalgal compounds after phase separation . . .	127
7.5.12	Error analysis of experiments . . . . .	128
7.5.13	Solvents . . . . .	128
7.6	Summary and conclusions . . . . .	128
<b>8</b>	<b>Conclusions and outlook</b>	<b>131</b>
<b>A</b>	<b>Phase equilibria</b>	<b>137</b>
A.1	Solubility of solids in liquids . . . . .	138
<b>B</b>	<b>Database</b>	<b>141</b>
<b>C</b>	<b>Lignocellulose</b>	<b>143</b>
C.1	Solvent screening . . . . .	143
C.1.1	Representative biomolecules . . . . .	143
C.1.2	Solvent ranking . . . . .	149
C.2	Solvent design . . . . .	150
C.2.1	Hyperparameters of <i>PSEvolve</i> . . . . .	150
C.3	Experimental . . . . .	157
C.3.1	2D HSQC NMR . . . . .	157
C.3.2	Lignin solubility measurements . . . . .	159
C.3.3	Biomass composition of Birch wood . . . . .	159
C.3.4	DPX and lignin monomer yields after AAF treatment . . . . .	159
<b>D</b>	<b>Microalgae</b>	<b>163</b>
D.1	Solvent screening . . . . .	163
D.1.1	Selection of representative molecules for <i>P. tricornutum</i> . . . . .	163
D.1.2	Manually added solvents after EHS screening step . . . . .	168
D.1.3	Solvent candidates identified for each biomass fraction . . . . .	169
D.2	COSMO-RS solubility prediction of representative molecules in 2- butanol/water-mixtures . . . . .	169
D.3	Experimental details . . . . .	170

D.3.1 Biomass composition of <i>P.tricornutum</i> . . . . .	170
D.4 Microscopic imaging of <i>P. tricornutum</i> cells . . . . .	173
D.5 Economic value of <i>P. tricornutum</i> biomass . . . . .	174
D.6 Mass balances of the 2-butanol-based biorefinery process . . . . .	174
<b>Bibliography</b>	<b>179</b>
<b>List of Figures</b>	<b>225</b>
<b>List of Tables</b>	<b>231</b>
<b>Afterword</b>	<b>235</b>
<b>Declaration</b>	<b>241</b>



## List of symbols

### Greek Symbols

$\gamma$	activity coefficient	-
$\epsilon$	permittivity	$F\text{ m}^{-1}$
$\mu$	chemical potential	$\text{J mol}^{-1}$
$\sigma$	screening charge density	$e\text{ \AA}^{-2}$
$\phi_{p,c}$	mass-based split ratio of mass flow $F_{p,c}$ that is transported to the subsequent process unit	

### Latin Symbols

$a$	activity	-
$\mathbf{A}$	incidence matrix	-
$\mathbf{b}$	vector of biomolecules	various
$d$	distance in property space	various
$f$	fitness	various
$F$	response factor (gas chromatography)	-
$F_{p,c}$	mass flow of process unit $p$ and compound $c$ on a dried biomass basis	$\text{wt.}\%_{\text{dry}}$
$\mathbf{F}$	vector of mass flows	-
$\mathcal{B}$	set of all biomolecules in the biomass	-
$E$	energy	$\text{J mol}^{-1}$
EHS score	measure for environmental, health, and safety risks of a substance	-
$\mathbf{e}$	vector of environmental, health, and safety properties	various
$\mathcal{E}$	set of all environmental, health, and safety properties	-
$\Delta G_{\text{fus}}$	molar Gibbs enthalpy of fusion	$\text{J mol}^{-1}$
$\Delta H_{\text{fus}}$	molar enthalpy of fusion	$\text{J mol}^{-1}$
$\Delta S_{\text{fus}}$	molar entropy of fusion	$\text{J mol}^{-1}\text{ K}^{-1}$
$\text{LD}_{50}$	dose of a substance with lethal effect on half of the test population	various

$n$	amount of substance	mol
$N$	number	various
$M$	molecular weight	$\text{g mol}^{-1}$ , kDa
$MC$	moisture content	wt. %
$\mathbf{o}$	vector of optimal properties	various
$p$	pressure, probability	bar, -
$\mathbf{p}$	vector of process conditions	various
$P$	partition coefficient, population	-, -
$\mathcal{P}$	set of process conditions	-
$Q$	net formal charge	-
$\mathbf{s}$	vector of solvents	various
$\mathcal{S}$	set of all feasible solvent structures	-
$t$	time	min
$T$	temperature	$^{\circ}\text{C}$ , K
$T_b$	boiling temperature	$^{\circ}\text{C}$ , K
$T_m$	melting temperature	$^{\circ}\text{C}$ , K
$\mathbf{t}$	vector of thermodynamic properties	various
$\mathcal{T}$	set of all thermodynamic properties	-
$x$	molar fraction	-
$\mathbf{x}$	vector of continuous variables	various
$\mathcal{X}$	set of all continuous variables	-
$Y$	yield	wt. %
$\mathbf{y}$	vector of variables describing the molecular structure	various
$\mathcal{Y}$	set of all variables describing the molecular structure	-
$w$	weight fraction	-

## Superscripts

org	organic phase
0	reference state
ex	excess
id	ideal
j	phase
*	pseudo
$\infty$	infinite dilution
$\bar{\square}$	average

## Subscripts

75 vol.%	75% based on volume
96 vol.%	96% based on volume
0	initial
b	biomolecules
c	chemical species
car	carotenoids
carb	carbohydrates
C	cellulose
dry	dried biomass
extr	extraction
gen	generation
H	hemicellulose
IS	internal standard
L	lignin
NL	neutral lipids
par	parents
PL	polar lipids
pop	population
prod	product
prot	protein
raw	raw biomass (including hydration, ash, extractives)
reactive	reactive functional groups
ref	reference
s	solvent
sel	selection

## Physical constants

$e$	elementary charge	$1.6021766 \cdot 10^{-19} \text{ C}$
$R$	universal gas constant	$8.3144626 \text{ J mol}^{-1} \text{ K}^{-1}$

## Acronyms and abbreviations

$T_b$	boiling temperature
$T_m$	melting temperature

2-BuOH	2-butanol
2-BuOH <sub>75</sub> vol.%	75 vol.% 2-butanol
2-BuOH <sub>79</sub> vol.%	79 vol.% 2-butanol
2-MeTHF	2-methyltetrahydrofuran
4-(2-HE)morpholine	4-(2-hydroxyethyl)morpholine
5-Br-1-Me-1H-imidazole	5-bromo-1-methyl-1H-imidazole
[BMIM]	1-butyl-3-methylimidazolium
[BTMP]	bis(2,4,4-trimethylpentyl)phosphinate
[Chol]	choline
[DBP]	dibutyl phosphate
[Dec]	decanoate
[DEP]	diethyl phosphate
[DMP]	dimethyl phosphate
[EMIM]	1-ethyl-3-methylimidazolium
[MMIM]	1,3-dimethylimidazolium
[OAc]	acetate
[OH]	hydroxide
[P666 14]	trihexyltetradecylphosphonium
CO <sub>2</sub> -eq	CO <sub>2</sub> equivalent
β-Car	β-carotene
<i>P. tricornutum</i>	<i>Phaeodactylum tricornutum</i>
AAF	aldehyde-assisted fractionation
ABS	aqueous biphasic systems
BHT	butylhydroxytoluol
BVE	butyl vinyl ether
C16:0	palmitic acid
C16:1	palmitoleic acid
CAMSD	computer-aided molecular screening and design
Chl a	chlorophyll a
Chl c	chlorophyll c
COSMO	COnductor like Screening MOdel
COSMO-RS	COnductor like Screening MOdel for Real Solvents
CPME	cyclopentyl methyl ether
CSM	continuum solvation model
Ddx	diadinoxanthin
DEGDEE	diethylene glycol diethyl ether



DEGDME	diethylene glycol dimethyl ether
DEMP	diethyl ethylphosphonate
DEMP	diethyl methylphosphonate
DES	deep eutectic solvent
DFT	density functional theory
DHA	docosahexaenoic acid
DMM-sulfonamide	n,n-dimethylmethanesulfonamide
DMMP	dimethyl methylphosphonate
DMSO	diethylsulfoxide
DMSO	dimethylsulfoxide
DPX	dipropylxylose
Dx	diatoxanthin
EHS	environmental, health, and safety
EoS	equation of state
EPA	eicosapentaenoic acid
EtOH	ethanol
EtOH <sub>96 vol.%</sub>	96 vol.% ethanol
FAME	fatty acid methyl ester
FCP	fucoxanthin-chlorophyll a/c-binding protein
FID	flame-ionisation detection
FP	flash point
Fx	fucoxanthin
G	guaiacyl
GA	genetic algorithm
GAN	generative adversarial network
GC	gas chromatography
GHG	greenhouse gas
GNN	graph neural network
H	<i>p</i> -hydroxyphenyl
HBA	hydrogen bond acceptor
HBD	hydrogen bond donor
HPLC	high-pressure liquid chromatography
HTL	hydrothermal liquefaction
IBVE	isobutyl vinyl ether
IG	integrated gradient
IL	ionic liquid

IPCC	Intergovernmental Panel on Climate Change
LCA	life cycle analysis
LLE	liquid-liquid-equilibrium
MAE	mean absolute error
MAL	mild acidolysis lignin
MI(N)LP	mixed-integer (non)linear program
ML	machine learning
MUFA	monounsaturated fatty acid
NL	neutral lipid
NMMO	N-methylmorpholine-N-oxide
NRTL	Non-Random-Two-Liquid
NWM	non water-miscible
PBR	photobioreactor
PC	phosphatidylcholine
PC-SAFT	perturbed-chain statistical associating fluid theory
PFE	pressurised fluid extraction
PL	polar lipid
PUFA	polyunsaturated fatty acid
PVE	propyl vinyl ether
PWM	partially water-miscible
QC	quantum chemistry
QM	quantum mechanical
QSAR	quantitative structure-activity relationship
QSPR	quantitative structure-property relationship
REACH	Registration, Evaluation, and Authorization of Chemicals
S	syngas
SAS	synthetic accessibility score
SDS	safety data sheet
SELFIES	Self-Referencing Embedded Strings
SFA	saturated fatty acid
SFE	supercritical fluid extraction
SLE	solid-liquid-equilibrium
SMARTS	SMILES arbitrary target specification
SMILES	simplified molecular-input line-entry system
SQDG	sulfoquinovosyl diacylglycerol
t-SNE	t-distributed stochastic neighbour embedding

TAG	triacylglycerol
UNIFAC	universal functional activity coefficient
UNIQUAC	Universal Quasichemical
VLE	vapour-liquid-equilibrium
WFT	wavefunction theory
WM	water-miscible



# 1 | Introduction and motivation

## 1.1 Planetary boundaries and the chemical industry

Currently, human activities massively exceed planet Earth's resources. In 2009, Rockström et al. identified nine planetary boundaries that are crucial for maintaining a stable and resilient Earth system: climate change, ocean acidification, stratospheric ozone depletion, biogeochemical flows (especially nitrogen and phosphorous cycles), freshwater use, land system change, biosphere integrity, atmospheric aerosol loadings, and novel entities such as pollutants.<sup>1</sup> Especially crossing the core boundaries - climate change, biosphere integrity, and spread of pollutants - poses a high risk of triggering irreversible tipping points and pushing the Earth system into a new state with adverse effects on the remaining boundaries.<sup>2</sup> Six out of nine boundaries are presently breached, and two further boundaries are either regionally transgressed or close to being exceeded.<sup>2</sup>

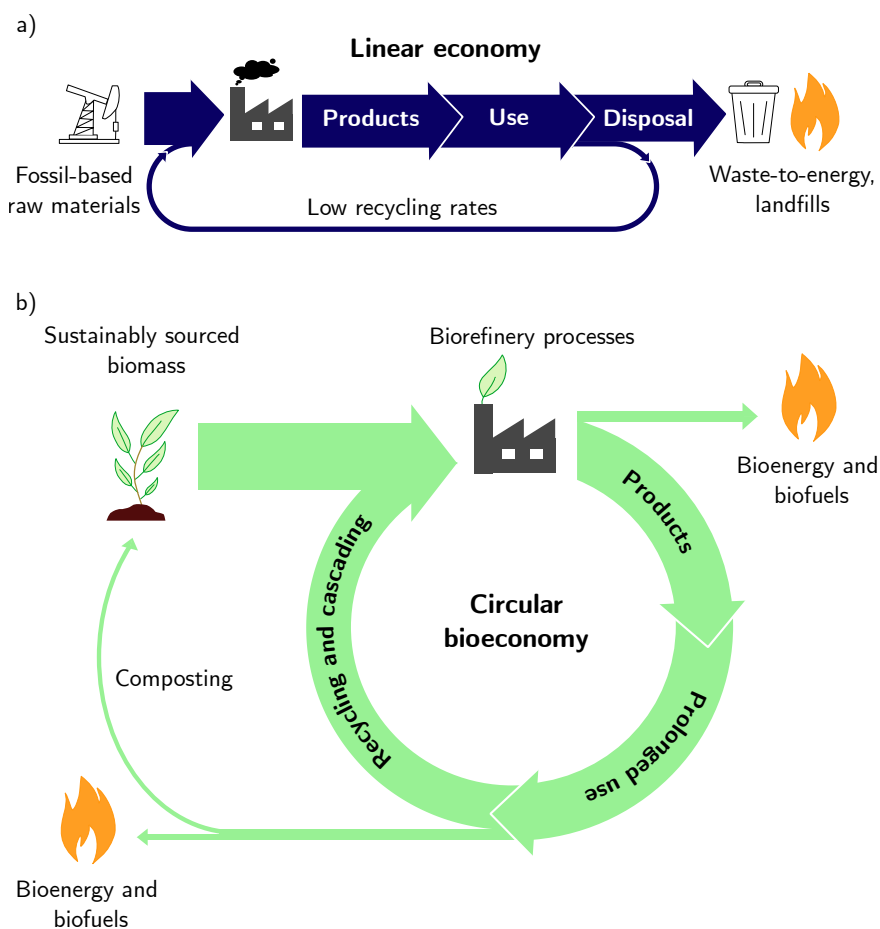
Greenhouse gases (GHGs), such as CO<sub>2</sub>, CH<sub>4</sub>, and NO<sub>2</sub>, are drivers of climate change.<sup>3-6</sup> Already now, with a global surface temperature increase of 1.1 °C compared to pre-industrial levels (1850-1900), climate change is causing weather and climate extremes in every region across the planet, endangering food and water security, human health, and damaging economies.<sup>7</sup> Studies of the Intergovernmental Panel on Climate Change (IPCC) revealed that if the temperature increase is held below 1.5 °C, adverse impacts of climate change can be limited.<sup>7</sup> Global policymakers signed the Paris Agreement in 2015, admitting to restrict the global temperature surge from the industrial revolution to 2100 to 2 °C and pursuing efforts to limit the increase further to 1.5 °C.<sup>8</sup> If GHG emissions, frequently expressed in terms of CO<sub>2</sub> equivalents (CO<sub>2</sub>-eqs), continue to follow current trends, the annual emissions of CO<sub>2</sub>-eqs will rise to ca. 60 Gt by 2050.<sup>7</sup> However, to limit the temperature increase to 1.5 °C, the GHG emissions must be reduced below 31 GtCO<sub>2</sub>a<sup>-1</sup> until 2030 and to 10 GtCO<sub>2</sub>a<sup>-1</sup> until 2050 (which equals net zero emissions).<sup>7</sup>

Products of the chemical industry are fundamental for our current lifestyle and present in 95% of all manufactured goods.<sup>9</sup> At the same time, the chemical industry is the third largest contributor to global GHG emissions (935 Mt direct CO<sub>2</sub> emissions in 2022) and

is the most energy-intensive industry sector.<sup>10,11</sup> The main energy supply is based on fossil resources and even 89% of the raw material streams comprise mineral oil and natural gas.<sup>12,13</sup> In view of climate change and depleting fossil resources, the chemical industry requires a fundamental transformation. Several strategies to develop a sustainable chemical industry were recently elaborated.<sup>14–17</sup> A pivotal approach is defossilisation in which both raw material and energy supply transition towards renewable resources. In the envisaged defossilised processes, biomass and renewable energy-based chemicals („green“ hydrogen, methanol, or ammonia) constitute the raw material base.

## 1.2 The need for a circular bioeconomy

Current patterns of production and consumption are predominantly linear: Primary material is sourced from Earth, subsequently converted into products, and finally disposed as waste (Fig. 1.1 a).<sup>18</sup> A linear economy not only contributes to exceeding the planetary boundaries<sup>18</sup> but also leads to a scarcity of critical raw materials.<sup>19</sup> To stop the current overexploitation of Earth's resources, a transition from the linear „take-make-dispose“ approach towards a circular economy is envisioned and promoted by policymakers. In a hypothetical fully circular scenario, all materials within the circular economy are reused or recycled. Consequently, no waste is generated, eliminating the demand for new primary materials. In practice, this fully circular scenario is limited by a lack of efficient recycling routes, energy-intensive recycling processes, and non-recyclable product design.<sup>20</sup> The Circularity Gap Report revealed that by increasing current circularity from 7 to 17%, 39% of the global GHG emissions could be mitigated by 2032.<sup>18,21</sup> Hence, even a slight increase in circularity is already an effective contribution to GHG mitigation.



**Fig. 1.1** Overview of a) a linear economy, and b) a circular bioeconomy. Adapted from Stegmann et al.<sup>22</sup>

To achieve defossilisation, the chemical industry needs to switch to a sustainable raw material base. Here, biomass as a renewable C-source offers conversion into a variety of chemicals and value-added products.<sup>15,23,24</sup> In this manner, fossil resources utilised as raw materials can be (at least partially) replaced by the use of biomass. Policymakers actively promote a sustainable transition by developing bioeconomy strategies.<sup>25</sup> The bioeconomy concept comprises the “production, exploitation, and use of biological resources, processes, and systems to provide products, processes and services across all economic sectors”.<sup>26</sup>

However, a bioeconomy implemented as a linear economy would increase the demand for biomass and consequently aggravate adverse environmental impacts (e.g. deforestation) and intensify competition for land use (e.g. cultivation of biomass for food vs. construction of wind farms for renewable electricity).<sup>22,27</sup> Additionally, resource conflicts between different industry sectors might intensify.<sup>27</sup> The projected biomass demand will significantly

surpass the amount of biomass that can be sustainably sourced.<sup>28</sup> Hence, although biomass is renewable, it is a limited resource that should be utilised in a resource-efficient manner. In contrast to a linear bioeconomy, a circular bioeconomy valorises biomass for a variety of products which are ideally reused and recycled multiple times before being composted at the end of their lifecycle (Fig. 1.1 b).<sup>22,29,30</sup>

### 1.3 Biorefineries as drivers of a circular bioeconomy

Biorefineries lie at the core of a circular bioeconomy aiming at sustainable conversion of biomass into a broad spectrum of marketable products and energy.<sup>31,32</sup> Beyond this primary goal, there exist various definitions for the term „biorefinery“ within the scientific literature and policy frameworks, each with slightly different nuances.<sup>22,33,34</sup>

First, the term „biomass“ should be further discussed within the context of biorefining. Biomass encompasses all organic material that is produced from living organisms, such as plants, animals, and microalgae. The most abundant source of terrestrial biomass comprises lignocellulosic feedstocks, such as straw, grasses, wood, agricultural residues. Microalgae are the dominant source of aquatic biomass. However, by this definition also coal and petroleum could be considered as potential feedstocks, since they were formed from dead plants and animal biomass by geological processes over the course of millions of years. Also, biomass originating from old-growth forests, peat bogs, nature conservation areas, or edible biomass would be suitable feedstocks according to this definition. Hence, an additional criterion needs to be fulfilled: the biomass must be sustainably sourced.<sup>22</sup> Sustainable biomass sourcing takes the regeneration time of the biomass into account, respects the planetary boundaries, and considers potential resource conflicts. Within this thesis, lignocellulose and microalgae, as the dominant sources of terrestrial and aquatic biomass, respectively, are investigated as promising feedstocks for biorefineries that can be sustainably sourced. Considering the biorefinery within the broader context of a circular economy, also undesired side and waste streams containing organic material could serve as potential feedstocks.

On the process level, biorefineries consist of a primary and secondary refining system including several extraction and other separation units.<sup>33</sup> In the primary refining system, the biomass is either fractionated into its main macromolecular constituents (e.g. lignocellulose into lignin, cellulose, and hemicellulose sugars) or the biomass is converted into a product mixture with high energy-density (e.g. lignocellulose gasification to synthesis gas). In both cases, the resulting molecules represent platform molecules or intermediates. In the secondary refining step, these intermediates are converted into the desired end-products.<sup>35</sup>



Typical biorefinery products include pharmaceuticals, chemicals, materials, food, feed, fuels, and energy. Biorefineries are classified according to the utilised feedstocks such as lignocellulose (straw, grasses, wood, agricultural residues) microalgae, or organic waste (e.g. spent cooking oil), the generated platform molecules (e.g. C<sub>5</sub> sugars, C<sub>6</sub> sugars, lignin, pyrolysis liquids, oils), the products (e.g. energy, fuels, ethanol, polymers) or the processing methods (e.g. thermochemical, mechanical, biochemical, chemical, or mechanical).<sup>34</sup>

In contrast to traditional chemical processes that operate with fossil-derived, clearly defined, and pure chemicals, biomass is a complex mixture of interacting biomolecules which poses challenges for further processing. Each bio-based feedstock is characterised by a distinct biochemical composition including carbohydrates, proteins, lipids, value-added molecules (e.g. pigments), and other biopolymers (e.g. lignin). The biochemical composition strongly influences the (physico)chemical and mechanical properties of the biomass. Furthermore, the chemical and thermal stability of the native biomolecules and the formed intermediates represent additional constraints for the biorefinery design. The connection between the biochemical structure and commonly applied process conditions is exemplified by lignocellulose and microalgae-based biorefineries. Both feedstocks are fundamentally different on the cellular and the molecular level. Lignocellulose owes its mechanical strength and recalcitrance to its high amounts of cellulose and, thus, requires harsh processing conditions, commonly involving high temperatures, acidic or alkaline pH, and organic solvents.<sup>36</sup> Due to potential degradation reactions, the process conditions must be carefully optimised.<sup>37,38</sup> Unicellular microalgae, on the other hand, contain considerably less amounts of cellulose than wood. Several microalgal species are even free of cellulose.<sup>39</sup> Therefore, the rigidity of microalgal cells is comparably low, and milder pretreatment methods are applicable. Due to the presence of light-, heat- and pH-sensitive pigments, ambient temperatures are commonly preferred. As highlighted by these particular examples, the process design of biorefineries is highly dependent on the specific feedstock and requires custom refining methods. In general, a biorefinery process aims at breaking the bonds of the native biomass such that the desired intermediates or end-products are directly released or become accessible for further processing. At the same time, the functionality of these molecules with respect to the desired end-product should be maintained.

## 1.4 Solvent selection for biorefinery processes: State of the art and open questions

To obtain a broad spectrum of products from the biomass, efficient separation strategies for intermediates and end-products are essential.<sup>40</sup> Commonly applied techniques include combinations of solid-liquid extraction, liquid-liquid extraction, precipitation, filtration,

distillation, and chromatographic methods.<sup>40,41</sup> Many of these operations rely heavily on the use of organic solvents.<sup>40</sup> Although solvent selection is a molecular-level decision, it is impacting the overall process system: The selected solvent determines the yield of the target compounds, and affects the energy demand, the economic feasibility as well as GHG emissions.<sup>42</sup> Furthermore, solvents influence reaction rates, prevalent reaction mechanisms, and the activity of catalysts.<sup>43</sup> Many solvents have adverse environmental, health, and safety (EHS) properties and their synthesis and disposal is commonly associated with a high environmental burden.<sup>44</sup> The solvent use contributes to more than half of the life cycle GHG emissions in chemical and pharmaceutical processes.<sup>45,46</sup> Therefore, the solvent selection is highly important and should be considered in early stages of the biorefinery design.

Traditionally, solvent selection for biomass processing is mainly guided by experimental work, relying on knowledge about solvent polarity and proticity, as well as empirics to estimate the solubility of target molecules, such as the famous „like dissolves like“-principle, Kamlet-Taft<sup>47</sup> or Hansen solubility parameters<sup>48</sup>. However, experiments are time- and resource-consuming, limiting the number of solvents that can be experimentally investigated. In contrast, computational methods allow for high-throughput exploration of potential solvents, thus guiding the experimentalists towards the most promising solvent candidates. Furthermore, thermodynamic models offer deeper insights into solution mechanism compared to empirical rules. In recent years, computational solvent screening and design methods were developed for several processes in chemical engineering.<sup>42,49–51</sup> Computer-guided screening methods offer an automated, high-throughput screening of a pre-defined solvent database, thus providing a broad overview of the performance of commercially available solvents. In computational solvent design methods, molecular structures are optimised towards the desired properties, thus exploring novel solvents beyond a limited database.

Despite their immense potential, computer-guided methods are not commonly used for solvent selection problems in biomass fractionation processes. To date, computational solvent screening for biomass processing is limited to certain solvent classes, considering only a limited number of candidates.<sup>52–55</sup> In particular, computational solvent design methods have yet to be developed for biorefinery processes. Filling this methodological gap could unveil previously unexplored solvents that potentially improve existing biorefinery processes (e.g. by replacing toxic solvents, increasing yields, lowering the energy demand and GHG emissions) or foster the development of new, innovative biomass fractionation strategies.

## 1.5 Aims and outline of this work

This thesis aims to promote resource-efficient biomass valorisation in biorefineries through the development of computer-aided solvent selection and design methods. The developed methods must be customisable to accommodate various solvent requirements (e.g. high solubility of target biomolecules and benign EHS properties). Furthermore, these methods must navigate the solvent space efficiently to identify the most promising candidates for biomass fractionation. The developed computational methods are applied as a first-line tool in lignocellulose and microalgae-based biorefineries to focus the experimental search on the most promising solvent candidates. Subsequently, the applicability of these solvents for biomass fractionation is experimentally validated. Through a close combination of computational and experimental investigations, promising solvent candidates are explored to gain deeper insights into how specific solvent properties influence the process performance. This interdisciplinary endeavour offers a unique opportunity to improve existing biorefinery processes and to facilitate the development of innovative biomass fractionation strategies.

The thesis is structured as follows: Chapter 2 and 3 provide fundamental knowledge about lignocellulose and microalgae processing, respectively. These feedstocks were chosen due to their abundance and their interesting product spectra. Furthermore, the choice of the feedstocks was motivated by their fundamentally different macromolecular composition which render these types of biomass ideal to thoroughly evaluate the applicability of the developed solvent selection methods. Both chapters start with a detailed description of the biochemical structure of the feedstocks. The biochemical structure of the biomass and the desired biorefinery end-products are closely related to the process conditions applied in the biorefinery. The link between feedstock biochemistry and the process level leads to a review of industrially relevant processing methods and state-of-the-art research. Chapter 4 reviews state-of-the-art methods for computer-guided solvent screening and design, as well as models for predicting solubilities, phase equilibria, and EHS properties. To fill the gap of computational solvent selection methods applicable to biomass fractionation processes, Chapter 5 proceeds with the development and the derivation of the computational screening and design methods, exceeding the state of the art. First, a solvent screening method, allowing to search a database of more than 8000 potential solvents with respect to thermodynamic and EHS properties, is presented. To enable the search for potential solvents beyond the predefined database, a graph-based genetic algorithm for solvent design is developed. In Chapter 6, the solvent screening method developed in this work is applied for solvent selection in lignocellulose-based biorefineries to provide an overview of potential solvent candidates and to identify promising, but not yet established solvents. Since in

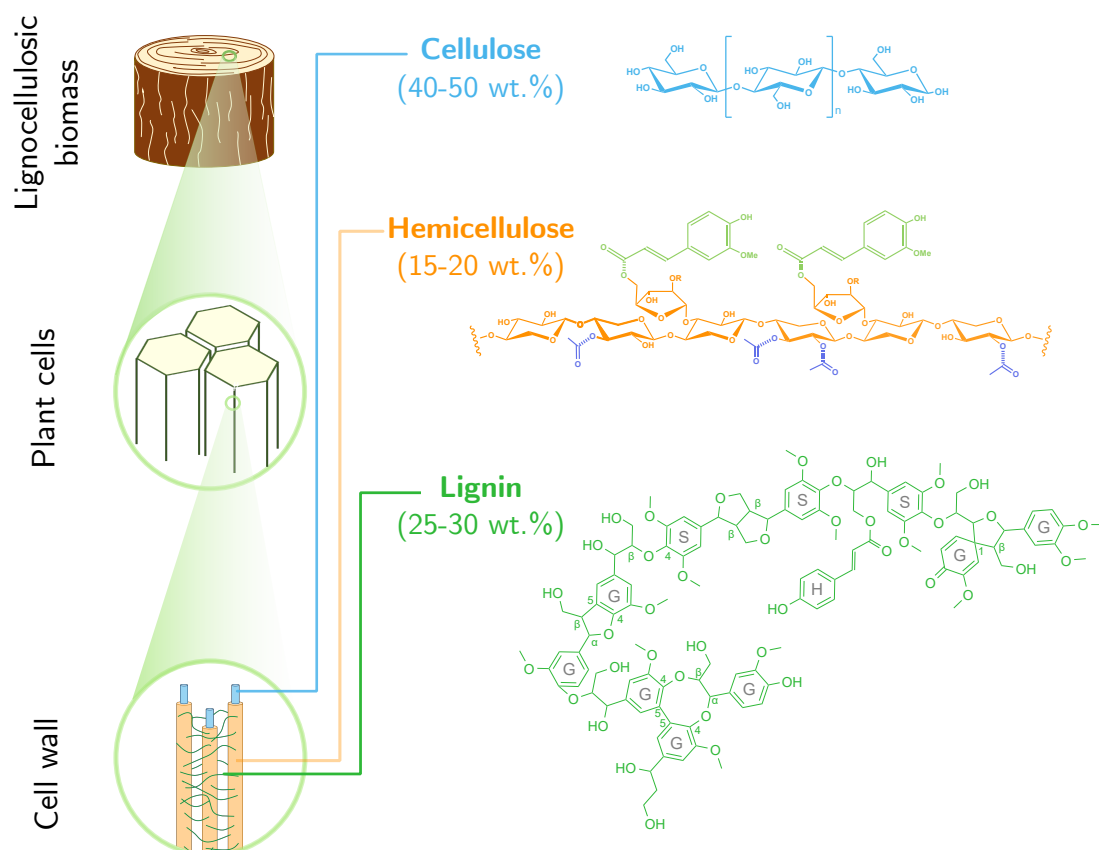
aldehyde-assisted fractionation (AAF), an innovative lignin-first approach, additional stabilising reagents considerably narrow the search space, the developed solvent design method is applied to generate tailor-made solvents. Especially the effect of structural solvent features on the lignin solubility, one of the main selection criteria in lignocellulose fractionation and lignin valorisation, are investigated. In Chapter 7, the developed solvent screening approach identified solvents applicable for the fractionation of undried microalgal biomass of the model species *Phaeodactylum tricornutum* (*P. tricornutum*), thus eliminating the energy-intensive biomass drying step. Based on the identified solvents, novel strategies for microalgal biorefining are explored. For both feedstocks, lignocellulose and microalgae, comprehensive experimental tests validate the applicability of the identified solvents for biomass fractionation. The thesis concludes with Chapter 8, which provides final remarks and highlights opportunities for further research.

## 2 | Lignocellulose processing

Lignocellulosic biomass (e.g. wood, straw, and grasses) represents the most abundant terrestrial source of renewable carbon.<sup>56</sup> Being renewable, widely available and inedible, lignocellulose is an attractive feedstock for biorefineries.<sup>36</sup> Since the valorisation of waste streams is particularly appealing in the context of a circular bioeconomy, lignocellulose biorefineries in Germany mainly utilise agricultural residues (cereal and corn straw), and residual wood (forest wood, poplar short-rotation wood).<sup>33</sup> The following chapter provides an overview over state-of-the-art processing technologies for lignocellulosic biomass. Since the biochemical structure of the biomass is directly related to the required process conditions, this chapter starts with a description of the biomass composition of lignocellulose (Section 2.1). Subsequently, industrially relevant methods and recently developed technologies for lignocellulose processing are reviewed (Section 2.2).

### 2.1 Biomass composition and molecular structure of lignocellulose

Lignocellulose consists of the three major fractions cellulose (30-50 wt.%<sub>dry</sub>), hemicellulose (20-30 wt.%<sub>dry</sub>), and lignin (15-30 wt.%<sub>dry</sub>).<sup>38,56</sup> These three biopolymers are tightly interconnected and form the lignin-carbohydrate complex located in the plant cell walls (Fig. 2.1).<sup>57</sup> The chemical structure renders lignocellulosic biomass resistant to (bio)chemical conversion.<sup>36,58</sup> This phenomenon, also known as recalcitrance, provides structural integrity and protects the plant against biotic and abiotic stressors.<sup>59</sup> Due to the recalcitrance of the feedstock, lignocellulose processing requires harsh process conditions including high temperatures, extreme pH, organic solvents, and catalysts.<sup>36</sup> The chemical structure and the resulting properties of lignocellulose are described in the following sections. Furthermore, potential intermediates and end products of each biomass fraction with significance for the chemical industry are reviewed.



**Fig. 2.1** Lignocellulosic biomass with the chemical structures of its three main components: cellulose, hemicellulose (here exemplified by xylan), and lignin. These three biopolymers are located in the cell wall where cellulose fibrils are surrounded by a matrix of hemicellulose and lignin. These three main components form the lignin-carbohydrate complex. The structure of lignin is based on G-, S-, and H-units linked by bonding motifs as proposed by Ralph et al.<sup>60</sup>

### 2.1.1 Cellulose

Cellulose, the main constituent of wood, is a linear polymer composed of  $\beta$ -1,4-glycosidically linked glucose monomers with a degree of polymerisation of up to 10,000 units.<sup>61</sup> Due to its equatorially positioned hydroxyl groups, the cellulose chains are stabilised by inter- and intraunit hydrogen bonds.<sup>62</sup> Within the plant cell wall, multiple cellulose chains bundle together to form microfibrils.<sup>63</sup> The microfibrils are encapsulated by a protective matrix of lignin and hemicellulose.<sup>58</sup> In wood, the microfibrils are predominantly crystalline with several amorphous domains.<sup>64</sup> In the crystalline region, cellulose chains are arranged in sheets, forming hydrophilic and hydrophobic domains.<sup>62</sup> In the hydrophilic domains, the hydroxyl-rich edge of each sheet is exposed, whereas in the hydrophobic domains, the faces of the sheets are exposed.<sup>65</sup> As a result, cellulose does barely dissolve in commonly

applied polar organic solvents, including water, despite the abundance of polar hydroxyl groups.<sup>66</sup> For a long time, the recalcitrance of cellulose was attributed to extensive inter- and intramolecular hydrogen bonding. More recently, London dispersion and hydrophobic interactions were reported to have a more dominant contribution to the recalcitrant behaviour.<sup>62,65</sup>

The cellulose fraction is the main target of the pulp and paper industry to produce high-quality paper products. Furthermore, cellulose is processed into bio-based materials, such as thermoplastics (e.g. cellulose acetate),<sup>67</sup> coatings, and films applicable as food packaging<sup>68</sup>. To produce cellulose coatings and films, cellulose needs to be effectively dissolved. However, only few solvent systems are known that are capable to dissolve this recalcitrant polymer, such as N-methylmorpholine-N-oxide (NMMO), dimethylsulfoxide (DMSO) in combination with ammonium salts, ionic liquids (ILs), and aqueous alkaline solutions.<sup>69</sup> Alternatively, cellulose can be hydrolysed to glucose monomers under acidic conditions, or by enzymes. Together with monomeric sugars originating from the hemicellulose fraction (Section 2.1.2), the glucose monomers serve as a sugar platform to produce alcohols (e.g. EtOH, 2,3-butanediol, 1-butanol), ketones (e.g. acetone), and organic acids (e.g. lactic acid, malic acid, succinic acid).<sup>70</sup> Furthermore, these sugars can be converted to furfural and 5-hydroxymethyl furfural to produce chemicals (e.g. levulinic acid,  $\gamma$ -valerolactone) or precursors for polymer production (e.g.  $\epsilon$ -caprolactam, 2,5-furandicarboxylic acid).<sup>58,70–73</sup> However, the carbohydrate-derived compounds form humins as undesired by-products in acid environment. Humins are polymers with stable C-C bonds that reduce the selectivity towards the desired chemicals, lead to catalyst deactivation, and cause separation problems.<sup>74</sup>

### 2.1.2 Hemicellulose

The hemicellulose fraction has a variable monosaccharide composition and linkage pattern. This fraction is classified according to the dominant monomer sugar in the backbone as xylan, glucuronoxylan, arabinoxylan, or glucomannan, with xylan being the most commonly encountered type of hemicellulose.<sup>75,76</sup> Commonly, their polymer backbone is composed of  $\beta$ -1,4-linked monomeric sugars in an equatorial configuration with numerous branches.<sup>75</sup> The degree of polymerisation ranges between 50 and 300 and is therefore substantially lower than that of cellulose.<sup>75,77</sup> Hemicellulose forms a complex network to tether the cellulose microfibrils, thus, providing additional mechanical strength to the plant.<sup>75,78</sup> In contrast to the rigid cellulose, hemicelluloses are amorphous polymers that are susceptible to chemical attack and can be dissolved under mild conditions.

Under acidic conditions, hemicellulose depolymerises to its constituent monomer sugars and serves as sugar-platform for the production of various chemicals (Section 2.1.1). Furthermore, hemicellulose sugars can be converted into aprotic solvents,<sup>79</sup> or biodegradable polymers.<sup>80</sup>

### 2.1.3 Lignin

Due to its antimicrobial properties and its hydrophobicity, lignin acts as a barrier within the plant to protect the carbohydrates from degradation.<sup>81,82</sup> Lignin is a highly cross-linked polymer composed of aromatic guaiacyl (G), *p*-hydroxyphenyl (H), and syringyl (S) monomer units. Its molecular weight ranges between 2,500 and 15,000 g/mol<sup>57,83–85</sup> depending on the biomass source and the applied lignin isolation method. The abundance of the monomeric G, H, and S units varies considerably between different biomass sources.<sup>86</sup> While softwood lignin is predominantly composed of G-units, lignin from hardwood and herbaceous species contains both G- and S-units. Additionally, substantial amounts of H-units are uniquely found in herbaceous lignin.<sup>86</sup> The predominant bonding patterns linking the subunits together are ether motifs ( $\beta$ -O-4,  $\alpha$ -O-4, 4-O-5) and C-C bonds (5-5,  $\beta$ -5,  $\beta$ -1, and  $\beta$ - $\beta$ ). The  $\beta$ -O-4 linkage represents the most abundant bond type across different biomass sources (40-80% of all bonds).<sup>87–89</sup>

In the pulp and paper industry, lignin has long been considered as an undesired impurity and was mainly used for heat production due to its high heating value.<sup>90</sup> However, given that lignin is one of the few renewable sources of aromatic hydrocarbon, there is immense interest in lignin depolymerisation to substitute fossil feedstocks for the production of aromatic chemicals and for the use as a drop-in fuel.<sup>58,91,92</sup>

To isolate high amounts of lignin from the biomass, high temperatures, and alkaline or acidic conditions are commonly applied.<sup>36</sup> However, such conditions induce a combination of de- and repolymerisation reactions, resulting in highly degraded, condensed lignin (see Fig. 2.3 b for detailed reaction mechanism under acidic conditions).<sup>36</sup> These structurally modified lignins are also termed technical lignins. Since lignin's ether bonds are most easily cleaved, they are susceptible to chemical attack.<sup>87,88,93</sup> As a result, the  $\beta$ -O-4 content of technical lignins is only 3–10% due to the formation of stable interunit C-C bonds caused by the harsh processing conditions.<sup>94</sup>

Lignin depolymerisation is typically performed *via* hydrogenolysis.<sup>95</sup> For hydrogenolysis, lignin is dissolved in an organic solvent, a transition metal catalyst is added, and the mixture is pressurised with H<sub>2</sub> ( $T = 180 - 280$  °C,  $p_{H_2} = 1 - 40$  bar)<sup>95</sup> to cleave the ether bonds. Since, the main targets for lignin depolymerisation are the  $\beta$ -O-4 bonds, the monomer yield



obtained from technical lignins is low (ca. 8%).<sup>38</sup> Recent advances in lignin research show, that also the stable C-C bonds can be selectively cleaved by oxidation,<sup>96</sup> autooxidation,<sup>97</sup> or the use of advanced catalysts.<sup>98</sup> However, current research efforts continue to explore C-C cleavage in lignin.

Due to the low monomer yields, technical lignins are commonly used to produce bio-based materials, such as lignin coatings, films, fibres, nanoparticles, or thermoplastics. Most of these applications require homogeneous lignin solutions in the initial processing steps. Hence, solvents with high lignin solubility are required. The fabrication of lignin coatings and films requires solvents with high lignin solubilities, such as DMSO.<sup>99–101</sup> For lignin nanoparticle formation, anti-solvent precipitation is a commonly applied method that relies on the relative difference in lignin solubility between the applied solvents.<sup>102,103</sup> Furthermore, the isolated lignin can be separated into fractions of homogeneous molecular weight using stepwise solvent fractionation.<sup>104</sup> Some of the resulting fractions have appealing thermo-mechanical properties for application in lignin-based resins.<sup>105</sup> To isolate lignin with high  $\beta$ -O-4 content from biomass that is suitable for hydrogenolysis, advanced processing technologies are required (Section 2.2.4).

## 2.2 State-of-the-art lignocellulose processing

In the following sections, the most relevant methods for lignocellulose processing in industry and research are reviewed.

### 2.2.1 Pyrolysis

Pyrolysis is a thermochemical process that converts dried lignocellulose to a solid (charcoal), liquid (bio-oil), and gaseous product stream ( $T = 500 - 700$  °C) in the absence of  $O_2$ .<sup>106</sup> The gaseous products are typically combusted to provide heat for the process, or are recycled back into the pyrolysis reactor to support fluidisation.<sup>107</sup> The solid charcoal can be used as a fertiliser or as low-grade fuel.<sup>107</sup> The obtained bio-oil is a mixture of multiple chemicals, including hydrocarbons, phenolic compounds, furans, ketones, sugars, aldehydes, and acids.<sup>107</sup> The direct use of the bio-oil as a drop-in fuel is challenging and requires further upgrading due to its immiscibility with petroleum-based fuels. Furthermore, the bio-oil is instable upon storage and has a low heating value.<sup>106,108–112</sup> Other approaches envision the separation of the bio-oil compounds to produce various chemicals,<sup>113</sup> for which efficient separation methods are currently being researched.<sup>114</sup> Despite intensive research in this area, large-scale lignocellulose pyrolysis has not yet reached economic viability.<sup>115</sup>

### 2.2.2 Kraft pulping

Lignocellulose processing in pulp and paper mills aims to produce high quality pulps for the paper and cardboard industry. Here, the main process objective is to obtain a high-purity cellulose stream, while particularly lignin is treated as undesired impurity. Currently, more than 90% of all chemical pulps are produced by the Kraft process.<sup>116</sup> In the Kraft process, wood is treated with an aqueous solution of NaOH and Na<sub>2</sub>S, also termed white liquor ( $T = 160\text{ }^{\circ}\text{C}$ ,  $t = 120 - 180\text{ min}$ ,  $\text{pH} = 11 - 14$ ).<sup>91,117,118</sup> Thus, HS<sup>-</sup>-ions are formed, promoting delignification without simultaneously accelerating carbohydrate solubilisation.<sup>119</sup> This treatment results in a high-quality pulp for the production of paper products with high mechanical strength. Hence, the German word 'Kraft' was chosen for the process, which translates to power or strength.<sup>120</sup> Lignin and parts of the hemicellulose fraction are dissolved in the process liquor, also termed black liquor. However, under the described alkaline conditions, lignin forms reactive quinone methides that undergo repolymerisation reactions, leading to severe lignin degradation.<sup>36,91,121</sup> The black liquor is mostly incinerated, thus serving as inexpensive energy supply to fuel the process.<sup>91,116</sup> Alternatively, the condensed lignin can be precipitated from the black liquor by acidification.<sup>122,123</sup> The precipitated lignin, also termed Kraft lignin, has a low  $\beta$ -O-4 content and contains thiol groups which constraint effective hydrogenolysis to aromatic monomers. Therefore, Kraft lignin valorisation in materials, such as resins,<sup>124</sup> is currently the more promising route.

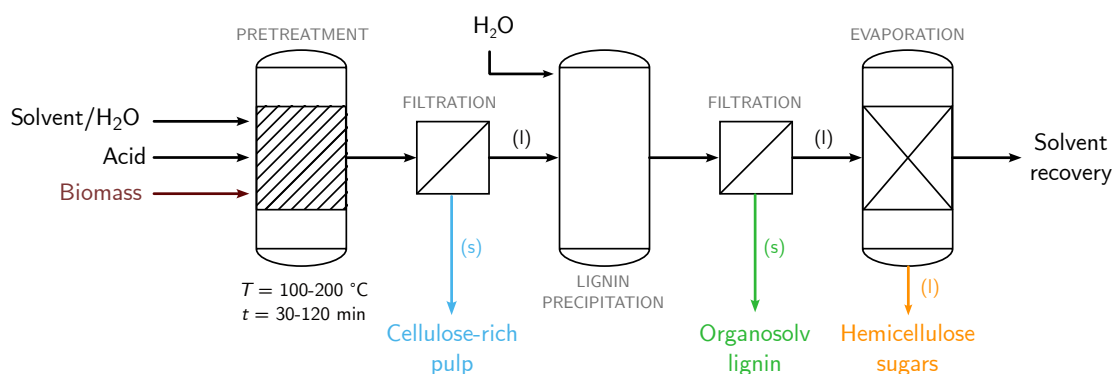
Furthermore, pulping causes a high environmental burden due to the high energy demand and pollutant emissions.<sup>125,126</sup> Since traditional paper mills produce pulp as their only end product, they are typically not considered as biorefineries *per se*.<sup>33</sup> However, methods for valorising the strongly degraded lignin are heavily researched. Connecting innovative strategies for lignin upgrading with traditional pulping processes provides interesting options to integrate the biorefinery concept into existing paper mills.

### 2.2.3 Organosolv processing

In organosolv processes, lignocellulosic biomass is treated with a mixture of an organic solvent, acid, and water ( $T = 100 - 200\text{ }^{\circ}\text{C}$ ,  $t = 30 - 120\text{ min}$ ,  $\text{pH} = 2 - 4$ ) resulting in three separate streams containing the cellulose-rich pulp, condensed organosolv lignin, and hemicellulose sugars.<sup>123,127-130</sup> Various process configurations have been developed, involving different organic solvents, product separation and solvent recycling systems (Fig. 2.2).<sup>130</sup> The acid catalyst promotes the hydrolytic cleavage of glycosidic bonds in hemicellulose. Thus, the hemicellulose depolymerises into its constituent monomer sugars that are solubilised in the process liquor. The use of the organic solvent facilitates lignin extraction. Ethanol (EtOH) and other low boiling alcohols are the most common

choice in organosolv processing due to their ease of recovery, low cost, and benign EHS properties.<sup>131–135</sup> Also polyols,<sup>136–139</sup> cyclic ethers,<sup>38,140</sup> organic acids,<sup>128,141–143</sup> and ketones<sup>144–147</sup> are applied. Ionosolv pulping is related to organosolv pulping but uses ILs instead of organic solvents.<sup>148</sup> After the pretreatment, the recalcitrant cellulose fraction remains as a solid residue (pulp) which can be easily removed from the process liquor by filtration or centrifugation.<sup>149</sup> The lignin fraction is separated from the process liquor by anti-solvent precipitation, commonly by water addition. The hemicellulose sugars remain in the aqueous process liquor and can be recovered by evaporation. In this manner, the biomass is fractionated into the three main components: Cellulose, hemicellulose sugars, and lignin,<sup>129</sup> thus reducing the complexity, heterogeneity, and recalcitrance of the biomass.

By organosolv pulping, each resulting stream becomes available for conversion into a wide range of products in line with the biorefinery concept. Cellulose and hemicellulose serve as a sugar platform to obtain various chemicals (Sections 2.1.1 and 2.1.2). However, due to its low  $\beta$ -O-4 content, the condensed organosolv lignin leads to low yields of aromatic lignin monomers after hydrogenolysis (Section 2.1.3). The acidic pretreatment conditions contribute to the formation of benzylium ions which ultimately lead to lignin condensation (see Fig. 2.3 b for details on the reaction mechanism). Since organosolv lignin has limited applicability for the production of aromatic monomers, its application in the materials sector is researched.<sup>150</sup>



**Fig. 2.2** Process flow diagram of organosolv pulping. As an alternative to filtration, centrifugation can be applied.

### 2.2.4 Lignin-first biorefineries

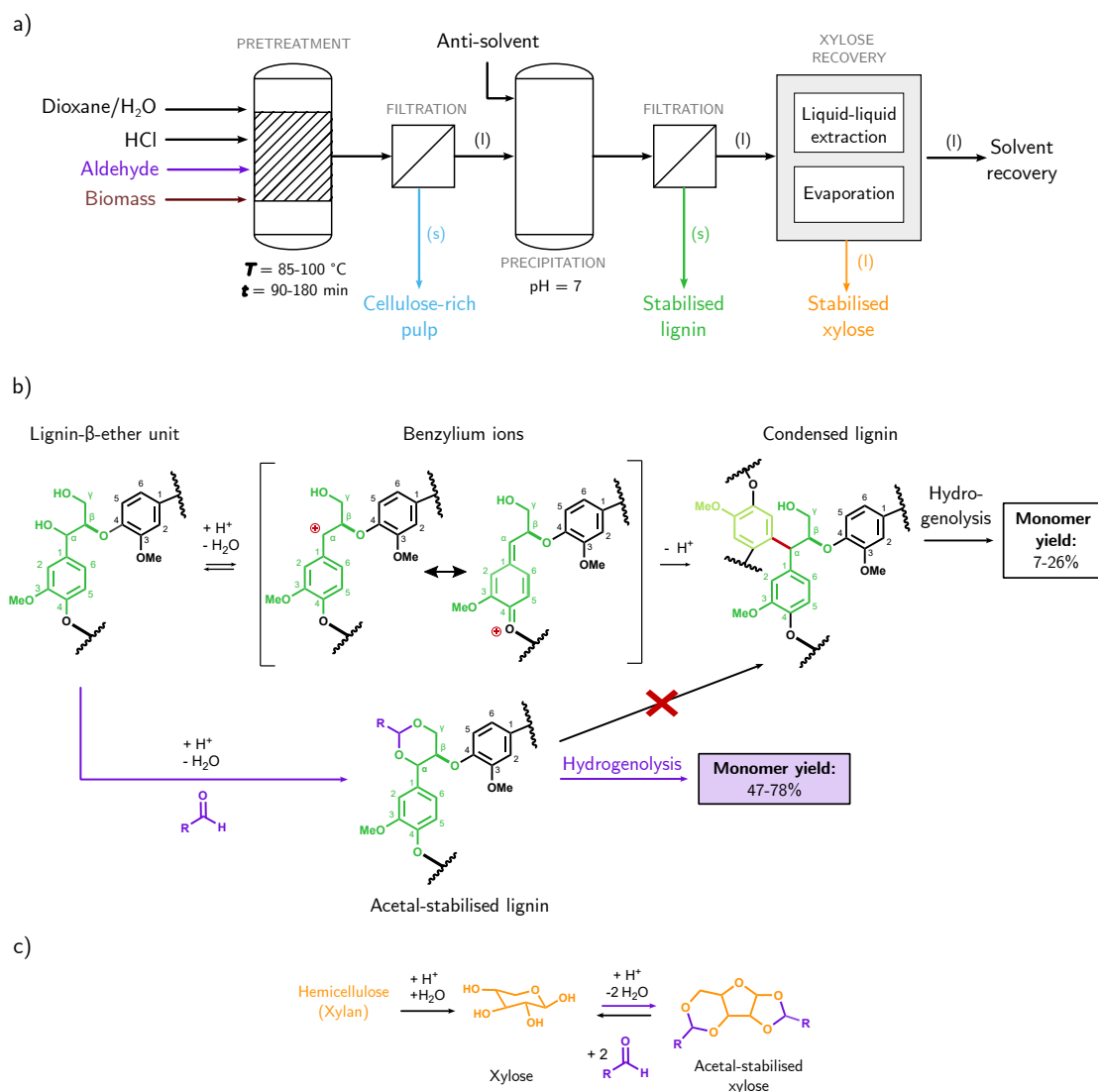
More recently, fractionation strategies with active lignin stabilisation, also termed lignin-first approaches, were developed. Lignin-first approaches aim to minimise lignin condensation during the treatment, thus enabling lignin upgrading to aromatic monomers.<sup>38,151–154</sup> One such approach is reductive catalytic fractionation, where lignocellulose is treated with an

organic solvent and a transition metal catalyst.<sup>151</sup> In the presence of a hydrogen donor, the dissolved lignin is directly converted to aromatic monomers by catalytic cleavage of the  $\beta$ -O-4 aryl ether motifs. Other approaches aim to stabilise the  $\beta$ -O-4 motifs by inserting protecting groups, e.g. by reaction with aldehydes<sup>38</sup> or ethylene glycol.<sup>152,154</sup>

#### 2.2.4.1 Aldehyde-assisted fractionation

AAF represents a protection chemistry-based approach where lignocellulosic biomass is treated with 1,4-dioxane, HCl, water, and an aldehyde (Fig. 2.3 a).<sup>38,155</sup> During pretreatment ( $T = 85 - 100$  °C,  $t = 90 - 180$  min), the aldehyde forms a stable acetal with the hydroxyl groups attached to the  $\alpha$ - , and  $\gamma$ - carbons (Fig. 2.3 b). The acetal formation outcompetes the protonation of the  $\alpha$ -carbon, thus minimising lignin condensation. Moreover, the aldehyde blocks reactive positions in the aromatic rings *ortho* and *para* to the methoxy groups. These protective mechanisms hinder the formation of interunit C-C bonds between the  $\alpha$ -carbons and the adjacent aromatic rings, and protect the  $\beta$ -O-4 motifs. Thus, the aldehyde-stabilised lignin can be effectively depolymerised *via* hydrogenolysis with near-theoretical monomer yields.<sup>38</sup> In addition to lignin, the aldehyde also stabilises the hemicellulose sugars. Under acidic conditions, hemicellulose is hydrolysed to monomer sugars, mainly xylose. The free hydroxyl groups of xylose react with the aldehyde to form stable acetals (Fig. 2.3 c), thus preventing humin formation.<sup>37,38</sup>

After pretreatment, the remaining cellulose-rich pulp is separated from the process liquor by filtration.<sup>155</sup> The stabilised lignin is precipitated from the process liquor by antisolvent addition (e.g. aqueous  $\text{NaHCO}_3$  solution, or n-hexane) and is separated by filtration.<sup>155</sup> The aldehyde-stabilised sugars are obtained either by liquid-liquid extraction, or by evaporation of the process liquor. The protected sugars are applied as bio-based, polar aprotic solvents,<sup>79</sup> or can be further processed to biodegradable plastics.<sup>80</sup> In this manner, three separate product streams, containing cellulose, stabilised lignin, and stabilised xylose are obtained, enabling wholistic valorisation of the biomass. However, the solvent choice needs to be critically assessed since 1,4-dioxane is toxic, mutagenic, and prone to peroxide formation, posing a risk for explosion.<sup>156</sup>



**Fig. 2.3** Overview over the AAF approach (lab-scale) based on Talebi Amiri et al.<sup>155</sup>

a) Process flow diagram. b) Reaction mechanism of a lignin fragment with and without aldehyde. For recovering the aldehyde-stabilised xylose, multiple process units are applicable as indicated by the gray rectangle. c) Reaction mechanism of the hemicellulose fraction (represented by xylan) during AAF.



### 3 | Microalgal processing

Microalgae are unicellular organisms with a size of around 2-50  $\mu\text{m}$  living in aqueous environments.<sup>157</sup> Similar to terrestrial plants, microalgae are able to grow by photosynthesis and produce almost half of the globally available  $\text{O}_2$ .<sup>158</sup> Microalgae do not require arable land for cultivation and grow at higher rates than terrestrial plants.<sup>159</sup> Some strains are able to grow on waste streams (photoautotrophically or mixotrophically).<sup>160,161</sup> In addition to microalgae, photosynthetic cyanobacteria, such as *Arthrospira platensis*, are potential feedstocks for biorefineries. In general, the processing of photosynthetic cyanobacteria is included under the umbrella-term „microalgal processing“. Microalgal biomass is mainly composed of proteins, carbohydrates, lipids, and pigments in varying ratios depending on the strain and the cultivation conditions.<sup>162</sup> These biomass fractions can be converted to produce biofuels, chemicals, biopolymers, or used as food or feed ingredients. Furthermore, several species produce high-value compounds, such as polyunsaturated fatty acids (PUFAs), pigments, antibodies, and other bioactive molecules, with applications as food supplements and in the pharmaceutical industry.<sup>163,164</sup> The ease of cultivation and the broad product spectrum render microalgae an interesting feedstock for biorefineries. At present, however, microalgae processing mainly targets lucrative niche markets, such as nutraceuticals, cosmetics, and specialised food.<sup>165,166</sup> For instance, *Schizochytrium* produces the PUFA docosahexaenoic acid (DHA) which is an important ingredient in infant formula.<sup>167</sup> *Haematococcus pluvialis* is a producer of the red pigment astaxanthin which has antioxidant properties and is marketed as a nutritional supplement.<sup>167</sup> The industrially most relevant strains are *Arthrospira platensis*, *Chlorella vulgaris*, and *Dunaliella salina*.<sup>167,168</sup> As these species are rich in protein, essential amino acids, essential fatty acids, antioxidants, and vitamins, their biomass is frequently marketed as „super food“. <sup>167,168</sup> The biomass of *Tetraselmis*, *Isochrysis*, *Pavlova*, and *Phaeodactylum* is commonly sold as animal feed.<sup>169–172</sup> However, the separation of multiple target compounds from microalgae remains challenging to date. Intensive research is necessary to spur the development of efficient extraction and fractionation methods, and to ultimately establish large-scale microalgal biorefineries.

The established model organism *P. tricornutum*<sup>173</sup> was chosen as a biomass source for

the aims of this work. *P. tricornutum* is an excellent feedstock for biorefinery processes due to its balanced ratio of proteins, carbohydrates, lipids, and pigments, including several high-value molecules. Section 3.1 describes the biomass composition of *P. tricornutum* and the macromolecular structure of each biomass fraction. Subsequently, established methods for microalgal processing and recently developed biorefinery approaches are reviewed in Section 3.2.

### 3.1 Biomass composition and molecular structure of the model alga *P. tricornutum*

The diatom *P. tricornutum* is composed of 18-54 wt.% proteins, 16-31 wt.% carbohydrates, 10-20 wt.% lipids, and 2-7 wt.% pigments, depending on the cultivation conditions.<sup>174-176</sup> *P. tricornutum* grows photoautotrophically, or mixotrophically on glucose, acetate, fructose, or glycerol.<sup>177</sup> Depending on the environmental conditions, *P. tricornutum* switches between a fusiform, triradiate, and oval morphotype.<sup>178</sup> The fusiform and oval morphotypes are dominant under commonly applied cultivation conditions.

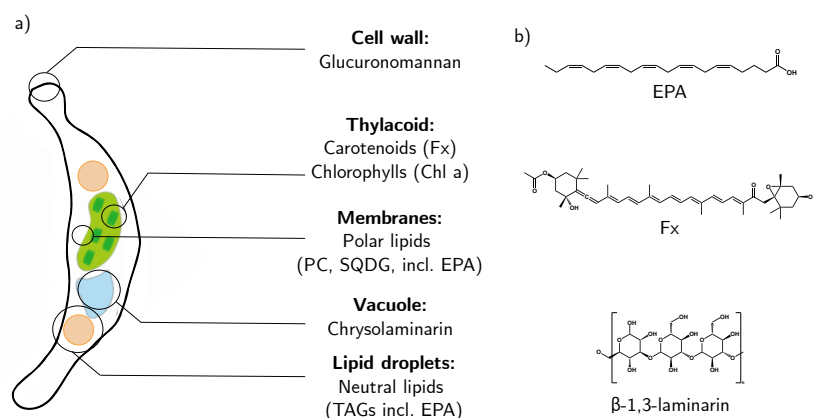
*P. tricornutum* produces several high-value compounds, such as fucoxanthin (Fx), a red carotenoid, and eicosapentaenoic acid (EPA), a PUFA with application in the nutraceutical industry. Fx contents between 0.5-5.92 wt.%<sup>179</sup> and EPA contents between 1.64-5 wt.%<sup>180-182</sup> on a dry matter basis were reported for *P. tricornutum*. Furthermore, *P. tricornutum* produces chrysolaminarin, a water-soluble storage carbohydrate, which is of high interest due to its bioactive and plant-protective properties.<sup>183-187</sup> To control the lipid:pigment ratio, the cultivation conditions can be adjusted, especially  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$ , and the light intensity.<sup>188-191</sup> In addition to the aforementioned value-added compounds, the protein and carbohydrate fractions can be utilised as animal feed, and for the production of biopolymers or biofuel.<sup>165</sup> The sale of *P. tricornutum* biomass for human nutrition is not yet approved by EU and USA authorities. However, EPA-rich oil from *P. tricornutum* is already being sold in the USA.<sup>192</sup>

Understanding the cellular structure of *P. tricornutum* and the chemical composition of the biomass fractions provides the basis for developing efficient fractionation approaches. The following sections describe the biochemical composition of the biomass fractions.

#### 3.1.1 Pigments

Photosynthetic organisms have evolved diverse light-harvesting complexes to capture solar energy.<sup>193</sup> In diatoms, fucoxanthin-chlorophyll a/c-binding proteins (FCPs) located in thylacoid membrane harvest the light that drives photosynthetic reactions (Fig. 3.1 a).<sup>193</sup> While





**Fig. 3.1** *P. triornutum* (fusiform shape) and its potential biorefinery products. a) Scheme of a *P. triornutum* cell (fusiform state). b) Chemical structures of Fx, EPA, and the  $\beta$ -1,3 backbone of chrysolaminarin.

chlorophyll a (Chl a) mainly absorbs violet-blue and orange-red light, Fx (Fig. 3.1 b) and chlorophyll c (Chl c) harvest blue-green light that reaches greater depths of water.<sup>194</sup> Absorbing excess light leads to photodamage and inhibition of photosynthetic activity.<sup>195,196</sup> In *P. triornutum*, the carotenoids Fx, diadinoxanthin (Ddx), diatoxanthin (Dx), and  $\beta$ -carotene ( $\beta$ -Car) reduce photodamage by scavenging reactive oxygen species and supporting the dissipation of the excess absorbed energy into heat by nonphotochemical quenching.<sup>194</sup> The most abundant carotenoid in *P. triornutum* is Fx and the main chlorophyll is Chl a.<sup>197</sup> Also the lipid droplets are enriched with Fx and  $\beta$ -Car with yet unknown functionality.<sup>198</sup> Pigments are sensitive to light, and elevated temperatures. Studies have shown that Fx is stable at a pH between 7-8 and temperatures below 40 °C.<sup>199</sup> Chl a degrades under acidic conditions or catalysed by chlorophyllases to several degradation products including pheophorbide a, pheophytin a and chlorophyllide a.<sup>200,201</sup>

### 3.1.2 Carbohydrates

Microalgae produce sugars by photosynthesis. Part of the produced sugar is stored in the vacuoles as chrysolaminarin.<sup>202,203</sup> Chrysolaminarin is a polysaccharide composed of  $\beta$ -1,3-glycosidically linked glucose units with  $\beta$ -1,6 branches and a molecular weight of about 10 kDa.<sup>202</sup> The dominant cell wall polysaccharide of *P. triornutum* is a sulphated  $\alpha$ -glucuronomannan.<sup>204–206</sup> Unlike other diatoms, the cell wall of *P. triornutum* is not surrounded by heavily silicified frustules.<sup>207</sup> Hence, *P. triornutum* cells can be more easily disintegrated.

### 3.1.3 Lipids

The lipids produced by *P. tricornutum* can be broadly divided into neutral lipids (NLs) and polar lipids (PLs). Almost all NLs of *P. tricornutum* are triacylglycerols (TAGs).<sup>198</sup> TAGs are synthesised at the endoplasmic reticulum, where three free fatty acids are attached to a glycerol backbone.<sup>177,208</sup> TAGs serve as energy storage<sup>209</sup> and are accumulated in lipid droplets located in the distal arms of fusiform cells.<sup>198,210</sup> When TAG accumulation is increased, chrysolaminarin synthesis is downregulated.<sup>177</sup> Palmitic acid (C16:0) and palmitoleic acid (C16:1) were reported as the most abundant fatty acids of *P. tricornutum*-TAGs.<sup>211</sup> PLs are composed of multiple fatty acids combined with polar functional groups, such as phosphatidylcholine (PC) and sulfoquinovosyl diacylglycerol (SQDG).<sup>198,211</sup> PLs are part of lipid organelle bilayers or precursor molecules in the lipid metabolism. Depending on the cultivation conditions, the high value fatty acid EPA was reported to be predominantly part of the TAGs<sup>212</sup> or predominantly incorporated in the PLs.<sup>211,213</sup> Dynamic changes in the EPA distribution during the cultivation stage were also reported.<sup>214</sup>

### 3.1.4 Proteins

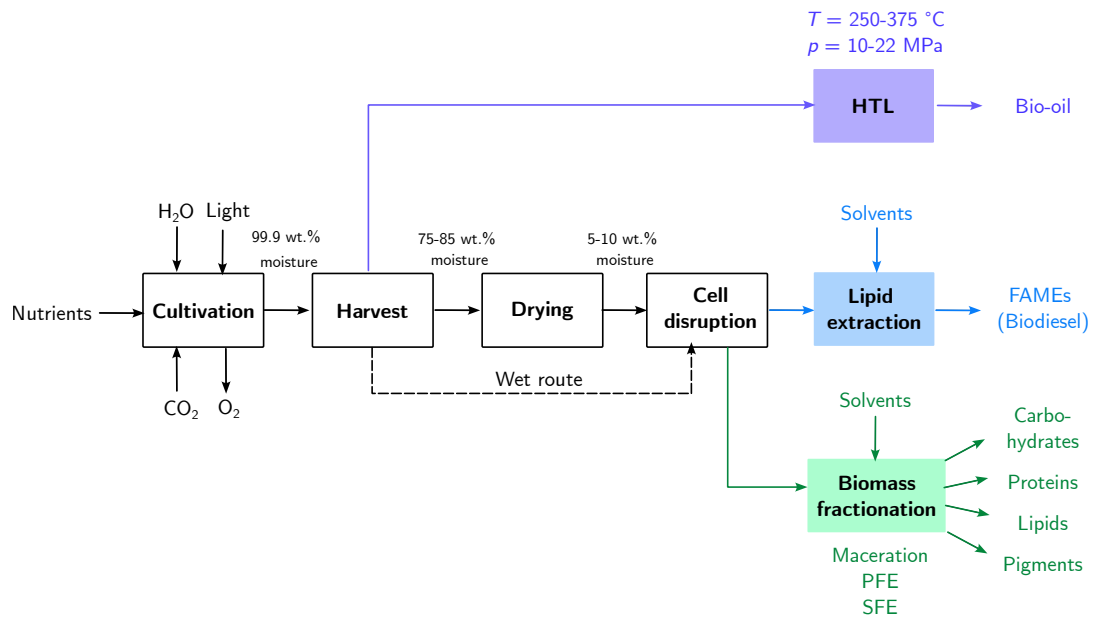
Proteins are uniquely folded biomolecules composed of peptide-bonded amino acid units. Proteins can be found at various locations in the cell as they play a central role in signaling, molecular transport, structural support, and the catalysis of metabolic reactions. In *P. tricornutum*, most of the proteins are water-soluble and have emulsifying properties with potential applications in the food industry.<sup>215</sup> Proteins denature under extreme pH, or elevated temperatures. In case that potential biorefinery products require the native protein structure, mild process conditions are required to prevent denaturation.

## 3.2 State-of-the-art microalgal processing

In a typical microalgal process, the biomass is harvested after cultivation to obtain a concentrated biomass suspension. There are many process routes to valorise the harvested biomass, e.g. for fuel production, or in a biorefinery process (Fig 3.2). In the following, the different process routes are described in detail.

### 3.2.1 Cultivation

For microalgal cultivation, light, CO<sub>2</sub>, and nutrient sources are required. Microalgal cultivation systems can be divided into open systems, such as raceway ponds, or closed photobioreactors (PBRs), such as bags, tubular reactors or flat panel reactors.<sup>216</sup> Natural sunlight can be used to illuminate the cultivation system but higher growth rates can be obtained by using with artificial light.<sup>216</sup> While open ponds have lower installation and operation costs



**Fig. 3.2** Process flow diagram for microalgal processing. After harvest, different process routes can be taken, such as hydrothermal liquefaction (HTL) (purple), lipid extraction for biodiesel production (blue), or biomass fractionation in a biorefinery (green).

compared to PBRs, there is a higher risk of contamination and the occurrence predators.<sup>217</sup> Closed systems are preferred over open systems when the production of high-value products compensates for the higher installation and operating costs.<sup>218</sup> The choice of the growth medium is highly dependent on the species and its individual demand of nutrients and trace elements. For *P. tricornutum*, high biomass, EPA, and Fx productivities were observed in flat-panel PBRs (PBRs) due to the low shear-stress and effective illumination.<sup>219</sup> Typical growth media (e.g.  $f/2$ <sup>220</sup> or the Mann and Myers medium<sup>221</sup>) contain carefully balanced concentrations of phosphates, nitrates, iron salts, and vitamins, and trace elements.

### 3.2.2 Harvest

After cultivation, the cell suspension is highly dilute, containing ca. 99 wt.% moisture.<sup>222</sup> To reduce the moisture content, the biomass is harvested from the aqueous growth medium by various techniques, including centrifugation, filtration or flocculation. After harvesting, the moisture content of the concentrated biomass suspension is 75 – 85 wt.%.<sup>222</sup> The choice of the harvesting method is highly dependent on the rigidity of the cell walls, the cell density, and the targeted moisture content.<sup>223</sup>

### 3.2.3 Drying

A drying step further reduces the moisture content to 5-10 wt.% and enhances the storability of the biomass.<sup>224,225</sup> Solar drying, convective drying, spray drying, and lyophilisation are suitable drying methods.<sup>225</sup> On the industrial scale, convective drying is most commonly applied (air,  $T = 50\text{ }^{\circ}\text{C}$ ).<sup>225</sup> Solar drying is the least expensive method and requires only solar energy. However, solar drying suffers from long drying times and a high area demand, and is therefore less relevant on an industrial scale.<sup>225</sup> Temperature-sensitive molecules as produced by microalgae, such as pigments or proteins, require mild drying methods, such as spray drying or lyophilisation, to prevent degradation. However, these methods have a high energy demand that can render the whole process economically infeasible.<sup>226</sup>

### 3.2.4 Cell disruption

After harvest (wet route) or drying (dry route, Fig. 3.2), a cell disruption step is commonly applied to facilitate the extraction of the target compounds. Solid-shear forces (e.g., bead mill, high speed homogenization) or liquid-shear forces (e.g. high pressure homogenization, microfluidization), can be leveraged to disrupt the algal cell wall. Other approaches for cell disruption are based on energy transfer through waves (e.g. ultrasonication, microwave), currents (e.g. pulsed electric field), heat (e.g. thermolysis, autoclaving), chemicals or enzymes.<sup>227</sup> Milling and high-pressure homogenisation are preferred on an industrial scale due to their high disruption efficiencies, high throughput, and simple temperature control.<sup>227</sup> Both methods are suitable for *P. tricornutum* biomass.<sup>228-230</sup> Furthermore, the cell disruption step can be integrated into the biomass fraction procedure for the targeted release of specific compounds.<sup>231</sup>

### 3.2.5 Biofuel production from microalgae

#### 3.2.5.1 Fatty acid methyl ester production from microalgal lipids

In the early 2000s, research and industry made large investments in microalgae cultivation for biodiesel production. The production of microalgae-based biodiesel requires several steps, including cultivation, harvest, drying, and cell disruption (Fig. 3.2). Subsequently, the microalgal TAGs are extracted with organic solvents, followed by transesterification of the lipids to fatty acid methyl esters (FAMES), the main component of biodiesel.<sup>232,233</sup> However, despite the strong interest from oil companies and extensive research, microalgae-based biodiesel could not reach economic feasibility since the costs of microalgal cultivation, low lipid productivities of wild-type microalgae, and the energy-intensive drying step led to high production costs.<sup>226,232,234-236</sup> A life-cycle assessment of Stephenson et al. revealed that especially the drying step is extremely energy-intensive and requires nearly 70% of the

overall energy input, leading to an overall negative energy balance.<sup>234</sup> Especially when the heat necessary for drying is generated from fossil fuels, the drying step leads to significant GHG emissions.<sup>234</sup> Thus, the production costs of the obtained biodiesel could not compete with inexpensive fossil-based fuels. Consequently, ExxonMobil, one of the last big oil companies that were still supporting microalgal fuel production, withdrew their investments.<sup>237</sup> Another drawback of FAME production from microalgae lies in the inefficient biomass utilisation. Here, the biorefinery concept was not fully adopted since only the TAGs were valorised to produce biodiesel as the main product. Further disadvantages were related to the organic solvents employed for the lipid extraction. The fossil-based n-hexane was proposed as a solvent due to its high lipid solubilities, low price, and ease of evaporation. However, n-hexane is highly toxic and its use is not recommended according to several solvent selection guides.<sup>238–240</sup>

### 3.2.5.2 Bio-oil production by hydrothermal liquefaction

Unlike the aforementioned biodiesel process, hydrothermal liquefaction (HTL) does not require biomass drying for fuel production. HTL is a thermochemical method converting wet algal biomass into bio-oil under subcritical conditions ( $T = 250 - 375$  °C,  $p = 10 - 22$  MPa, in the absence of  $O_2$ )<sup>241,242</sup>. Under these conditions, the entire biomass is converted into a liquid bio-oil without the need for organic solvents. However, due to the high heat demand, HTL has not yet reached industrial relevance<sup>241,242</sup>. Also in HTL, the biorefinery concept is not implemented, since the obtained bio-oil is the only main product. However, HTL can be integrated into an overall biorefinery process, e.g. for the valorisation of the residual biomass after the extraction of valuable compounds<sup>243</sup>.

### 3.2.6 Separation methods in microalgal biorefineries

Due to the limitations of the processes described in Section 3.2.5, more recently, the valorisation of microalgal biomass in a biorefinery has been proposed to increase resource efficiency. In the envisioned „zero-waste“ biorefinery, microalgal biomass is separated into its macromolecular fractions to produce high-value pigments and fatty acids, proteins, and carbohydrates.<sup>163–165,244</sup> Furthermore, wet algal biomass was proposed as a feedstock to eliminate the energy-demanding drying step.<sup>226,244,245</sup> Here, the wet algal paste obtained after the harvesting step (moisture content: 75–85 wt.%) should be used directly as a feedstock for biomass fractionation. The efficient extraction and separation of multiple target compounds is still challenging,<sup>163,165,246</sup> which is further complicated by the high moisture content of the biomass ranging between 75–85 wt.%.<sup>247,248</sup> Currently, microalgae processing in biorefineries is still at an early stage. Most commercial microalgal processing facilities do not fully adopt the biorefinery approach as they focus on a single high-value product

(e.g. pigments or PUFAs).<sup>165</sup> Therefore, extraction and separation strategies applicable to microalgal biorefineries are currently being researched. These methods are reviewed in the following sections.

### 3.2.6.1 Maceration

Maceration is one of the oldest and simplest techniques for solid-liquid extraction.<sup>249</sup> In this technique, microalgal biomass (wet suspension or dried powder) is mixed with organic solvents or a solvent mixture in a vessel to extract the target compounds. Shaking or stirring facilitates the extraction and reduces the incubation time.<sup>249</sup> Typically, maceration is performed at room temperature and atmospheric pressure for several minutes up to days.<sup>249</sup> Increased temperatures promote the extraction of the target compounds, however, the heat-sensitivity of pigments and proteins must be considered. The advantages of maceration lie in the simplicity of the process and low equipment costs.<sup>249</sup> However, high solvent consumption and long incubation times may be required to achieve high yields.

For lipid extraction on the lab scale, mixtures of chloroform and methanol are commonly chosen, as in the Bligh and Dyer<sup>250</sup> or Folch method.<sup>251</sup> However, these solvents are not recommended for industrial use due to high toxicity and other critical EHS properties.<sup>238–240</sup> For the extraction of chlorophylls and carotenoids, alcohols (e.g. methanol, EtOH) are applied,<sup>252–254</sup> or polar aprotic solvents (e.g. acetone, DMSO, dimethylformamide).<sup>253,254</sup> Kim et al. reported that EtOH extracted Fx from dried *P. tricornutum* biomass more efficiently than acetone and ethyl acetate<sup>255</sup>. EtOH and acetone can be synthesised from renewable resources and have benign EHS properties.<sup>238–240</sup>

Several studies have shown that wet extraction results in remarkably lower lipid yields than dry extraction.<sup>247,248</sup> This observation spurred the search for solvents that i) are effective on wet biomass, and ii) have benign EHS properties. It was frequently hypothesised that nonpolar solvents, such as hexane, cannot efficiently penetrate the polar moisture contained in the biomass, resulting in insufficient contact between the solvent and the non-polar target molecules.<sup>247,256,257</sup> Increased lipid yields were reported for mixtures of nonpolar and polar solvents, such as hexane/EtOH.<sup>248,256</sup> Furthermore, novel solvent classes such as ILs,<sup>258–261</sup> DESs,<sup>262,263</sup> or polarity-switchable solvents<sup>264–267</sup> were investigated for their applicability in extractions from wet and dried microalgal biomass as potentially benign alternatives to commonly applied, toxic solvents.

### 3.2.6.2 Supercritical fluid extraction and pressurised fluid extraction

Supercritical fluid extraction (SFE) is applied in the extraction of lipids and pigments from microalgae. In SFE, gases or liquids are used as solvents at temperatures and pressures

above their critical point.<sup>268,269</sup> The most commonly applied solvent in SFE is CO<sub>2</sub>, which is non-toxic, volatile, non-flammable, and has a low critical temperature.<sup>270</sup> CO<sub>2</sub> can be easily recovered by decompression which is beneficial for heat-sensitive molecules, such as pigments.<sup>271</sup> In this manner, no solvent traces remain in the final product. The low polarity of supercritical CO<sub>2</sub> renders this solvent particularly suitable for the extraction of nonpolar components, such as TAGs and carotenoids.<sup>272</sup> However, SFE using CO<sub>2</sub> was not efficient for wet biomass, as the nonpolar CO<sub>2</sub> could not sufficiently penetrate the wet biomass.<sup>180</sup> To increase the polarity, more polar co-solvents such as EtOH can be applied, as in the extraction of astaxanthin from *Haematococcus pluvialis*.<sup>273</sup>

Pressurised fluid extraction (PFE), also known as subcritical fluid extraction, involves the extraction of wet or dry algal biomass using gases or liquids below their critical point.<sup>274,275</sup> Derwenskus et al. showed that Fx and EPA can be efficiently extracted from wet *P. tricornutum* biomass using pressurised EtOH. Both high-value products were efficiently separated by precipitation of Fx and subsequent filtration.<sup>180,276</sup>

Both, PLE and SFE allow for rapid extraction of valuable compounds from microalgal biomass and facile recovery of the target compounds. However, both extraction methods require costly equipment and have a high energy demand caused by pressurisation.<sup>277–280</sup>

### 3.2.6.3 Methods for biomass fractionation

Several approaches for the fractionation of microalgal biomass are currently being investigated. One such approach involves sequential solid-liquid extractions of the biomass using solvents of different polarity. With this approach, lipids, carbohydrates, and proteins were sequentially extracted from dried *Scenedesmus obliquus* biomass.<sup>281</sup> However, the toxic solvents chloroform and methanol were applied to extract the lipids and should be replaced by more benign alternatives. In another study, pigments, lipids, carbohydrates and proteins were sequentially extracted from *Chlorella variabilis*.<sup>282</sup> Also here, toxic solvents were applied. Moreover, a sequence of SCF and PFE extractions was applied to first extract the lipid and pigment fractions, followed by carbohydrate extraction.<sup>283</sup> Furthermore, pulsed electric fields can be applied for the targeted release of the biomass fractions. Papachristou et al.<sup>284</sup> selectively released proteins and carbohydrates, before the lipid fraction was extracted with a hexane/EtOH mixture from *Scenedesmus almeriensis*.

In other studies, multiple target components were extracted from the biomass, and separated by sequential liquid-liquid extractions. Suarez Ruiz et al. demonstrated a multiproduct biorefinery approach for *Neochloris oleoabundans* by sequential separation steps using IL-based aqueous biphasic systems (ABS) and ultrafiltration.<sup>285</sup> ABSs consist of two

distinct aqueous solutions capable of forming two liquid phases when mixed beyond a critical concentration, including water-soluble polymers, salts, alcohols, micelles, or ILs.<sup>286,287</sup> In the context of microalgal biorefining, ABSs are especially useful when the functionality of the protein fraction needs to be preserved. ABSs were used for the separation of algal carbohydrates,<sup>288</sup> proteins and pigments,<sup>289</sup> and proteins and carbohydrates.<sup>290</sup>



## 4 | Computer-aided solvent screening and design: state of the art

When choosing the optimal molecule for a specific task, one of the first questions that can be asked is: How many organic molecules exist in total? The chemical space, describing the set of all potentially existing molecules, was estimated to comprise between  $10^{18}$  and  $10^{200}$  organic molecules (depending on the considered maximum molecule size and atom types).<sup>291,292</sup> To date, only a small fraction of the chemical space is synthetically available.  $2.2 \cdot 10^8$  molecules are listed in the ZINC15 database of which around half are commercially available.<sup>293</sup> Identifying molecules with desired properties for a specific application, such as solvents, catalysts or drugs, has been an active field of research for decades, involving laborious experimental procedures that are limited by temporal and human resources.<sup>294</sup> So far, solvent selection for lignocellulose and microalgal biorefineries was mainly based on experimental tests, not involving more than 10-15 solvents per study.<sup>100,101,247,248,295,296</sup> Considering the gigantic size of the chemical space, the need for systematic, computer-guided solvent selection becomes evident. Computer-aided molecular screening and design (CAMSD) methods aim to steer the search towards molecular structures with desired properties by using computational techniques, such as molecular modelling, property predictions, and mathematical optimisation.<sup>294</sup> Unlike direct problems, where properties and the performance are calculated based on a defined molecular structure, CAMSD algorithms aim to solve the inverse problem where, given a desired performance or property, the optimal molecular structure is sought after (Fig. 4.1). CAMSD algorithms consist of three fundamental elements: i) a measure for the performance of the screened or designed molecular structures, ii) property prediction models with sufficient accuracy and computational speed, and iii) algorithms for effective exploration of the chemical space. While screening methods search a defined database of molecular structures, design methods have the potential to drive the search towards completely undiscovered compounds. Both methods have the potential to transform current processing by proposing highly promising, so-far unexplored molecules for a given application.

In Section 4.1, the inverse problem is formulated as a mathematical optimisation problem. Suitable solution methods are explained in Section 4.2. Furthermore, property prediction

models especially important in the field of biomass processing are reviewed in Section 4.3. These property models require varying molecular representations which are explained in Section 4.4.

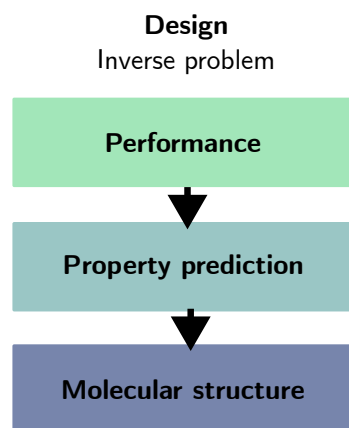


Fig. 4.1 Scheme of molecular design as an inverse problem.

## 4.1 The inverse problem as an optimisation problem

The inverse problem can be expressed as follows: Given a process (e.g. biorefinery process) that requires a molecule (e.g. solvent), find the optimal molecular structure minimising a cost function that is related to the performance in the considered process. This statement can be translated to the following optimisation problem<sup>297–299</sup>:

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} F(\mathbf{x}, \mathbf{y}) \quad \text{objective function} \quad (4.1)$$

$$\text{s.t.} \quad \mathbf{g}(\mathbf{y}) \leq \mathbf{0} \quad \text{structural constraints,} \quad (4.2)$$

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad \text{property constraints} \quad (4.3)$$

where the objective function  $F$  measures the performance of the molecules with structures  $\mathbf{y}$  depending on continuous variables  $\mathbf{x}$  (e.g. process conditions, such as temperature or pressure). The objective function  $F$  may contain one or multiple performance objectives. In screening methods, the vector  $\mathbf{y}$  often comprises complete molecular structures, whereas in molecular design methods,  $\mathbf{y}$  are molecular building blocks (e.g. functional groups). The continuous variables  $\mathbf{x}$  lie within the set  $\mathcal{X}$ . The set  $\mathcal{Y}$  represents the attainable chemical space. In screening methods, the attainable chemical space is given by a predefined database which is searched for the optimal candidate. In molecular design methods, the attainable space comprises all molecules that can be generated from the combination of the molecular

building blocks contained in  $\mathbf{y}$ . The choice of molecular building blocks are discrete decisions frequently expressed by integer variables. Property predictions are not only necessary to evaluate the performance of the molecules in  $F$  but are commonly applied to evaluate whether the property constraints  $\mathbf{h}$  are satisfied. In this manner, molecules with unsuitable properties (e.g. solvents being solid at process conditions) are effectively excluded to narrow the search space. The structural constraints  $\mathbf{g}$  ensure structural feasibility of the molecules and allow incorporating *a priori* knowledge regarding unsuitable functional groups.

## 4.2 Methods for chemical space exploration

The success of solving the inverse optimisation problem in Eqs. 4.1-4.3 highly depends on the applied method for chemical space exploration. In the following sections, state-of-the-art screening and design methods are reviewed.

### 4.2.1 Screening methods

In screening methods, the inverse problem is solved by systematically searching in a fixed set of molecular structures  $\mathcal{Y}$  to identify the molecule with optimal performance as measured by the objective function  $F$ . Hence, to solve the overall inverse problem, screening methods solve multiple direct problems. Here,  $\mathcal{Y}$  is given as a database containing molecular structures derived from expert knowledge. Molecules that do not satisfy the structural and property constraints  $\mathbf{g}$  and  $\mathbf{h}$  are excluded from  $\mathcal{Y}$  to narrow the search space. In this manner, the search is focused on the most promising structures included in the confined subspace  $\tilde{\mathcal{Y}}$ . For each  $\mathbf{y}$  in  $\tilde{\mathcal{Y}}$  the performance is measured by evaluating  $F$ . Based on the predicted performance, the molecular structures are ranked.<sup>50,51</sup> Frequently, the best performing molecules are selected for more detailed computational investigations or experimental validation. The choice of  $\mathcal{Y}$  is critical step for the success of screening methods. Since  $\mathcal{Y}$  does not comprise all molecules of the vast chemical space, the globally optimal molecular structure might be missing. To increase the probability of identifying molecular structures with near-optimal properties, it is recommended to screen a comprehensive, structurally diverse  $\mathcal{Y}$ .

Screening methods were recently applied for solvent selection in the field of chemical engineering<sup>50,300–303</sup> and for biomass processing. Motlagh et al. screened 352 ILs for the extraction of fatty acids from microalgae.<sup>53–55</sup> Several studies focused on the selection of ILs for lignin, cellulose, or hemicellulose sugars.<sup>52,304–312</sup> For both biomass sources, microalgae and lignocellulose, these studies commonly focused on single target biomolecules rather than incorporating the biorefinery approach into the solvent selection, or were limited to ILs. Therefore, there is a significant research gap tackling the solvent selection for biorefinery processes.

## 4.2.2 Molecular design methods

Unlike screening methods where the molecular structures remain unchanged, design methods modify the molecular structures in  $\mathbf{y}$  towards optimal performance evaluated by  $F$ .<sup>294</sup> Hence, in design methods,  $\mathcal{Y}$  is defined by the combination of all molecular building blocks in  $\mathbf{y}$ . In design methods, structural and property constraints  $\mathbf{g}$  and  $\mathbf{h}$  ensure the generation of molecules with desired structural features and properties. Hence, similar to screening methods,  $\mathcal{Y}$  is narrowed to  $\tilde{\mathcal{Y}}$ . However, in screening methods,  $\mathcal{Y}$  and  $\tilde{\mathcal{Y}}$  are considerably smaller sets. State-of-the-art molecular design methods are reviewed in the following sections.

### 4.2.2.1 Generate-and-test methods

In generate-and-test methods, a large number of novel molecular structures  $\mathbf{y}$  is created by enumerative combination of molecular building blocks. For each generated molecule, the objective function  $F$  is evaluated to predict the performance.<sup>313–315</sup> Increasing the number of different molecular building blocks leads to combinatorial explosion of the molecular design space. To limit the size of  $\mathcal{Y}$  and to lower the computational time, combination rules can be introduced.<sup>313</sup>

### 4.2.2.2 Deterministic optimisation

The inverse problem frequently involves several continuous variables (e.g. desired melting point  $T_m$  and boiling point  $T_b$  ranges), and integer variables (e.g. discrete decisions for given molecular building blocks) in combination with linear and/or nonlinear objective functions or constraints (e.g. thermodynamic models) (Eqs. 4.1–4.3). In such cases, the optimisation problem is formulated as a mixed-integer (non)linear program (MI(N)LP). Frequently, deterministic algorithms break the optimisation problem into smaller subproblems that can be more easily solved than the original problem,<sup>316</sup> e.g. by fixing variables, relaxing constraints, and linearising nonlinear functions. Examples for such solution strategies are outer approximation,<sup>317</sup> branch-and-bound,<sup>318,319</sup> generalised disjunctive programming, and interval analysis.<sup>320</sup> A comprehensive overview about deterministic optimisation methods in molecular design is provided by Papadopoulos et al.<sup>321</sup>

### 4.2.2.3 Genetic algorithms

Genetic algorithms (GAs) are a commonly chosen method for molecular design.<sup>322–326</sup> GAs are a metaheuristic inspired by Darwin's idea of biological evolution processes, such as selection, cross-over, and mutation. In GAs, the molecular structure is iteratively changed towards the desired properties. In each iteration  $i$ , the fittest molecules have a higher probability to be selected to create offspring molecules in the so-called cross-over step.<sup>327</sup> During cross-over, the selected parent molecules exchange molecular fragments with each other. These

fragments are recombined to generate children. The selection pressure drives the molecular population towards maximal fitness as measured by the objective function  $F(\mathbf{x}, \mathbf{y}_i)$ .<sup>328</sup> Additionally, random mutations on the molecular structure reduce the risk of premature convergence. The generation of new molecular structures  $\mathbf{y}_{i+1}$  is guided by their performance as measured by  $F(\mathbf{x}, \mathbf{y}_i)$ . This iterative procedure continues until a specified termination criterium is reached (e.g a finite number of iterations, or performance improvement close to a pre-defined threshold).

#### 4.2.2.4 Machine learning-based methods

In the recent years, machine learning (ML) approaches were developed for molecule generation and property prediction. This section focuses on the use of ML methods for molecule generation (see Section 4.3 for a review of ML models for property predictions). Frequently employed techniques for molecule generation include variational autoencoders,<sup>329,330</sup> generative adversarial networks,<sup>331,332</sup> and recurrent neural network.<sup>333</sup> These methods generate novel structures based on the training on extensive datasets. During the training phase, these models capture essential molecular features for generating new structures. To evaluate the molecules with respect to the desired properties, the ML models are usually combined with property prediction models.<sup>330</sup> The architecture of generative adversarial networks, variational autoencoders, and recurrent neural network is briefly explained in the following.

**4.2.2.4.1 Generative adversarial networks** Generative adversarial networks are built from two neural networks that are simultaneously trained in a competitive manner: a generator and a discriminator. The generator aims to create realistic molecules, while the discriminator aims to distinguish between real and generated data. The trained GANs produce novel structures that resemble the molecules of the training dataset.<sup>332,334</sup> To identify the most suitable molecule, the GAN can be combined with additional models for property predictions.<sup>335</sup>

**4.2.2.4.2 Variational autoencoders** Variational autoencoders consist of an encoding and a decoding network. The encoder maps a molecular representation into a lower-dimensional space, the so-called latent space, according to a probability distribution (usually the Gaussian distribution).<sup>336</sup> The decoder subsequently maps the latent vector to the original representation. In this way, essential molecular features are captured in the latent space, where molecules are represented as continuous and differentiable vectors residing on a probabilistic manifold.<sup>336</sup> Therefore, the latent space is especially interesting for molecular design, as it allows to identify molecules with the target properties using mathematical optimisation.<sup>330,337,338</sup>

**4.2.2.4.3 Recurrent neural network** Recurrent neural networks are typically applied in molecular design for processing and generating sequential molecular representations, such as simplified molecular-input line-entry system (SMILES) strings<sup>339</sup> (see Section 4.4 for information about molecular representations). In contrast to feedforward networks, where the information flows in a linear manner from input to the output prediction, RNNs contain cyclic connections allowing them to maintain information from previous input data.<sup>340</sup> The thus implemented memory allows RNNs to capture dependencies and patterns within the input sequence and generate novel SMILES strings. In combination with methods such as transfer learning or Monte Carlo tree search, RNNs can generate novel molecules with desired properties.<sup>340</sup>

## 4.3 Property predictions

In biorefinery processes, solvents with defined solubilities of biomolecules, phase behaviour, and benign EHS properties are sought after. Predicting these properties is a key step for solving the inverse problem (Eqs. 4.1-4.3). Since the property predictions are performed for each evaluation of the objective function  $F$  and the property constraints  $h$ , they largely determine the computational time, and thus, the size of the chemical space that can be explored. Therefore, property models with a suitable trade-off between accuracy and computational time should be selected. In the following sections, models for the prediction of thermodynamic properties and EHS properties that are important for the solvent selection in biorefineries are reviewed.

### 4.3.1 Thermodynamic models

On the thermodynamic level, several important process units of a biorefinery are governed by phase equilibria. In the initial stage of a biorefinery process, one or several target compounds are extracted from the biomass. Here, solvents with high solubility of these target compounds are commonly desired. In general, solubility is described by the amount of a solid solute that dissolves in the liquid solvents when the solid and liquid phase are in solid-liquid-equilibrium (SLE). Frequently, the obtained extract contains multiple target compounds or undesired impurities. In a later stage of the biorefinery process, the extracted compounds must be separated. Liquid-liquid extraction is a commonly applied method for this purpose. Here, the formation of two (or seldomly more) liquid phases is triggered by adding a solvent or a solvent mixture that is practically immiscible or only partially miscible with the solvent containing the target molecules. When the polarity difference between the target molecules is large enough, they partition to opposite phases and sufficient separation can be achieved. Typically, the liquid phases are in liquid-liquid-equilibrium (LLE). Another method to separate several target compounds is crystallisation which is based on SLEs.

Finally, the applied solvents should be recycled, e.g. by distillation, which exploits the vapour-liquid-equilibrium (VLE).

When multiple phases  $j = 1, 2, \dots, J$  irrespective of their state (solid, liquid, or gaseous) are in thermodynamic equilibrium, the conditions for thermal

$$T^1 = T^2 = \dots = T^j = \dots T^J, \quad (4.4)$$

mechanical

$$p^1 = p^2 = \dots = p^j = \dots p^J, \quad (4.5)$$

and chemical equilibrium

$$\mu_c^1 = \mu_c^2 = \dots = \mu_c^j = \dots \mu_c^J. \quad (4.6)$$

must be simultaneously fulfilled.<sup>341</sup> Here,  $T$  denotes the temperature,  $p$  describes the pressure, and  $\mu_c$  is the chemical potential of chemical species  $c = 1, 2, \dots, C$ . The chemical potential is derived from the Gibbs free energy  $G$  (see Appendix A for a detailed derivation). Mixtures encountered in biorefinery systems frequently involve one or multiple solvents and multiple biomolecules. Such complex mixtures typically show non-ideal thermodynamic behaviour. For non-ideal mixtures, the chemical potential  $\mu_c^j$  is defined as

$$\mu_c^j(p, T, \mathbf{x}) = \mu_c^{j,0}(p, T, \mathbf{x}) + RT \ln a_c^j. \quad (4.7)$$

Here,  $\mu_c^{j,0}$  is the chemical potential of a reference state and  $R$  refers to the universal gas constant. The activity  $a_c^j$  is defined as

$$a_c^j = \gamma_c^j x_c^j, \quad (4.8)$$

where the activity coefficient  $\gamma_c^j$  acts as a scaling factor to account for deviations from ideal thermodynamic behaviour.<sup>342</sup> Hence, the activity  $a_c^j$  can be interpreted as „effective“ concentration of a component in non-ideal mixtures. In biorefinery processes, often the molar composition  $x_c^j$  in phase equilibrium is sought after, e.g. the solubility of a solid solute in a solvent (SLE), or the mutual solubility of two solvents (LLE). To solve the system of equations arising from Eqs. 4.6 and 4.7 for  $x_c^j$ , the activity coefficients  $\gamma_c^j$  must be known. Activity coefficients can be obtained from experiments, which is however inconvenient for high-throughput CAMSD. To enable high-throughput solvent selection, predictive thermodynamic models are required to estimate activity coefficients.

Models of the excess Gibbs energy, such as the Wilson model,<sup>343</sup> the Non-Random-Two-Liquid (NRTL) model,<sup>344</sup> universal functional activity coefficient (UNIFAC),<sup>345</sup> and Universal

Quasichemical (UNIQUAC)<sup>346</sup> are established models for activity coefficients predictions on the basis of the excess Gibbs energy. Furthermore, results obtained from equation of state (EoS) models, such as Soave-Redlich-Kwong,<sup>347</sup> Peng-Robinson,<sup>348</sup> perturbed-chain statistical associating fluid theory (PC-SAFT),<sup>349,350</sup> can be used to determine activity coefficients. Although all of these models have been successfully applied to the prediction of phase equilibria in chemical engineering, their applicability to biorefinery processes is rather limited for several reasons. First, these models require component-specific parameters. These parameters were determined for commonly used chemicals for which they are available in large databases.<sup>351</sup> However, component-specific parameters are generally not widely available for complex biomolecules<sup>352</sup> or novel solvent classes, such as ILs.<sup>353</sup>

For biological systems, the Hansen solubility model<sup>48,354</sup> provides a simple and helpful approach to roughly estimate whether a solvent will dissolve a solute. The Hansen model is derived from EoS and assumes that the solubility of a component in a solvent is determined by intermolecular dispersion  $\delta_d$ , dipolar interactions  $\delta_p$ , and hydrogen-bonding interactions  $\delta_h$ .<sup>48,354</sup> These parameters span a three-dimensional space, the so-called Hansen space. The smaller the distance of the solvent and the solute in the Hansen space, the more likely the solute will dissolve. However, the model cannot sufficiently capture the complex dissolution process for many polymers.<sup>355</sup> For lignin, inaccuracies were reported even for standard solvent systems.<sup>356,357</sup> A major drawback of using the Hansen solubility model in high-throughput CAMSD is the unavailability of the model parameters.<sup>355</sup> While these parameters can be found in the literature for standard solvents, they must be experimentally or computationally derived for less frequently applied solvents or biopolymers.

#### 4.3.1.1 Quantum chemical models

For systems including complex biopolymers, a viable alternative to the aforementioned models are quantum chemistry (QC)-based methods. Due to their high accuracy, QC-based methods are considered the current gold standard for complex molecules, and have been frequently applied for biological systems.<sup>52–55,306,310</sup> Another advantage of QC-based methods is their independence from compound-specific model parameters. In QC-based methods, the molecules are described at a sub-atomic level as an assembly of nuclei and electrons. The basis for electronic structure calculations is given by the electronic Schrödinger equation

$$\hat{H}_{\text{el}}(\mathbf{R}, \mathbf{r})|\Psi_i(\mathbf{R}, \mathbf{r})\rangle = E_i|\Psi_i(\mathbf{R}, \mathbf{r})\rangle, \quad (4.9)$$

where  $\hat{H}_{\text{el}}$  and  $E$  denote the electronic Hamilton operator and the energy, respectively.<sup>358</sup> The  $N$ -electron wavefunction  $\Psi_i(\mathbf{R}, \mathbf{r})$  of electronic state  $i$  depends on the electronic coordinates  $\mathbf{r}$  and nuclear coordinates  $\mathbf{R}$ . From the wavefunction, which is the eigenvector of this



eigenvalue equation, any property of the system can be deduced. Provided that the nature of nuclei and numbers of electrons in the system are known, the equation can be formulated for molecular systems of any chemical composition. For solving the Schrödinger equation, only fundamental physical constants (e.g. the Planck constant) and no empirical parameters are required. However, solving the Schrödinger analytically, is only possible for very simple systems, e.g.  $H_2$ . Larger systems require approximations to reduce their complexity. Hohenberg and Kohn derived the density functional theory (DFT)<sup>359</sup> where the Schrödinger equation is not directly solved. Instead, one-particle electron densities are determined, denoted as  $\eta(\mathbf{r})$ . These densities describe the probability distribution of electrons in space. A crucial aspect is the one-to-one relationship between the electron density and an external potential, represented by  $v(\mathbf{r})$  which acts on the electron density. The electron density uniquely determines the external potential and, consequently, the ground state wavefunction. As a result, the system's energy is expressed using a universal functional of the electron density, denoted as  $F[\eta(\mathbf{r})]$ . Importantly, this functional depends solely on the electron density and is independent of the external potential. In this way, the energy of a systems' ground state is computed as

$$E[\eta(\mathbf{r})] = \int v(\mathbf{r})\eta(\mathbf{r}) + F[\eta(\mathbf{r})]. \quad (4.10)$$

Compared to wavefunction-based QC methods which give results within chemical accuracy ( $1 \text{ kcal mol}^{-1}$ ), DFT methods are slightly less accurate ( $2\text{-}10 \text{ kcal mol}^{-1}$ ) but up to 100 times faster.<sup>360</sup> Therefore, DFT methods became the workhorse of computational quantum chemistry.

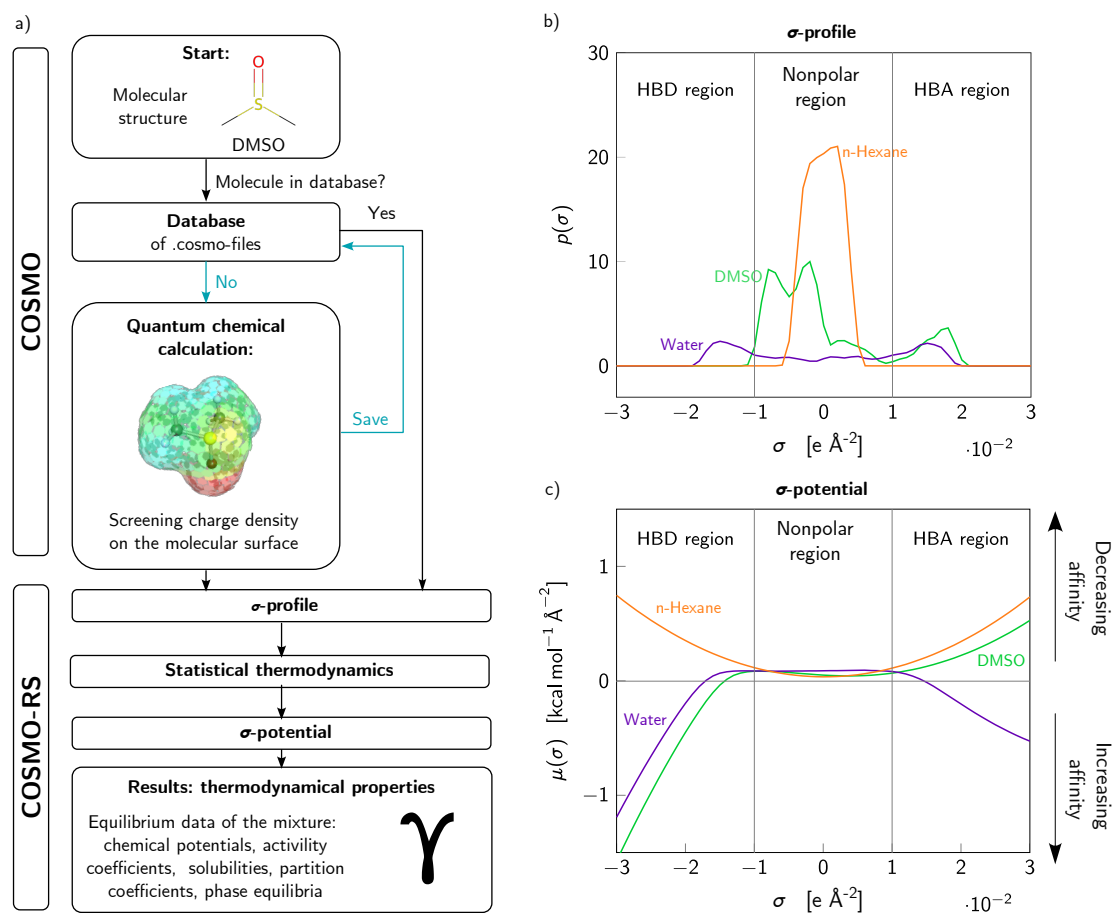
Classical quantum chemistry considers isolated systems, neglecting the influence of the surrounding environment. However, many chemical processes of interest, particularly in biology, occur in solution. Continuum solvation models (CSMs) serve as a bridge between classical quantum chemistry and the condensed phase, and incorporate solvent effects into QC calculations.<sup>361</sup> Instead of describing the solvent as individual molecules, CSMs represent the solvent as a continuous dielectric medium. The solute is treated on the quantum mechanical (QM) level and is embedded in the dielectric continuum. A molecular surface, also known as cavity, is constructed around the solute.<sup>361</sup> The solute's charge distribution polarises the solvent, and vice-versa, the dielectric continuum polarises the solute's charge distribution. This solute-solvent interaction gives rise to an iterative mathematical problem that can be easily integrated into QC methods.<sup>361</sup> CSMs are particularly valuable for studying biomolecular systems, where the influence of the surrounding solvent is critical, e.g. for proteins and other biomacromolecules.

**4.3.1.1.1 Conductor-like screening model and conductor-like screening model for real solvents** CSMs enable QM-based property predictions that are useful in CAMSD e.g.

for solvent selection,<sup>50,362–364</sup> polymers,<sup>365</sup> and complex biomolecules.<sup>52–55,304–312</sup> COnductor like Screening MOdel (COSMO) is a CSM that considers solute molecules placed in a virtual, perfect conductor ( $\epsilon = \infty$ ).<sup>366</sup> COSMO calculations are performed for all meaningful conformers of the molecules using standard QC software, e.g. TURBOMOLE.<sup>367</sup> The COSMO framework discretises the molecular surface into segments and provides each segments' screening charge density  $\sigma$  (Fig. 4.2 a).<sup>368</sup> The  $\sigma$ -profile is inferred from all discrete segments, representing the likelihood that a specific surface element of a molecule possesses a particular surface charge density  $\sigma$ . The COSMO calculation is based on DFT and requires computational times ranging from several minutes up to a week, depending on the size of the molecular size and complexity. COnductor like Screening MOdel for Real Solvents (COSMO-RS), converts the  $\sigma$ -profiles obtained from the COSMO calculation into  $\sigma$ -potentials by means of statistical thermodynamics (Fig. 4.2 b and c).<sup>369</sup> The  $\sigma$ -potentials describe the chemical potential of a surface segment with screening charge density  $\sigma$  and, thus, the affinity of the system to a surface of a certain polarity. The  $\sigma$ -potentials are the basis to determine activity coefficients necessary for the prediction of solubilities and phase equilibria.<sup>368</sup> COSMO-RS calculations are performed within seconds and are therefore, remarkably faster than the COSMO calculations. COSMO-RS is implemented in the software COSMOthermX19.<sup>368–370</sup> For this software, additional databases containing finished COSMO calculations for around 7000 molecules are available. These databases can be expanded with additional molecules by the user. In case that the molecule of interest is already stored in the database, the COSMO step is not necessary, saving computational time. COSMO-RS was recently applied for the selection of ILs for the dissolution of complex biomolecules such as fatty acids,<sup>53–55</sup> lignin,<sup>52</sup> cellulose,<sup>306,308,309</sup> and hemicellulose.<sup>308,309</sup>

#### 4.3.1.2 Machine learning-based models

ML models are extensively applied for the prediction of thermodynamic properties and were recently applied for solubility predictions,<sup>371,372</sup> the prediction of activity coefficients,<sup>373–376</sup> and phase equilibria.<sup>377</sup> Based on large-scale datasets, these models learn a mapping of the molecular representation (see Section 4.4) to the target property.<sup>378</sup> During the training phase, these algorithms learn by adjusting their parameters through iterative performance evaluations, thereby increasing their accuracy. Furthermore, the accuracy of the predictions heavily relies on the quality of the training set. Commonly applied ML models are black-box models which are difficult to interpret. To explain the predictions, additional algorithms are required.<sup>379,380</sup> Even though correlations between the molecular descriptors and the target properties do not imply causality, these methods are useful guides to further investigate interesting relationships.



**Fig. 4.2** Workflow for thermodynamic property predictions using the COSMO-RS. a) COSMO-RS predictions rely on QC-based COSMO calculations that can be stored in a database. b)  $\sigma$ -profiles and c)  $\sigma$ -potentials of water, DMSO, and n-hexane. Adapted from Eckert and Klamt.<sup>368</sup>

### 4.3.2 EHS property models

In the past, the design of chemical processes was mainly focused on selecting the best technological alternative to meet a particular product specification, to minimise operational costs, and to maximise the profit.<sup>381</sup> Environmental impacts, and the safety and the health of the workers played a rather subordinate role and influenced design decisions at a later design stage at best. For sustainable process design, various EHS properties (e.g. mutagenicity, carcinogenicity, biodegradability, toxicity), should be considered in an early design phase. The importance of EHS properties is also reflected on a political level. In 2006, the EU passed the Registration, Evaluation, and Authorization of Chemicals (REACH) legislation that regulates the handling of chemicals with respect to their EHS properties to protect the human health and the environment against harmful substances.<sup>382</sup> Since the fundamental concept

of biorefinery processes revolves around enhancing the overall sustainability of biomass processing, it is imperative to consider the EHS properties of the applied chemicals. For the predictions of EHS properties, different approaches can be applied, such as group contribution methods,<sup>383</sup> machine learning-based methods,<sup>384</sup> or quantitative structure activity relationship models.<sup>385</sup>

## 4.4 Molecular representation

The most accurate molecular representation is obtained from solving the Schrödinger equation (see Eq. 4.9) using QM-methods for the electronic Hamiltonian which relates to the Cartesian coordinates of the atomic positions in three-dimensional space. QM methods are computationally intensive and such detailed information of the molecular structure is not always required. Simpler molecular representations capture the inherent features and relationships of atoms and bonds while saving computational time. Text-based molecular representations, such as the SMILES,<sup>386,387</sup> SMILES arbitrary target specification (SMARTS),<sup>388,389</sup> or Self-Referencing Embedded Strings (SELFIES).<sup>390,391</sup> SMILES and SMARTS provide human-readable molecular embedding to describe the arrangement of atoms and bonds in a molecule while respecting a certain chemical grammar. The SMARTS language was specifically developed for substructure searches within molecules. In CAMSD, ensuring the chemical validity of generated molecules is a critical for efficient exploration of the chemical space. However, SMILES and SMARTS are only machine-readable when a distinct chemical grammar is followed. Due to this limitation, SELFIES were developed as a text-based representation that always provides structural feasibility.

As an alternative to text-based representations, molecules can be described by a multitude of chemical descriptors, of which the molecular fingerprint is the most prominent example. Molecular fingerprints are numerical, vectorial representations of molecular structures that encode information about structural features, such as the occurrence of particular functional groups.

In graph-based representations, atoms are considered as nodes that are connected by edges, the chemical bonds. Thus, molecules are treated as an undirected graph. Representing molecules as graphs follows the molecular structure very naturally. Text-based representations, such as SMILES, SMARTS or SELFIES, can be seen as one-dimensional text encodings of the molecular graph. More advanced graph representations, such as weighted graphs, include a variety of atom and bond features such as bonding type, aromaticity, charge, and distance.<sup>373,374,392,393</sup> Furthermore, molecules can be represented by discrete building blocks, e.g. functional groups.

The molecular representation of choice should capture the most important molecular features that are related to the property of interest, should be compatible with the property

model, and offer the desired level of interpretability.



## 5 | Development of computer-aided solvent screening and design methods for biorefinery processes

In this chapter, solvent screening and solvent design methods are developed to fill the gap in computational methods for biomass fractionation in biorefinery processes. In Section 5.1, the solvent selection problem is formulated as a mathematical optimisation problem. Based on this general formulation, a screening method (Section 5.2) and a solvent design method (Section 5.3) applicable to solvent selection problems in biorefineries are derived.

*This chapter is based on the publications of König-Mattern et al.<sup>394–396</sup> and includes direct as well as modified excerpts used under Creative Commons license CC BY 4.0.*

### 5.1 Solvent selection in biorefineries as a CAMSD optimisation problem

The general CAMSD problem described in Eqs. 4.1-4.3 is the foundation to formulate the optimisation problem tailored towards biorefinery processes. For a general biorefinery process with thermodynamic properties  $\mathbf{t}$  (e.g. solubilities of the biomolecules  $\mathbf{b}$  in the solvent structures  $\mathbf{s}$ , partition coefficients, molar compositions in phase equilibria), EHS properties  $\mathbf{e}$  (e.g. acute toxicity, mutagenicity, biodegradability), and process conditions  $\mathbf{p}$  (e.g. temperature, pressure), the CAMSD problem can be formulated as

$$\min_{\mathbf{s} \in \mathcal{S}} \quad F(\mathbf{t}(\mathbf{b}, \mathbf{s}), \mathbf{e}(\mathbf{s}), \mathbf{s}) \quad \text{objective function} \quad (5.1)$$

$$\text{s.t.} \quad \mathbf{g}(\mathbf{s}) \leq \mathbf{0} \quad \text{structural constraints,} \quad (5.2)$$

$$\mathbf{h}(\mathbf{t}(\mathbf{b}, \mathbf{s}), \mathbf{e}(\mathbf{s}), \mathbf{p}(\mathbf{b})) \leq \mathbf{0} \quad \text{property constraints.} \quad (5.3)$$

Here, the objective function  $F$  takes into account the desired solvent properties, such as thermodynamic properties  $\mathbf{t} \in \mathcal{T}$ , the EHS properties  $\mathbf{e} \in \mathcal{E}$ , and/or structural features of the solvents  $\mathbf{s} \in \mathcal{S}$ . Depending on the solvent selection problem, either all of the mentioned properties, or only selected properties can be included in the objective function  $F$ . To prevent unnecessary evaluations of  $F$ , solvents with unsuitable structures should be excluded based on the structural constraints  $\mathbf{g}$  and property constraints  $\mathbf{h}$ . In this manner, solvents with unsuitable structures (e.g. functional groups that are prone to undesired reactions with parts of the biomass) and undesired properties (e.g. low solubilities or harmful EHS properties) are excluded. Furthermore, the  $T_m$  and  $T_b$  limits of the solvents might serve as exclusion criterium in  $\mathbf{h}$ . Excluding solvents with unsuitable melting and boiling point ensures the solvents being liquid at the process conditions  $\mathbf{p} \in \mathcal{P}$ . The process conditions are mainly determined by the structure of the biomolecules  $\mathbf{b} \in \mathcal{B}$  (e.g. heat-sensitive pigments require mild temperatures close to room temperature, while for the processing of lignocellulose extreme high temperatures are applied). In this manner, the search space  $\mathcal{S}$  is narrowed to a significantly smaller  $\tilde{\mathcal{S}}$ . The set  $\tilde{\mathcal{S}}$  contains only molecules meeting the constraints  $\mathbf{g}$  and  $\mathbf{h}$ . Properties that are required to evaluate  $F$  or  $\mathbf{h}$  must be predicted by suitable models, which must be applicable for complex biomolecules and a broad range of different solvents, such as QC- or ML-based methods. To solve the solvent selection problem for biorefineries as given in Eqs. 5.1-5.3, a screening method (Section 5.2) and a design method (Section 5.3) were developed.

## 5.2 Solvent screening method for biorefinery processes

The solvent screening method screens a defined database  $\mathcal{S}_{\text{database}}$  and excludes all molecules not satisfying structural constraints  $\mathbf{g}$  and property constraints  $\mathbf{h}$ . For all solvent candidates  $\mathbf{s}$  included in the pre-screened search space  $\tilde{\mathcal{S}}_{\text{database}}$ , the objective function  $F$  (Eq. 5.1) is evaluated. Thermodynamic and EHS property models are necessary to evaluate the objective function  $F$  and the constraints  $\mathbf{h}$ . EHS properties of the solvents are predicted by available QSAR models. The thermodynamic properties of the pure solvents, solvent mixtures, or biomolecule/solvent mixtures are predicted using COSMO-RS which is seen as the current gold-standard for complex molecules and mixtures. For COSMO-RS property predictions of biomolecule/solvent mixtures, e.g. solubility predictions or partition coefficients, QC-calculations of the biomolecules are required. Frequently, results of these calculations are not available in publicly open or commercially available databases. Therefore, biomolecules that are representative for a specific biomass fraction must be modeled on a quantum chemical level. In some cases, the size of these representative biomolecules would lead to infeasibly high computation times, e.g. for large polymers such as cellulose or lignin. In this case,



simplified model molecules must be determined. Finally, the solvent candidates are ranked according to their performance determined by  $F$ .

The developed solvent screening methodology was implemented as an automated solvent screening framework in Python by the following steps (Fig. 5.1):

1. Database construction
2. Identification and QC modeling of representative biomolecules for each biomass fraction
3. Screening: In this step, the number of solvent candidates is gradually reduced. Here, unsuitable solvents are determined and removed based on the constraints  $\mathbf{g}$  and  $\mathbf{h}$ . It is advisable to perform the property predictions with the lowest computational demand first, to reduce the number of solvents that is admissible in subsequent property predictions. After excluding solvent candidates not satisfying the constraints, the objective function  $F$  is evaluated for each remaining solvent.
4. Solvent ranking: The identified solvent candidates are ranked according to their performance given by  $F$ .

In the following sections, the database construction, the choice of the representative molecules, the prediction of thermodynamic and EHS properties, the solvent ranking, and computational methods of the solvent screening method are described.

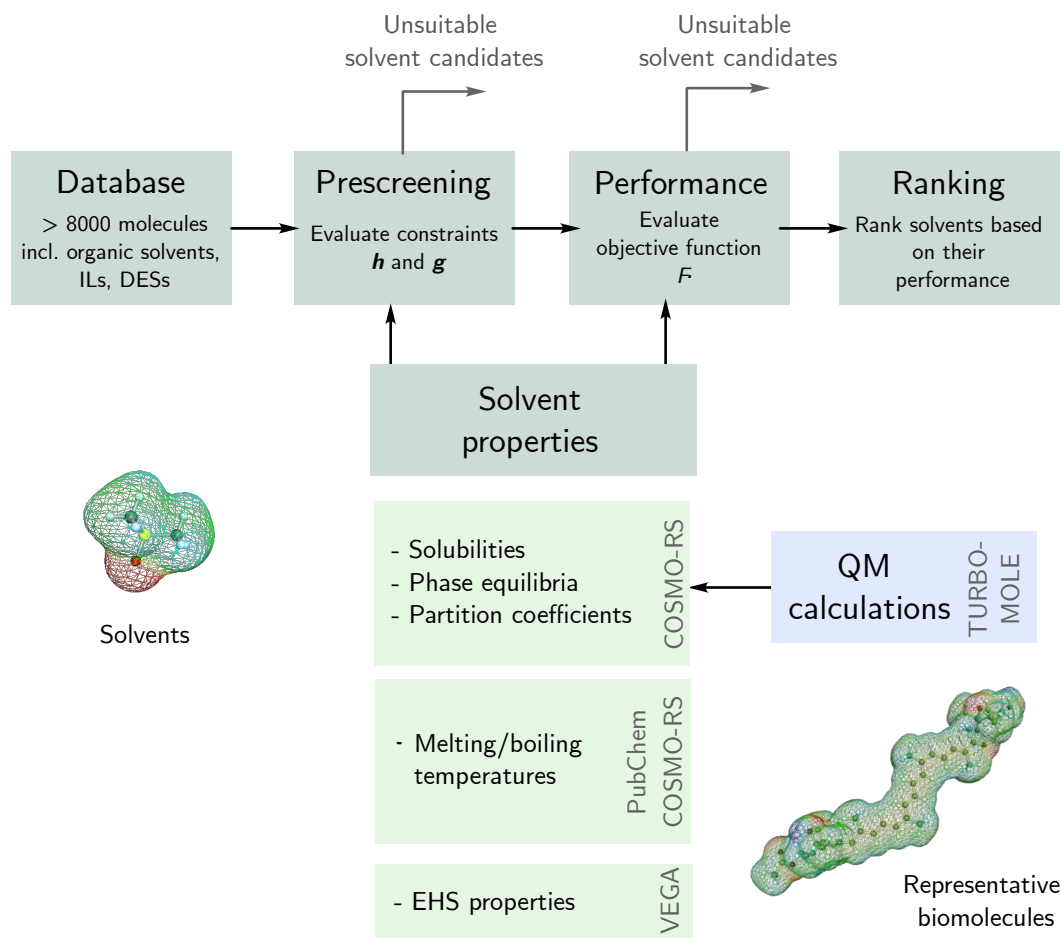


Fig. 5.1 Overview of the solvent screening procedure.

### 5.2.1 Database

The compiled database consists of the COSMOthermX19 integrated database, COSMObase13-01, and COSMObaseLL19-01. Additionally, green solvents from Moity et al.,<sup>397</sup> and the green solvent Cyrene<sup>398</sup> were modeled by Linke et al.<sup>50</sup> and added to the database. Duplicate solvents were deleted. Furthermore, 178 pairs of DESs<sup>399–409</sup> and 143 commercially available IL pairs<sup>410,411</sup> were modeled and added to the search space, leading to a total of 8011 molecules defining the solvent search space  $\mathcal{S}_{\text{database}}$ . A complete list of the database is provided in Appendix B.

### 5.2.2 Representative biomolecules

Each biomass fraction is a complex mixture of different biomacromolecules with various interactions, which are far beyond to be modeled as a whole using QC methods. Therefore,

several representative biomolecules from each fraction are chosen and modeled individually, assuming that the influence between each molecule is negligible. For common DFT methods, the computation time scales with the size of the molecule  $N_{\text{mol}}$  in the order of  $\mathcal{O}(N_{\text{mol}}^3)$ ,<sup>360</sup> leading to a high computational effort for large molecules. Therefore, for large biopolymers, representative parts of the original structure can be modelled to reduce the computational cost. The computation details of the modelling of the representative molecules is described in Section 5.2.5. The choice of representative molecules for each biomass fraction is guided by

- high importance for the overall process (e.g. high-value molecules),
- high abundance (e.g. on a weight-basis),
- high solvent accessibility (e.g. molecules stored in solvent-inaccessible locations are not representative due to a lack of solvent contact),
- computing effort (e.g. a suitable simplification of the molecules must be made for large molecules).

It is strongly recommended to verify the validity of representative molecules by comparing experimental and computational results for the property of interest for selected solvents.

### 5.2.3 Thermodynamic property predictions

COSMO-RS is the gold standard method for thermodynamic property predictions involving biomolecules (see Section 4.3.1), and was therefore applied in the solvent screening approach. In the following, the thermodynamic properties  $\mathbf{t}$  with high importance in biorefinery processes are described and their calculation using COSMO-RS is explained.

#### 5.2.3.1 Solubility

In COSMOtherm, the solubility of the representative molecules in the solvent candidates calculated as the logarithmic, molar solubility  $\log_{10}(x_b)$ . The molar solubility of the representative biomolecule  $b$  in a solvent  $s$  is defined as

$$x_b = \frac{n_b}{n_b + n_s}. \quad (5.4)$$

The solubility predictions in COSMOtherm are derived from the SLE. Here, COSMO-RS utilises the pseudochemical potentials  $\mu^*$  obtained from the SCDs of  $b$  and  $s$  to approximate the molar fraction of the representative molecules in the liquid phase (see Appendix A for a detailed derivation).

$$\log_{10}x_b^l = \frac{\mu_b^{*,0}(T, p, x_b^l = 1) - \mu_b^{*,\infty,l}(T, p, x_b^l) - \Delta G_{\text{fus}}(T)}{RT \ln(10)} \quad (5.5)$$

Here,  $\mu_{\text{rep}}^{*,\infty,l}$  describes the pseudochemical potential of the representative biomolecule at infinite dilution in solvent  $s$ . The obtained solubility can be considered as a zeroth-order approximation. However, the accuracy of the solubility predictions can be increased if the solubility  $x_b^l$  is used to determine  $\mu_b^{*,l}(T, \rho, x_b^l)$ , and inserted into Eq. 5.5 iteratively. The free enthalpy of fusion  $\Delta G_{\text{fus}}(T)$  is zero for liquid compounds and is estimated for solid compounds based on a quantitative structure-property relationship (QSPR) approach implemented in COSMOtherm (see Appendix A.1).

Since often several biomolecules are used to represent a single biomass fraction, the solubility for a the overall fraction was calculated as the average solubility of several representative molecules denoted as  $\log_{10}(x_{\text{frac}})$ . In this case, the solubilities of the respective single representative molecules were first predicted according to Eq. 5.5, and subsequently averaged:

$$\log_{10}(x_{\text{frac}}) = \frac{1}{n_b} \sum_{i=1}^{n_b} \log_{10}(x_b^{(i)}), \quad (5.6)$$

where  $n_b$  describes the number of representative molecules in the respective fraction.

### 5.2.3.2 LLEs

In biorefineries, liquid-liquid extraction is an important separation process. When two liquid phases  $j = 1, 2$  are in LLE, the molar fraction of the solvent  $s$  in each phase are calculated based on Eqs. 4.6-4.8 as

$$x_s^1 \gamma_s^1 = x_s^2 \gamma_s^2. \quad (5.7)$$

Here, the activity coefficients  $\gamma_s$  are obtained from the screening charge densities  $\sigma$  of the solvents calculated by COSMO-RS. Within COSMOthermX19, the molar fractions are determined iteratively.

### 5.2.3.3 Partition coefficients

In liquid-liquid extraction, not only the phase solvent composition of the liquid phases is of interest. Also, the partition coefficients of the biomolecules, describing the efficiency of the separation, are important. In COSMOtherm, the solute is treated as infinitely diluted in the phases  $j = 1, 2$ . The solvent composition at LLE is obtained from Eq. 5.7. The partition coefficient is calculated as<sup>412</sup>

$$\log_{10} P_b^{(2,1)} = \frac{\mu_b^{(*,\infty,1)} - \mu_b^{(*,\infty,2)}}{RT \ln(10)} = \log_{10} \left( \frac{x_b^{(2)}}{x_b^{(1)}} \right). \quad (5.8)$$

### 5.2.4 Prediction of EHS properties

The prediction of EHS properties is necessary to select benign solvents. Here, the EHS properties listed in Tab. 5.1 were predicted using the quantitative structure-activity relationship (QSAR) models implemented in the VEGA software.<sup>385</sup> In addition to the EHS property predictions, VEGA calculates a reliability score, enabling further assessment of the obtained predictions. Using the results and the reliability scores, the EHS score was predicted as proposed by Linke et al.<sup>50</sup> The EHS score describes how „benign“ a solvent is, providing a value in the interval  $[0, 1]$ . Here, an EHS score closer to one indicates more benign EHS properties. In brief, for each solvent candidate, the EHS properties summarised in Tab. 5.1 are predicted by the VEGA models. In addition, the flammability of a solvent was evaluated in terms of the flash temperature  $T_f$ , which is defined as the temperature at which vapour over a liquid combusts completely upon ignition. COSMOtherm estimates the  $T_f$  iteratively based on the vapour pressure  $p_c^0$  and the pressure  $p_c(T_f)$ .

The results for each predicted EHS property are represented by different data types, such as numeric values for fish toxicity, or qualitative responses, such as „mutagenic“ or „toxic“. These outputs were converted to a colour-code accompanied score. This score is element of the set  $\{0, 0.25, 0.5, 0.75, 1\}$ , where „green - 1“ corresponds to desirable („green“) EHS properties, „green-yellow - 0.75“, „yellow - 0.5“, „yellow-red - 0.25“, and „red - 0“. Some of the EHS properties can be predicted by multiple models (e.g. mutagenicity or carcinogenicity, see Tab. 5.1). In this case, the individual model results were weighted with the predicted reliability and unified as described by Linke et al.<sup>50</sup> The VEGA software additionally contains experimentally determined data points which were preferred over the values predicted by the QSAR models. The score of each unified model result was averaged by the number of models applied to generate the EHS score.

VEGA is not applicable to ILs since these salt-like molecules were not included in the training-set of the VEGA models. To prevent false exclusion, their EHS score was set to 1. In case that an IL is obtained as a suitable solvent candidate at the end of the screening, their EHS properties were manually checked based on the safety data sheet (SDS). For DESs, the EHS score for the hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) were predicted separately. To obtain the EHS score for the whole DES, the score of the HBA and HBD were averaged and weighted by their molar fractions in the DES. VEGA models were recently applied for a screening of DESs by Song et al.<sup>301</sup> and are a reasonable choice given the number of solvent candidates. However, recent studies regarding the toxicity of DESs suggest, that the toxicity of the DES might diverge from that of its single components.<sup>413</sup>

**Tab. 5.1** QSAR models from VEGA applied for the prediction of selected EHS properties.

Property	Model	Version
Mutagenicity	CONSENSUS	1.0.3
	CAESAR	2.1.13
	SarPy/IRFMN	1.0.7
	ISS	1.0.2
	KNN/reas-across	1.0.0
Carcinogenicity	CAESAR	2.1.9
	ISS	1.0.2
	IRFMN/Antares	1.0.0
	IRFMN/ISSCAN-CSX	1.0.0
Developmental toxicity	CAESAR	2.1.7
	PG	1.1.0
Endocrine disruptor potential	IRFMN	1.0.1
	IRFMN/CERAPP	1.0.0
Skin sensitization	CAESAR	2.1.6
Hepatotoxicity	IRFMN	1.0.0
Fish acute toxicity	SarPy/IRFMN	1.0.2
	KNN/read-across	1.0.0
	NIC	1.0.0
	EPA (96h)	1.0.7
Daphnia toxicity	EPA (48h)	1.0.7
	DEMETRA	1.0.4
Bee acute toxicity	KNN/IRFMN	1.0.0
Bioaccumulation factor	CAESAR	2.1.14
	Meylan	1.0.3
	KNN/read-across	1.1.0
Biodegradability	Arnot/Episuite	1.0.0
	IRFMN	1.0.9
Persistence (sediment)	IRFMN	1.0.0
Persistence (soil)	IRFMN	1.0.0
Persistence (water)	IRFMN	1.0.0
Log <sub>10</sub> P(octanol/water)	Meylan/Kowwin	1.1.4
	MlogP	1.0.0
	AlogP	1.0.0

#### 5.2.4.1 Determination of melting and boiling points

The  $T_m$  and  $T_b$  ranges considered in  $\mathbf{h}$  are usually defined based on the process conditions  $\mathbf{p}$ . For many substances,  $T_m$  and  $T_b$  (at atmospheric pressure) are already given in the COSMOtherm database. However, for solvents that were added to the database later on,  $T_m$  and  $T_b$  were lacking. In this case, an automated PubChem database query retrieved missing experimental data.<sup>414</sup> If there was no experimental data available in PubChem, the missing  $T_b$  was predicted using COSMO-RS. Solvents with missing entries were kept in the screening to prevent false exclusion.

#### 5.2.4.2 Structural constraints

The structural constraints defined by  $\mathbf{g}$  ensure that solvents containing unsuitable functional groups, e.g. due to reactivity with the biomass or the process liquor, are eliminated from the list of potential solvent candidates. Solvents containing the defined groups were identified using RDKit.<sup>415</sup>

#### 5.2.4.3 Solvent ranking

After predicting all properties that are required to evaluate the objective function  $F$  for the solvent candidates identified in the screening procedure, the identified solvent candidates are ranked to determine deriable candidates. Often, the optimal values  $\mathbf{o}$  of the properties of interest are known, e.g. ideal EHS score = 1 and solubility  $x_b = 1$ . Based on  $\mathbf{o}$  the solvents can be ranked according to the Euclidian distance  $d$  of the properties of interest  $\mathbf{i}$  from the optimal values  $\mathbf{o}$

$$d(\mathbf{o}, \mathbf{i}) = \sqrt{\sum_k^K (o_k - i_k)^2}. \quad (5.9)$$

Note that in this case the values for all considered properties must be normalised. For all highly ranked solvents, the predicted thermodynamic properties and EHS properties were manually checked based on literature data, publicly available databases, and SDSs.

#### 5.2.5 Computational details

For molecules not contained in the COSMOtherm databases (COSMOtherm integrated database, COSMObase13-01, COSMObaseLL19-01), QC calculations were performed. Molecular conformers were generated using RDKit<sup>415</sup> with a force field according to Eberjer et al.<sup>416</sup> Subsequently, the resulting structures were further optimised at a QM level using TURBOMOLE 3.7 and its *calculate* interface (version 2.1, 2009). The BP-86 functional with the def-TZVP bases set was applied using the COSMO boundary condition and the

standard COSMO cavity construction. With the optimised geometries, a single point calculation was performed, using the more accurate def2-TZVPD bases set. Cavities were constructed at the FINE level. The automated screening procedure was implemented in a python script (Python 3.7). In this procedure, the database described in Section 5.2.1, was loaded as a pandas dataframe (pandas 0.25.1). Subsequently, a pre-screening step was performed as given by the constraints  $\mathbf{g}$  and  $\mathbf{h}$ , and unsuitable molecules were deleted. PubChemPy 1.0.4 was used to read the  $T_m$  and  $T_b$  of the compounds from PubChem. Missing  $T_b$  and  $T_f$  (used for predicting EHS properties) were predicted using CosmoPy 19.10. EHS properties were predicted using VEGA QSAR 1.1.5. For solubility predictions, the python script called COSMOthermX19<sup>368–370,417,418</sup> via the command line. For these calculations, input files were automatically generated and the BP\_TZVPD.FINE\_19.ctd parameterisation was applied. The results of each solubility prediction was stored in SQLite3 databases (sqlite 3.30.0). All COSMO-RS predictions including ILs and DESs were handled using the so-called electroneutral approach.<sup>419–421</sup> In this approach, the single IL and DES constituents were implemented as a stoichiometric mixture. The electroneutral approach is commonly applied for predictions concerning ILs and DESs. A Linux Ubuntu 16.04. system (Intel i5-8500 processor at 3.00 GHz and 16 GB RAM) was used to perform all calculations.

## 5.2.6 Advantages and limitations

The high-throughput screening framework is a tool specifically developed to facilitate solvent selection for biorefinery processes. The tool enables systematic computational solvent screening of a database containing more than 8000 potential solvents. During the screening, unsuitable solvent candidates are eliminated as given by structural constraints  $\mathbf{g}$  (e.g. unsuitable functional groups) or property constraints  $\mathbf{h}$  (e.g. solubilities, EHS properties,  $T_m$  and  $T_b$  limits). For all solvents meeting the constraints, the objective function  $F$  that measures the performance is evaluated. Based on the performance the solvents are ranked, enabling to identify the optimal solvent candidate of the database.

The prediction of solubilities, LLEs, and partition coefficients is performed using COSMO-RS, the gold-standard method for complex biomolecules. For each biomass fraction, representative molecules are defined and modelled on a quantum chemical level to obtain SCDs. COSMO-RS converts the SCDs into  $\sigma$ -profiles and  $\sigma$ -potentials, providing additional chemical information about hydrogen bonding behaviour and polarity which are important features for the solvent selection. However, several simplifications are made. The biomass is far too complex to be modelled as a whole on a quantum chemical level. Therefore, the representative molecules for each biomass fraction are modelled individually. For large biopolymers, simplified structures are used to reduce the computational time. The quantumchemical calculations are performed at the reference state of an ideal conductor,



providing enormously efficient property predictions. However, the response of the solute electron density to the solvent cannot be captured. During the prediction of solubilities and partition coefficients, interactions between the biopolymers are neglected. These predictions are based on the chemical potential at infinite dilution  $\mu_b^\infty$ , implying that the representative biomolecule  $b$  is assumed to be only surrounded by solvent molecules. Furthermore, in COSMO-RS, polymers are treated as „pseudo-liquids“. As a consequence, the applicability of COSMO-RS for crystalline polymers is limited. Also, polymer swelling is not taken into account.

Although the screened database is larger than the search space in previously published solvent screening approaches,<sup>52–55,307</sup> it represents only a small subset of the chemical space with an estimated size of around  $10^{18}$  -  $10^{200}$  molecules. Considering the enormous size of the chemical space, the globally optimal solvent candidate is likely not included in the solvent database. However, the solvent database represents a large collection of chemically diverse molecules of which many are commercially available. Therefore, the screening framework has the potential to identify meaningful solvent candidates in a short time.

### 5.3 Solvent design method: The genetic algorithm *PSEvolve*

To explore the chemical space beyond an existing database and to expand the search for so-far undiscovered solvents, a GA for molecular design was developed. The GA, called *PSEvolve*, was implemented as a Python framework which is publicly available under: <https://github.com/koenigmattern/PSEvolve>. In *PSEvolve*, a molecule is represented as a graph  $G(V, E)$ , where the vertices  $V$ , also called nodes, represent atoms and the edges  $E$  represent bonds. The graph representation is preferred since commonly applied text-based representations, such as SMILES or SMARTS, can lead to structural infeasibility of the designed molecules (see Section 4.4).

In the following, each step of the *PSEvolve* algorithm is explained (Fig. 5.2). First, the hyperparameters are set, including population size  $N_{\text{pop}}$ , mutation rate  $m$ , number of parent molecules  $N_{\text{par}}$ , number of generated children  $N_{\text{fili}}$ , and the ending criterion (e.g. maximum number of generations  $N_{\text{gen}}$ ). Furthermore, the objective function  $F$ , the constraints  $\mathbf{h}$  and  $\mathbf{g}$  (Eq. 5.1-5.3) must be defined. Subsequently, a start population  $P_0$  is generated. The start population can be freely chosen (e.g. randomly chosen molecules). Afterwards, the iterative, stochastic optimisation procedure starts with the property predictions necessary to evaluate  $\mathbf{h}$ ,  $\mathbf{g}$ , and  $F$ . Here, models with a suitable trade-off between speed and accuracy are necessary for an efficient, high-throughput exploration of the chemical space. In the screening

method described in Section 5.2, COSMO-RS was used for this purpose. However, COSMO-RS requires computationally intense QM-calculations for each newly designed solvent that is not already contained in the database that is described in Section 5.2.1. Hence, for these predictions, suitable surrogate models must be developed. Based on the property predictions, the objective function  $F$  is evaluated and the performance of each molecule is measured. In the context of GAs, the performance is also called fitness. In resemblance to Darwin's theory of natural selection, the fitness determines the reproductive success. Based on the determined fitness,  $N_{\text{par}}$  parent molecules are selected. Here, roulette-wheel selection is applied, where the probability for selecting an individual is proportional to its fitness. During roulette wheel selection, the fitness values  $f_c = F(t_s, e_s, s)$  for each compound  $s$  in the population must be normalised (Eq. 5.10). Subsequently, the selection probability  $p_{\text{sel}}$  is determined

$$\tilde{f}_c = \frac{f - \min(x)}{\max(x) - \min(x)} \quad (5.10)$$

$$p_{\text{sel}} = \frac{\tilde{f}_c}{\sum_{c=0}^C \tilde{f}_c} \quad (5.11)$$

After selecting molecules for reproduction, the cross-over step follows. Here, the selected molecules are split into fragments as described in Section 5.3.1. The obtained fragments are stored in the fragment pool, from which  $N_{\text{fil}}$  child molecules are generated. To generate a child molecule, two parent fragments are randomly chosen from the mating pool. The chosen fragments are randomly combined to form a child molecule as described in Section 5.3.1. If the generated child molecule does not satisfy the constraints  $\mathbf{h}$  and  $\mathbf{g}$ , it is not added to the population. The cross-over procedure continues until  $N_{\text{fil}}$  child molecules are generated that meet the constraints  $\mathbf{h}$  and  $\mathbf{g}$ .

To reduce the risk of premature convergence of the GA, the mutation operation maintains diversity within the population. The probability that a molecule in the population mutates is given by the mutation rate  $m$ . *PSEvolve* provides eight mutation operations: atom and bond addition, atom and bond deletion, atom and bond substitution, relocation of functional groups, and addition of functional groups (see Section 5.3.1 for details). In case that a molecule is selected for mutation, the mutation operation to be performed is randomly chosen. When the respective mutation operation is performed, the validity of the mutated molecules with respect to the constraints  $\mathbf{g}$  and  $\mathbf{h}$  is checked. In the case that a particular mutation operation cannot be performed on the molecule, e.g. because no structurally feasible a valid structure arises, another mutation operation is randomly chosen.

After the mutation step, the fitness of each molecule in the population is evaluated and

individuals with the lowest fitness are deleted to maintain a constant population size  $N_{\text{pop}}$ . The  $N_{\text{pop}}$  fittest molecules comprise the new population  $P_1$ . Subsequently, the iterative loop of roulette-wheel selection, cross-over, and mutation and fitness evaluation continues until the ending criterium is reached (e.g. maximum number of iterations, or a constant average fitness of the population). The computational details of the solvent design method are described in Section 5.3.2.

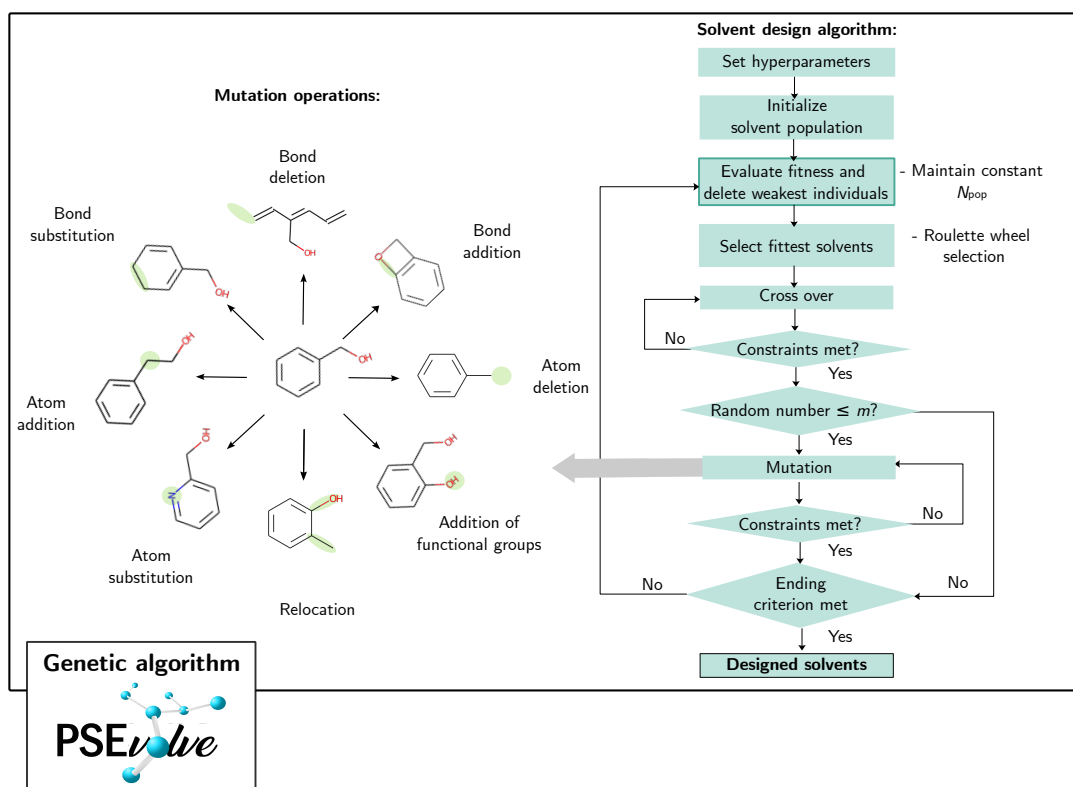


Fig. 5.2 Overview of the PSEvolve algorithm.

### 5.3.1 Structure-altering operations

In the following, all structure altering operations are described. Note, that in *PSEvolve*, H-atoms are treated implicitly.

#### 5.3.1.1 Fragmentation algorithm

In the fragmentation algorithm, a molecule is split into two or more fragments. Since splitting a molecule at a randomly chosen position does not always lead to the desired fragmentation (e.g. due to the occurrence of rings), first, feasible locations for the molecule fragmentation are identified. For this purpose, graph theory is applied. In graph theory, bridges are edges that split the graph into two or more fragments if deleted. Using the

python package networkx<sup>422</sup> and the implemented chain decomposition algorithm,<sup>423</sup> the set of all bridges is identified. From this set, one bridge is randomly chosen and deleted. The concept of bridges is illustrated for two molecules, hexane and benzene (Fig 5.3 a and b). All edges in the hexane graph are considered bridges. Here, any bond could be deleted to split the graph. The benzene graph, on the other hand, is circular, and therefore does not contain any bridges.

To enable the fragmentation of ring-containing molecules, a ring splitting function was implemented. If the molecular graph does not contain any bridges, the graph is checked for the occurrence of rings. Of the determined rings, one ring was chosen randomly to be fragmented. For ring splitting, two conditions must be fulfilled i) at least two edges must be deletable to split the graph into fragments, ii) the chosen bonds must not be neighbouring bonds, iii) the selected bonds must not be intersecting with a neighbouring ring. Of the edges fulfilling these criteria, two were chosen randomly and deleted to split the graph.

### 5.3.1.2 Combination algorithm

The combination algorithm combines two or more molecular fragments to obtain one single molecule. First, the implicit valences of all atoms in each fragment are determined. The implicit valence of an atom describes the number of implicit H-atoms attached to it. Since H-atoms are treated implicitly, the implicit valence describes the number of bonds that can be additionally formed. Subsequently, from each fragment, one atom with an implicit valence  $\geq 1$  is randomly chosen. For combination, a new bond is formed which can either be a single or double bond (triple bonds are neglected), depending on the valence. If for both bonding atoms the implicit valence is  $\geq 2$ , it is randomly decided whether a single or double bond is formed. Otherwise, a single bond is formed. If in one fragment no atoms with the ability to form additional bonds exist, the combination algorithm fails.

### 5.3.1.3 Cross-over

In the cross-over operation, the graph of the selected parent molecules is split using the fragmentation algorithm. The resulting fragments are stored in a fragment pool, from which  $N_{\text{offspring}}$  offspring molecules are generated. To form an offspring molecule, two fragments from the fragment pool are randomly selected and combined using the combination algorithm. The generated offspring molecules must fulfill the constraints  $\mathbf{g}(\mathbf{s})$  and  $\mathbf{h}$ . If some of the offsprings do not meet the constraints, new offspring molecules are generated until  $N_{\text{offspring}}$  offspring molecules fulfill the constraints.

#### 5.3.1.4 Bond addition

For bond addition, the set of atoms with the ability to form at least one additional bond (implicit valence  $\geq 1$ ) are determined. From all possible combinations of two atoms within this set, one combination is randomly chosen and a bond is formed. If both bonding atoms have an implicit valence  $\geq 2$ , then it is randomly decided whether a single or a double bond is formed. In case that no additional bond can be formed, the bond addition operation fails and a different mutation operation is randomly chosen.

#### 5.3.1.5 Bond deletion

The deletion of bonds can only be performed, if the deletion does not lead to a fragmentation of the graph. Therefore, all bridges are determined using `networkx`. From all non-bridge edges, one edge is randomly chosen and deleted. If all bonds are bridges, the bond deletion operation fails and another mutation operation is chosen.

#### 5.3.1.6 Bond substitution

During bond substitution, the bond type is changed, e.g. a single bond becomes a double bond, or vice-versa. A double bond can always be changed to a single bond. However, a single bond can only be changed to a double bond, if both connected atoms have an implicit valence  $\geq 2$ . In case no bond suitable for substitution can be determined, another mutation operation is randomly selected.

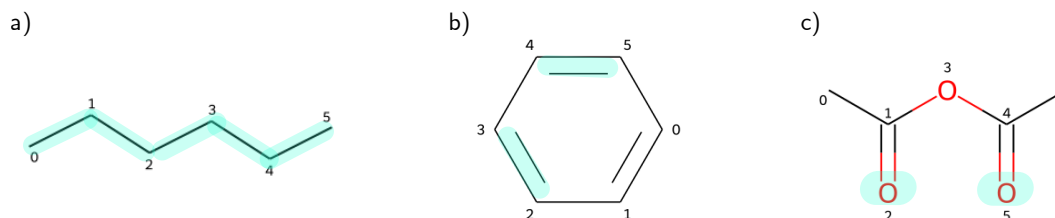
#### 5.3.1.7 Atom addition

For the addition of atoms, the considered molecule is fragmented using the fragmentation algorithm. The atom to be added is randomly chosen and treated as an additional fragment. Then, the combination algorithm randomly combines the all fragments. If the atom addition operation fails, and another mutation operation is randomly chosen.

#### 5.3.1.8 Atom deletion

During atom deletion, single vertices are deleted. However, not every vertex can be safely deleted due to the risk of fragmenting the molecular graph. To prevent the graph from splitting, the vertex connectivity algorithm<sup>424</sup> as implemented in the `networkx` package<sup>422</sup> was applied. The algorithm identifies the set of vertices that, if deleted, split the graph into two or more fragments, also called vertex cut. Therefore, all vertices that are not element of vertex cut can be safely deleted. The importance of identifying atoms suitable for deletion is exemplified for acetic anhydride (Fig. 5.3 c). Here, only the oxygen atoms of the carbonyl

groups can be deleted, since the deletion of all other atoms would lead to a fragmentation of the molecule.



**Fig. 5.3** Deleting edges and nodes from the molecular graph. a) In the hexane graph, all edges are bridges. Therefore, the fragmentation algorithm could delete each bond to achieve the desired fragmentation of the graph (blue highlights). In contrast, in the bond deletion operation, fragmentation is not desired. Therefore, this operation is not applicable to the hexane graph. b) In the benzene graph, there are no bridges. Therefore, a pair of non-neighbouring edges must be selected to achieve a fragmentation, as exemplified by the blue highlights. In contrast, in the bond deletion operation, each bond could be chosen, since no fragmentation would occur. c) For atom deletion, a fragmentation of the graph is not desired. Therefore, only the oxygen-atoms attached to the branches (blue highlights) can be deleted.

### 5.3.1.9 Atom substitution

During atom substitution, one atom is randomly chosen and its atom type is changed (e.g. a carbon atom is replaced by an oxygen atom). To ensure structural feasibility, the replacing atom must have an equal or higher explicit valence than the replaced atom. The explicit valence refers to the actual number of bonds that an atom forms in a specific molecule. If the substitution operation cannot be performed because no suitable replacing atom can be identified, the procedure starts over again. In case the substitution operation is not successful for ten tries, the operation fails to prevent the algorithm from getting stuck, and different mutation operation is randomly chosen.

### 5.3.1.10 Relocation

For the relocation operation, the fragmentation and combination algorithm are successively applied on a molecule. In this way, the molecule is split into fragments which are randomly recombined. If one of the two algorithm fails, the whole relocation operation fails, and another mutation operation is randomly selected.

### 5.3.1.11 Addition of functional groups

Functional groups that can be added to a molecule are stored in a pre-defined fragment pool. From the pool of fragments, one fragment is randomly chosen and randomly combined with the original molecule using the combination algorithm. If the combination algorithm fails, the addition of functional group fails, and another mutation operation is randomly chosen.

### 5.3.2 Computational details

The GA *PSEvolve* was implemented in Python 3.7 using the graph theory package `networkx`<sup>422</sup> (version 2.3.6) and the cheminformatic package `RDKit`<sup>415</sup> (2022.03.5). A Lenovo IdeaPad 5 14ALC05 (AMD Ryzen 7 5700U with Radeon Graphics 1.80 GHz and 16 GB RAM) with a Windows operating system was used to perform all calculations.

### 5.3.3 Advantages and limitations

The graph-based GA *PSEvolve* developed in this thesis is a stochastic molecular design algorithm. By combining graph and valence theory, only structurally feasible molecules are generated, which is highly advantageous for the efficient exploration of the chemical space. In contrast to other graph-based GAs, as proposed by Jensen et al.,<sup>325</sup> *PSEvolve* allows for a higher number of structure altering operations and stochastic elements. Furthermore, *PSEvolve* can be flexibly adjusted e.g. by selecting preferred atom types or functional groups that should be explored. A further advantage of *PSEvolve* is the possibility for integrating different types of property models, e.g. QSPR models, neural networks, or group contribution models. Especially neural networks gained popularity in recent years due to their high speed and accuracy (when trained on a high-quality dataset). However, their integration into deterministic optimisation methods requires translating the trained neural network into optimisation formulations suitable for the solver software which is tedious and not always possible. In *PSEvolve*, no reformulation is required and neural networks can be seamlessly integrated. Besides solvents, *PSEvolve* can potentially be applied for other substance classes, e.g. drugs, when suitable constraints are introduced and appropriate property predictions models are available. However, genetic algorithms such as *PSEvolve* are not global optimisation methods. Therefore, the overall best performing molecule might not be always generated. Furthermore, the efficiency is highly determined by the selected hyperparameters. In contrast to the proposed solvent screening method, *PSEvolve* does not rely on a pre-defined solvent database and navigates the chemical space autonomously based on principles similar to natural selection.

## 5.4 Summary

In this chapter, two complementary methodologies for solvent selection for biomass fractionation processes were developed: a solvent screening, and a solvent design approach. Using the solvent screening approach, a database containing more than 8000 solvent molecules is screened for the optimal solvent candidate. From the database, unsuitable molecules are successively deleted. For thermodynamic property predictions (e.g. solubility of the biomolecules, partition coefficients, or LLE phase compositions), COSMO-RS is applied. For COSMO-RS predictions, representative model molecules for each biomass fractions are defined and modeled using DFT methods. The performance of the solvent candidates as defined by the objective function is evaluated. To identify the optimal solvent candidate within database, candidates are ranked based on their performance. In this manner, readily available solvents with high potential application in biorefinery processes can be identified. The developed GA for solvent design, called *PSEvolve*, allows for the generation of molecules with desired target properties. Since the explored space is not limited to a predefined database, also undiscovered molecular structures can be explored. Hence, the solvent design method allows for a broader exploration of the chemical space than the screening method. *PSEvolve* is a GA that is based on evolutionary strategies (selection, cross-over, and mutation) on the graph of the molecules and optimises the molecular structures towards the target properties. By combining graph and valence theory, structural feasibility of the designed structures is ensured and the chemical space can be efficiently explored. The generation of so-far undiscovered structures offers opportunities for the targeted synthesis of highly promising solvent candidates. Even if no suitable synthesis route is available or the designed solvent is not commercially available, further analysis of generated structures enhance the understanding of structural features and their relation to the desired properties. In the following chapters, the developed methods are applied to solvent selection problems in lignocellulose and microalgal biorefineries.



## 6 | Solvent screening and design for ligno-cellulose biorefineries

Solvent selection is highly important for environmentally friendly, safely operating, and economically feasible lignocellulose biorefineries. Organosolv processing and lignin-first approaches, including AAF, rely on the use of solvents to obtain separate cellulose, hemicellulose sugars, and lignin streams (Section 2.2). In all of these processes, as a first step, lignin is extracted from the biomass. Conventional organosolv pulping leads to lignin condensation, mainly caused by the acidic conditions and high temperatures applied in the process (Section 2.2.3). Subjecting such condensed lignins to lignin depolymerisation *via* hydrogenolysis leads to low monomer yields due to the low residual  $\beta$ -O-4 content. Condensed lignins, such as Kraft or organosolv lignins, are therefore not suitable for conversion to aromatic chemicals. Therefore, these condensed lignins are rather applied as materials, such as thermoplastics, films, or nanoparticles. In AAF, aldehydes are added to minimise lignin condensation (Section 2.2.4.1). Here, lignin's  $\alpha$ - and  $\gamma$ -hydroxyl groups react with the aldehyde to form a stable acetal, thus protecting the  $\beta$ -O-4 bonds.

Mainly alcohols (EtOH, 2-propanol, 1-butanol) are applied in organosolv pulping and reductive catalytic fractionation due to their low price, and benign EHS properties. However, the lignin solubility in alcohols is rather low. In AAF, solvent selection is particularly difficult due to potential reactions between the solvent and the aldehyde. So far, mainly cyclic ethers are known to offer stability under AAF conditions.<sup>38,155</sup> Here, the solvent selection is a trade-off between the toxic and carcinogenic 1,4-dioxane with high lignin solubility, and the more benign 2-MeTHF with a remarkably lower lignin solubility. Moreover, halogenated solvents such as chloroform are not recommended for application on the industrial scale due to their health hazards and low lignin solubility. Solvents are not only applied for lignin isolation from biomass, but also for lignin upgrading to materials. Here, solvents with high lignin solubilities such as DMSO, are commonly applied.<sup>100,101</sup> Lignin dissolution is a crucial step in lignocellulose fractionation and subsequent lignin upgrading. Hence, in both fields, solvents with high lignin solubilities are sought after.

In this chapter, the developed solvent screening (Section 5.2) and design method (Section 5.3) are applied to identify solvents for lignocellulose biorefineries. In Section 6.1, a database containing more than 8000 structures is screened for solvents with high solubilities of the biomass fractions and benign EHS properties. The solvent screening is evaluated for different process configurations, including targeted lignin extraction (as performed in organosolv, reductive catalytic fractionation, and AAF processing), and for the joint dissolution of multiple fractions (followed by subsequent precipitation of cellulose<sup>425–427</sup> and lignin from the process liquor.<sup>36</sup>) The solvent screening is useful to obtain an overview of already synthesised, commercially available structures. However, the screening results showed that for AAF where the solvent selection is additionally constrained by stability towards aldehydes, the search space must be expanded. In Section 6.2, the developed design algorithm is applied to design tailored solvents for lignin dissolution in lignin upgrading and AAF.

*This chapter is based on the publications of König-Mattern et al.<sup>395,396</sup> and includes direct as well as modified excerpts used under Creative Commons license CC BY 4.0. I kindly acknowledge the implementation, training, and testing of a graph neural network for the prediction of lignin solubilities by Edgar I. Sanchez Medina. He also applied the attribution method to the solvent structures. Furthermore, I would like to acknowledge the experimental work of Anastasia O. Komarova, who provided several datapoints of experimentally determined lignin solubilities, as well as the yields of aldehyde-protected sugars and lignin monomers from AAF.*

## 6.1 Solvent screening

In the following section, the solvent screening procedure is explained. As a first step, the solvent screening was formulated as an optimisation problem which guides the solvent selection (Section 6.1.1). Model molecules representing the original structures of cellulose, hemicellulose, and lignin were identified (Section 6.1.2) and later used in COSMO-RS solubility predictions. Finally, the database was screened and the identified candidates were ranked with respect to different process configurations, including targeted lignin dissolution, and the joint dissolution of multiple biomass fractions. (Section 6.1.3).

### 6.1.1 Optimisation problem and screening procedure

Based on the optimisation problem stated in Eqs. 5.1-5.3, the solvent screening for lignocellulose biorefineries was formulated as follows

$$\min_{\mathbf{s} \in \mathcal{S}_{\text{database}}} F(\mathbf{t}(\mathbf{b}, \mathbf{s}), \mathbf{e}(\mathbf{s})) \quad (6.1)$$

$$\text{s.t.} \quad T_m(\mathbf{s}) \leq 70 \text{ }^\circ\text{C}, \quad (6.2)$$

$$70 \text{ }^\circ\text{C} \leq T_b(\mathbf{s}) \leq 200 \text{ }^\circ\text{C}, \quad (6.3)$$

$$N_{\text{amines}}(\mathbf{s}) = 0, \quad (6.4)$$

$$Q(\mathbf{s}) = 0. \quad (6.5)$$

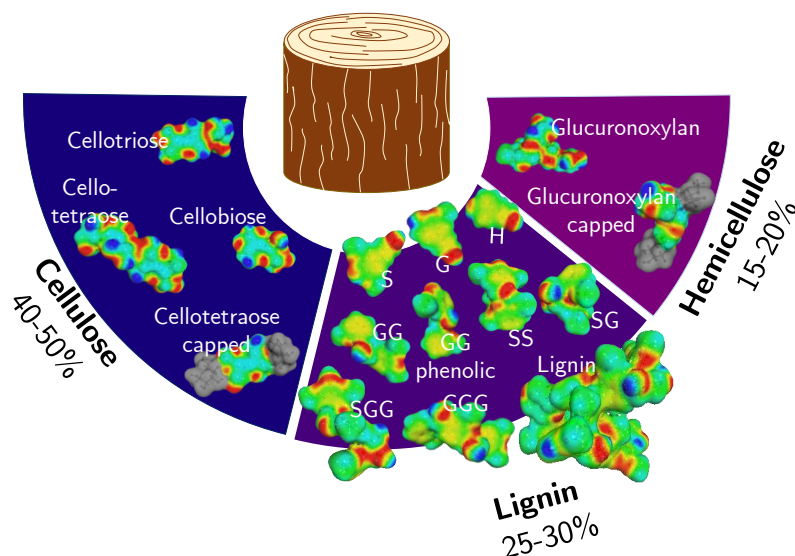
where the thermodynamic properties  $\mathbf{t}$  are defined as the molar solubilities  $x_C$ ,  $x_H$  and  $x_L$  of the representative biomolecules  $\mathbf{b}$  (cellulose C, hemicellulose H, and lignin L) in the solvents  $\mathbf{s}$  contained in the database  $\mathcal{S}_{\text{database}}$  (see Section 5.2.1 for details regarding the constructed database). The EHS properties  $\mathbf{e}$  are defined by the EHS score (Section 5.2.4). Since the process temperatures in lignocellulose fractionation usually range between 70 and 200 °C (Section 2.2), the melting temperature of the solvents  $T_m$  was limited to  $\leq 70$  °C and the boiling temperature of the solvents  $T_b$  (at atmospheric pressure) was limited to  $70 \text{ }^\circ\text{C} \leq T_b \leq 200 \text{ }^\circ\text{C}$ . Furthermore, structural constraints were introduced. Due to their potential reactivity with the biomass, solvents containing primary and secondary amines were excluded during the screening procedure ( $N_{\text{amines}} = 0$ , see Eq. 6.4). Primary and secondary amines are alkaline, nucleophilic functional groups that likely react with carbonyl compounds. Additionally, aromatic amines are highly reactive in electrophilic aromatic substitution. Furthermore, the solvent should have a charge  $Q$  of net zero (Eq. 6.5). Therefore, electroneutral mixtures of ions (e.g. ILs) were eligible solvent candidates, however, single ions were excluded.

The objective function  $F$  was defined as the Euclidian distance  $d(\mathbf{o}, \mathbf{i})$  of the properties of interest  $\mathbf{i}$  from the optimal point  $\mathbf{o}$  (see Eq. 5.9). Here, different properties of interest were studied depending on the desired process configurations (which are explained in more detail in the Sections 6.1.3.1-6.1.3.3). The objective function  $F$  includes the solubilities of the biomass fractions and the EHS score, and measures the performance of each solvent necessary for the solvent ranking. The solubilities  $x_C$ ,  $x_H$ , and  $x_L$  were predicted using COSMO-RS at  $T = 70$  °C according to Eq. 5.5-5.6. This temperature reflects preferred, rather mild processing conditions. For the solubility predictions, representative molecules for each biomass fraction were identified (Section 6.1.2). The EHS score was predicted as described in Section 5.2.4.

To solve the optimisation problem, first, solvent candidates not fulfilling the constraints given in Eqs. 6.2-6.5 were removed from the database  $S_{\text{database}}$ . First, the number of primary and secondary amines  $N_{\text{amines}}$  and the net charge  $Q$  of the solvent were determined and unsuitable compounds were eliminated. Subsequently, solvent candidates with unsuitable  $T_m$  and  $T_b$  were removed. For each solvent candidate meeting the constraints, the objective function  $F$  is evaluated for each scenario described in the Sections 6.1.3.1-6.1.3.3, and the solvents are ranked according to the minimum distance  $d(\mathbf{o}, \mathbf{i})$ . After the ranking,  $T_m$ ,  $T_b$ , and EHS properties of the identified candidates are compared to literature data to reassure that the constraints imposed on the optimisation problem were indeed fulfilled.

### 6.1.2 Representative lignocellulose molecules

The COSMO-RS solubility predictions of  $x_C$ ,  $x_H$ , and  $x_L$  require QC calculations of the biomass compounds as described in Section 5.2.5. Since cellulose, hemicellulose, and lignin are large, complex molecules (Section 2.1) with high molecular weights ( $> 150$  kDa for cellulose,  $> 30$  kDa for hemicelluloses,  $> 4$  kDa for lignin)<sup>83,428</sup> modeling their native structures is computationally infeasible. To reduce the computational time and structural complexity, fragments of the original biomass fractions were modeled individually (Fig. 6.1).



**Fig. 6.1** Representative molecules investigated for the cellulose, hemicellulose, and lignin fraction. The abundance of each fraction within the biomass is given as wt.% on a dry matter basis. Adapted from König-Mattern et al.<sup>395</sup> with permission from Elsevier.

Several representative molecules for the cellulose fraction were proposed in the literature,

including glucose,<sup>307,311</sup> cellobiose,<sup>307–309</sup> cellotriose,<sup>306,307</sup> and cellotetraose.<sup>307</sup> Furthermore, capped cellotriose<sup>306</sup> and cellotetraose<sup>310</sup> molecules were used as representative structures. Here, QM-calculations were performed for the respective polymer fragment and the end-groups were subsequently truncated, since their influence compared to the original polymer chain is usually low. Chu et al. reported a higher correlation between the COSMO-RS predicted excess enthalpies and experimentally determined cellulose solubilities of cellobiose and cellotetraose in ILs than with glucose and cellotriose.<sup>307</sup> Other studies highlighted the importance of intramolecular hydrogen-bonding which differs between the different cellulose-representing molecules.<sup>308,309</sup> According to Yamin, the hydrogen-bonding is better captured when not only one mid glucose-unit as in the truncated cellotriose molecule is considered, but two mid glucose-units as in cellotetraose.<sup>310</sup> A more complex cellulose-representing structure was modelled by Casas et al.<sup>312</sup> Here, not only a single cellulose chain was considered, but several fragments of adjacent cellulose chains to capture intracellular and intermolecular interactions of cellulose.<sup>312</sup> However, conformers were not taken into account, probably due to the time-consuming calculations. In this thesis, cellobiose, cellotriose, cellotetraose, and a cellotetraose molecule with capped end-groups were assessed for their ability to represent cellulose in COSMO-RS solubility predictions. Conformational search was applied to all modeled biomolecules (Section 5.2.5). Glucose was not considered as a representative molecule since its thermodynamic properties remarkably differ from cellulose, e.g. the solubility of glucose in water is 47.8 wt.% ( $T = 20\text{ }^{\circ}\text{C}$ ),<sup>429</sup> while cellulose is insoluble in water at the same conditions.<sup>66</sup>

For the lignin fraction, predominantly monolignols were selected as representative structures in recent studies selected.<sup>52,304,308,309,311</sup> Also, the dimeric structures pinoresinol and guaiacylglycerol-2-coniferyl ether were considered as lignin representatives.<sup>312</sup> A more complex structure was modelled by Achinivu et al., consisting of ten aromatic units connected *via* different linkage-motifs aiming to mimic herbaceous biomass.<sup>305</sup> In this thesis, several lignin-representing molecules were assessed, including monolignols, as well as dimers and trimers composed of S- and G-units connected *via*  $\beta$ -O-4 bonds were modelled. G- and S-units are present in several lignocellulosic feedstocks, including hardwood, softwood, and herbaceous lignin. Only herbaceous lignin contains noteworthy amounts of H-units (up to 35% of all aromatic units). Additionally, one conformer of a  $1,500\text{ g mol}^{-1}$  lignin fragment (8 G- and S-units, connected *via*  $\alpha$ -O-4,  $\beta$ -O-4,  $\beta$ - $\beta$ , and  $\beta$ -5 bonds) was modelled to study differences compared to the truncated representatives in COSMO-RS solubility predictions.

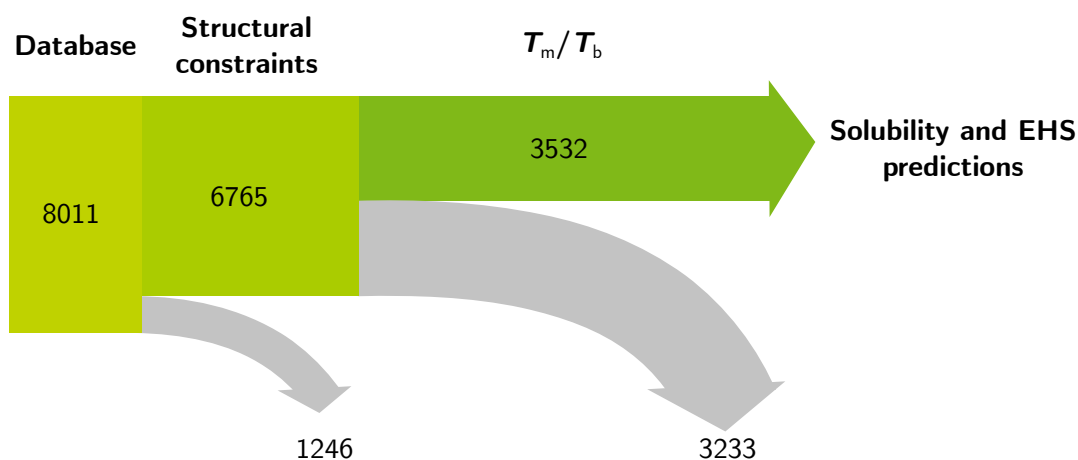
In the recent literature, hemicelluloses were represented by a mixture of glucose and xylose monomers.<sup>304,308,309</sup> However, the polymeric character of hemicelluloses was not taken into account so far. Therefore, in this work, glucuronoxylan with and without capped

end-groups were chosen as a representative structures. Glucuronoxylan is primarily present in dicots.

The chemical structures of all representative molecules are summarised in the Appendix (Tab. C.1). These structures were modeled on a quantumchemical level as described in Section 5.2.5. From these 16 initial structures, the most suitable representative molecules were identified by analysing their  $\sigma$ -profiles (Appendix C.1.1.1) and by correlating COSMO-RS solubility predictions with experimental solubility data (Appendix C.1.1.2). Based on these analyses, the capped cellotetraose molecules was selected as a cellulose-representative, the capped glucuronoxylan molecules was chosen as a model molecule for the hemicellulose fraction, and all monolignols, dimers and trimers were selected as lignin-representatives. Overall, the COSMO-RS solubility predictions and the experimental data were in qualitative agreement. Although COSMO-RS was not able to accurately predict absolute solubility values, nevertheless, a qualitative solvent comparison was possible (Appendix C.1.1.2). Since a relative solvent comparison is sufficient for the aims of a solvent screening, COSMO-RS is a suitable method for solubility predictions.

### 6.1.3 Screening results and solvent ranking

84% of the initial 8011 solvent candidates met the structural constraints described in Section 6.1.1. After screening for solvents with favourable  $T_m$  and  $T_b$ , 3525 solvent candidates remained for COSMO-RS solubility predictions using the selected representative molecules (Fig. 6.2). The solubility predictions revealed distinct solubility profiles (Fig. 6.3 a), opening possibilities for various process configurations. Several solvent candidates were predicted to simultaneously dissolve multiple biomass fractions. In such a biorefinery scenario, first, all biomass fractions are dissolved and subsequently separated, e.g. by precipitation from the process liquor. Other solvent classes are rather applicable for selective lignin dissolution. Such solvents could be integrated into organosolv processing or AAF, provided that these solvents are non-reactive towards other compounds (e.g. acids, or aldehydes) the process liquor. In addition, solvents with high lignin solubilities have high potential for dissolution-based lignin upgrading (e.g. nanoparticles, lignin films). Interestingly, no solvent candidate within the search space was able to selectively dissolve cellulose. Based on these results, three different objective functions are formulated to evaluate the solvent candidates for different process configurations: In Section 6.1.3.1, solvents with high solubilities of cellulose-, hemicellulose-, and lignin representing molecules, and a high EHS score were ranked. Here, only few solvent candidates with lignin solubilities comparable to the benchmark solvent 1,4-dioxane were discovered. To explore more suitable candidates, the objective function  $F$  was simplified by considering only cellulose and lignin solubilities (Section 6.1.3.2). The hemicellulose solubility was neglected since hemicellulose is depolymerised, and its constituent



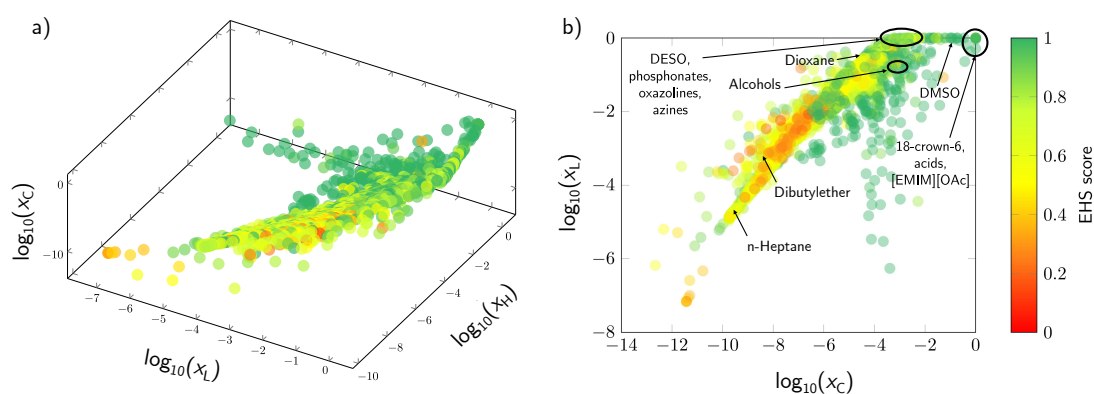
**Fig. 6.2** Numbers of eligible solvent candidates after each screening step. Reprinted from König-Mattern et al.<sup>395</sup> with permission from Elsevier.

sugars are solubilised under commonly applied process conditions. Furthermore, the EHS score was not used as a ranking criterion. In Section 6.1.3.3, solvents with high lignin solubility and low solubilities of the carbohydrate fractions were sought after, as commonly applied in organosolv processing, reductive catalytic fractionation, or AAF. In the following sections, the best performing solvent candidates for each objective are presented and potential process configurations are discussed. A solvent ranking for each objective is included in Appendix C.1.2.

### 6.1.3.1 Objective 1: Solvents with high EHS score for the joint dissolution of all lignocellulose fractions

To identify solvents with benign EHS properties and high solubilities for all biomass fractions, the optimal point was defined as  $\mathbf{o}(\mathbf{t}, \mathbf{e}) = [x_C, x_L, x_H, \text{EHS score}]^T = [1, 1, 1, 1]^T$  and the solvent candidates were ranked according to their minimal distance from that point (Eq. 5.9). Fig. 6.3 a shows that few candidates exist in proximity of the optimal point. Most of these candidates were ILs which are listed in Tab. 6.1. The two best performing ILs were [Chol][OH] and [P666,14][BTMP]. [Chol][OH] is an alkaline IL which was already proposed for the valorisation of agricultural waste.<sup>430</sup> No experimental solubility data for [P666,14][BTMP] was found in the literature. However, phosphonium-based ILs are known to dissolve lignocellulosic biomass well.<sup>431,432</sup> Furthermore, the IL [EMIM][OAc] was predicted to solubilise all biomass fractions. In agreement with literature reports, this IL is able to completely dissolve wood chips,<sup>433</sup> and also other classes of ILs are able to dissolve all biomass fractions.<sup>425</sup> Furthermore, Zavrel et al. dissolved different wood species in [EMIM][OAc] and [MMIM][DMP].<sup>434</sup> High cellulose solubilities for [EMIM][OAc] and [BMIM][OAc] were reported, dissolving 8

and 14 wt.% cellulose, respectively.<sup>435</sup> For the DES included in the database, generally low cellulose solubilities were predicted, leading to a large  $d(\mathbf{o}, \mathbf{i})$ . Nevertheless, choline chloride in combination with imidazole was ranked the highest among the DESs and was reported to dissolve 2.48 wt.% of cellulose.<sup>436</sup> Also for the screened DES, the COSMO-RS solubility predictions were qualitatively in line with literature data. Furthermore, the organic solvent 18-crown-6 ether was predicted to dissolve all biomass fractions. According to its SDS, the crown ether is only slightly toxic (acute toxicity category 4 according to the Environmental Protection Agency) and was not listed as a carcinogen.<sup>437</sup> Data regarding ecotoxicity and flammability is still lacking. Other organic solvent candidates were predicted to have remarkably lower cellulose solubilities, leading to a lower rank. Overall, of the 8011 potential solvent candidates, only 8 ILs and the 18-crown-6 ether were close to the defined optimal point. The distance from the optimal point increased rapidly after the first four ranks, mainly caused by low cellulose solubilities.



**Fig. 6.3** Solubilities of the biomass fraction for all of the 3525 solvent candidates that were identified using the screening procedure. The datapoints are colour coded corresponding to their EHS score. a) Logarithmic molar solubilities of cellulose ( $\log_{10}(x_C)$ ), lignin ( $\log_{10}(x_L)$ ), and hemicellulose ( $\log_{10}(x_H)$ ). b) Logarithmic molar solubilities of cellulose ( $\log_{10}(x_C)$ ), lignin ( $\log_{10}(x_L)$ ). The locations of distinct solvent candidates are highlighted. Adapted and reprinted from König-Mattern et al.<sup>395</sup> with permission from Elsevier.

### 6.1.3.2 Objective 2: Solvents for the joint dissolution of lignin and cellulose

To promote the exploration of further solvent candidates, the search criteria were relaxed and the optimal point was defined as  $\mathbf{o}(\mathbf{t}) = [x_C, x_L]^T = [1, 1]^T$ , neglecting the hemicellulose solubility and the EHS score. Similar to the previous section, the ILs given in Tab. 6.1, and additionally the solvent candidates 1-methylpiperidine-1-oxide, 18-crown-6 ether, and dimethylselenoxide were identified (see Tab. 6.2 for their chemical structures).



**Tab. 6.1** Ranking of ILs for the joint dissolution of cellulose, lignin and hemicellulose. The distance from the optimal point  $\mathbf{o}(\mathbf{t}, \mathbf{e}) = [x_C, x_L, x_H, \text{EHS score}]^T = [1, 1, 1, 1]^T$  is given as  $d(\mathbf{o}, \mathbf{i})$ . Solubilities for each biomass fraction are given in logarithmic scale. The  $T_m$  was obtained from lolitec,<sup>410</sup>  $T_b$  was unknown. RT indicates room temperature. EHS properties were taken from PubChem.<sup>414</sup> <sup>a</sup>

IL	$d(\mathbf{o}, \mathbf{i})$	$\log_{10}(x_C)$	$\log_{10}(x_L)$	$\log_{10}(x_H)$	$T_m$ [°C]	EHS properties
[Chol][OH]	0.00	0.00	0.00	0.00	n.a.	n.a.
[P666,14][BTMP]	0.00	0.00	0.00	0.00	< RT	irritant, corrosive, environmental hazard
[BMIM][OAc]	0.08	0.00	-0.04	0.00	< RT	n.a.
[EMIM][OAc]	0.49	0.00	-0.29	0.00	< RT	irritant
[MMIM][DMP]	0.70	-0.20	-0.37	-0.07	< RT	corrosive, irritant
[EMIM][DMP]	0.75	-0.37	-0.18	-0.18	23	corrosive, irritant
[EMIM][DEP]	0.77	-0.46	0.00	-0.23	< RT	corrosive, irritant
[BMIM][DBP]	1.01	-0.72	0.00	-0.40	< RT	irritant
[P666,14][Dec]	1.05	-0.96	0.00	-0.35	~ RT	corrosive




<sup>a</sup> Choline ([Chol]), hydroxide ([OH]), trihexyltetradecylphosphonium ([P666 14]), bis(2,4,4-trimethylpentyl)phosphinate ([BTMP]), 1-butyl-3-methylimidazolium ([BMIM]), acetate ([OAc]), 1-ethyl-3-methylimidazolium ([EMIM]), 1,3-dimethylimidazolium ([MMIM]), dimethyl phosphate ([DMP]), diethyl phosphate ([DEP]), dibutyl phosphate ([DBP]), decanoate ([Dec]).

1-methylpiperidine-1-oxide is an interesting solvent candidate due to its structural similarities to the solvent NMMO which is applied for cellulose dissolution in the Lyocell process for fibre spinning.<sup>438</sup> However, 1-methylpiperidine-1-oxide suffers from low thermal stability and a high price. To date, there is no experimental data available for lignocellulose processing with 1-methylpiperidine-1-oxide available. The identified dimethylselenoxide has even stronger HBA properties than its sulfur analog DMSO,<sup>439</sup> potentially increasing the solubility of lignin. However, this candidate is highly toxic and an environmental hazard,<sup>414</sup> and is for this reason not further considered. Similar to objective 1, no DES was able to compete with the solvents listed in Tab. 6.1 and 6.2 for the given criteria.

### 6.1.3.3 Objective 3: Solvents for selective lignin extraction

Selective lignin extraction is a key step in several lignocellulose-based biorefinery processes, such as organosolv processing, AAF, or reductive catalytic fractionation (see Section 2.2). In these processes, lignin is extracted from the biomass and dissolved by the employed solvent, while the cellulose fraction remains as an easily separable solid. Hence, for this objective, the optimal point was defined as  $\mathbf{o}(\mathbf{t}) = [x_C, x_L]^T = [0, 1]^T$ . To explore many solvents with matching solubility characteristics, the EHS score was neglected at this stage. However, the EHS properties as given in the SDSs were considered during the solvent

**Tab. 6.2** Solvent ranking for joint lignin and cellulose dissolution. EHS criteria and hemicellulose solubilities were neglected in the ranking. The distance from the optimal point  $\mathbf{o}(t) = [x_C, x_L]^T = [1, 1]^T$  was zero for all solvents presented in the table. Unless otherwise stated, all properties were obtained from the respective safety data sheets.

Solvent	$T_m$ [°C]	$T_b$ [°C]	EHS properties	Structure
1-Methylpiperidine-1-oxide	n.a.	n.a.	n.a.	
18-Crown-6 ether	42-45	118 (0.07 mbar) <sup>440</sup>	irritant	
Dimethylselenoxide	n.a.	n.a.	acute toxicity, health hazard, environmental hazard	

selection for experimental testing (see Section 6.3).

In total, 104 organic solvents, one IL, and four tetrabutyl ammonium-based DESs with a distance of  $\leq 0.5$  from the optimal point were identified. Commercially available solvents with a price  $\leq 100$  € per 10 g were identified for this objective (Tab. 6.3). Fig. 6.3 b shows that several solvent candidates suitable for selective lignin dissolution also have benign EHS properties. Furthermore, many structural similarities among the most promising solvent candidates were identified. These candidates belong to four major classes: azines (pyridines, pyrazines, pyrimidines, pyridazines, triazines), sulfoxides, oxazolines, and phosphonates. Many of these candidates were predicted to have higher lignin solubilities and more benign EHS properties than the toxic benchmark 1,4-dioxane (the EHS score for most of the identified chemicals from Tab. 6.3 is  $> 0.8$ ). Alcohols, such as EtOH, 2-propanol or 1-butanol, that are commonly applied in reductive catalytic fractionation and in industrial organosolv processes, showed comparably low lignin solubilities.

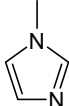
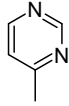
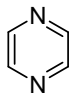
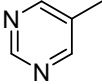


The screening procedure was additionally applied to identify solvents with high lignin solubilities, regardless of the solubilities of other fractions. This additional objective was defined by the optimal point  $\mathbf{o}(t) = x_L = 1$ . However, the same solvents as in Tab. 6.3 were discovered, only with slightly altered ranking due to the neglected cellulose solubilities (data not shown). The identified solvents could also be applied in many areas of lignin upgrading, including the formation of lignin films or nanoparticle fabrication, where solvents with high lignin solubilities are essential.

**Tab. 6.3** Solvent ranking for selective lignin extraction. The distance from the optimal point  $\mathbf{o}(\mathbf{t}) = [x_C, x_L]^T = [1, 0]^T$  is given as  $d(\mathbf{o}, \mathbf{i})$ . Only solvent candidates with  $d(\mathbf{o}, \mathbf{i}) \leq 0.2$  are shown. Unless otherwise indicated,  $T_m$ ,  $T_b$ , and EHS properties were obtained from the respective SDSs.

Solvent	$d(\mathbf{o}, \mathbf{i})$	$\log_{10}(x_C)$	$\log_{10}(x_L)$	$T_m$ [°C]	$T_b$ [°C]	EHS properties	Structure
2-Methyl-2-oxazoline	0.00	-2.89	0.00	n.a.	100	flammable	
Pyrrolidine-1-carbaldehyde	0.00	-3.21	0.00	n.a.	92	irritant <sup>414</sup>	
Diethyl-methyl-phosphonate	0.00	-2.74	0.00	n.a.	194	irritant	
Diethyl-sulfoxide	0.00	-2.51	0.00	14 414	n.a.	n.a.	
4-Piperidino-pyridine	0.01	-2.33	0.00	78	n.a.	toxic, corrosive	
Dimethyl-methyl-phosphonate	0.01	-2.22	0.00	< 50 414	181 414	flammable, irritant, health hazard 414	
4-Pyrrolidino-pyridine	0.01	-2.31	0.00	54	n.a.	corrosive, acute toxicity	
Dimethyl-sulfoxide	0.02	-1.66	0.00	16	189	irritant	
4-Methoxy-pyridine	0.04	-2.84	-0.02	n.a.	n.a.	irritant	
Dimethyl ethyl-phosphonate	0.04	-3.24	-0.02	n.a.	n.a.	toxic	
5-bromo-1-methyl-1h-imidazole	0.05	-2.46	-0.02	40	110	irritant	
n,n-Dimethyl-acetamide	0.07	-3.41	-0.03	-20	164	irritant, health hazard	
Pyrimidine	0.08	-2.57	-0.04	19	123	flammable	
5,5-Dimethyl-1-pyrrolin-n-oxide	0.10	-3.49	-0.04	25	78 (0.5 hPa)	n.a.	

Continued on next page

Tab. 6.3 – Continued from previous page

Solvent	$d(\sigma, i)$	$\log_{10}(x_C)$	$\log_{10}(x_L)$	$T_m$ [°C]	$T_b$ [°C]	EHS properties	Structure
1-Methylimidazole	0.12	-0.94	0.00	-6	198	toxic, corrosive	
4-Methylpyrimidine	0.12	-3.14	-0.06	n.a.	141	flammable	
Pyrazine	0.14	-2.82	-0.06	50	115	flammable	
5-Methylpyrimidine	0.15	-3.16	-0.07	33	n.a.	flammable, irritant	
Dimethylformamide	0.15	-3.25	-0.07	-61	153	flammable, health hazard, irritant	
Pyridine	0.20	-3.14	-0.10	-42	115	flammable, irritant	

## 6.2 Solvent design for lignin dissolution in lignin-first biorefineries and lignin upgrading

The solvent screening approach identified azines, sulfoxides, oxazolines, and phosphonates as promising solvent candidates for selective lignin dissolution. Solvents from these classes are suitable for application in the fabrication of lignin nanoparticles, films, or fibres. However, oxazolines and azines are unstable in the presence of acids, limiting their use in organosolv pulping and AAF. To explore potential solvents independent of a predefined database, the *GA PSEvolve* (Section 5.3) was applied for solvent design tailored towards lignin dissolution and AAF. Lignin dissolution is an essential step in lignin isolation from biomass but also for dissolution-based lignin upgrading (e.g. nanoparticle, film, and fibre production). Designing tailor-made solvents with high lignin solubility could not only improve the lignin yield in lignocellulose fractionation but also open new perspectives in the field of lignin upgrading, and hence, effectively promote the utilisation of lignocellulosic feedstocks within the circular economy.

For this purpose, the solvent design problem was stated as an optimisation problem and the genetic algorithm *PSEvolve* was applied for generating tailored solvent structures (Section 6.2.1) aiming at maximum lignin solubility. For lignin solubility predictions, COSMO-RS was not suitable due to its need for time-intensive quantum chemical calculations for each newly designed solvent candidate. Therefore, a GNN for lignin solubility predictions was developed, serving as a surrogate model (Section 6.2.2). *PSEvolve* and the integrated GNN were applied to generate potential solvents aiming at maximum lignin solubility with potential application in dissolution-based lignin upgrading (Section 6.2.3). Subsequently, tailor-made solvent candidates for AAF were designed (Section 6.2.4). Here, structural constraints were introduced, preventing the design of solvent structures that are potentially reactive towards the process liquor and biomass compounds.

### 6.2.1 Optimisation problem and solvent design algorithm

Based on Eqs. 5.1-5.3, the optimisation problem for the solvent design was formulated as

$$\min_{\mathbf{s} \in \mathcal{S}} \quad -x_L(\mathbf{s}) \quad (6.6)$$

$$\text{s.t.} \quad M_s(\mathbf{s}) \leq 200 \text{ g mol}^{-1}, \quad (6.7)$$

$$\text{SAS}(\mathbf{s}) \leq 3.5, \quad (6.8)$$

$$\text{atom type}(\mathbf{s}) \in \text{C, H, N, S, O, P}, \quad (6.9)$$

$$N_{\text{reactive}}(\mathbf{s}) = 0. \quad (6.10)$$

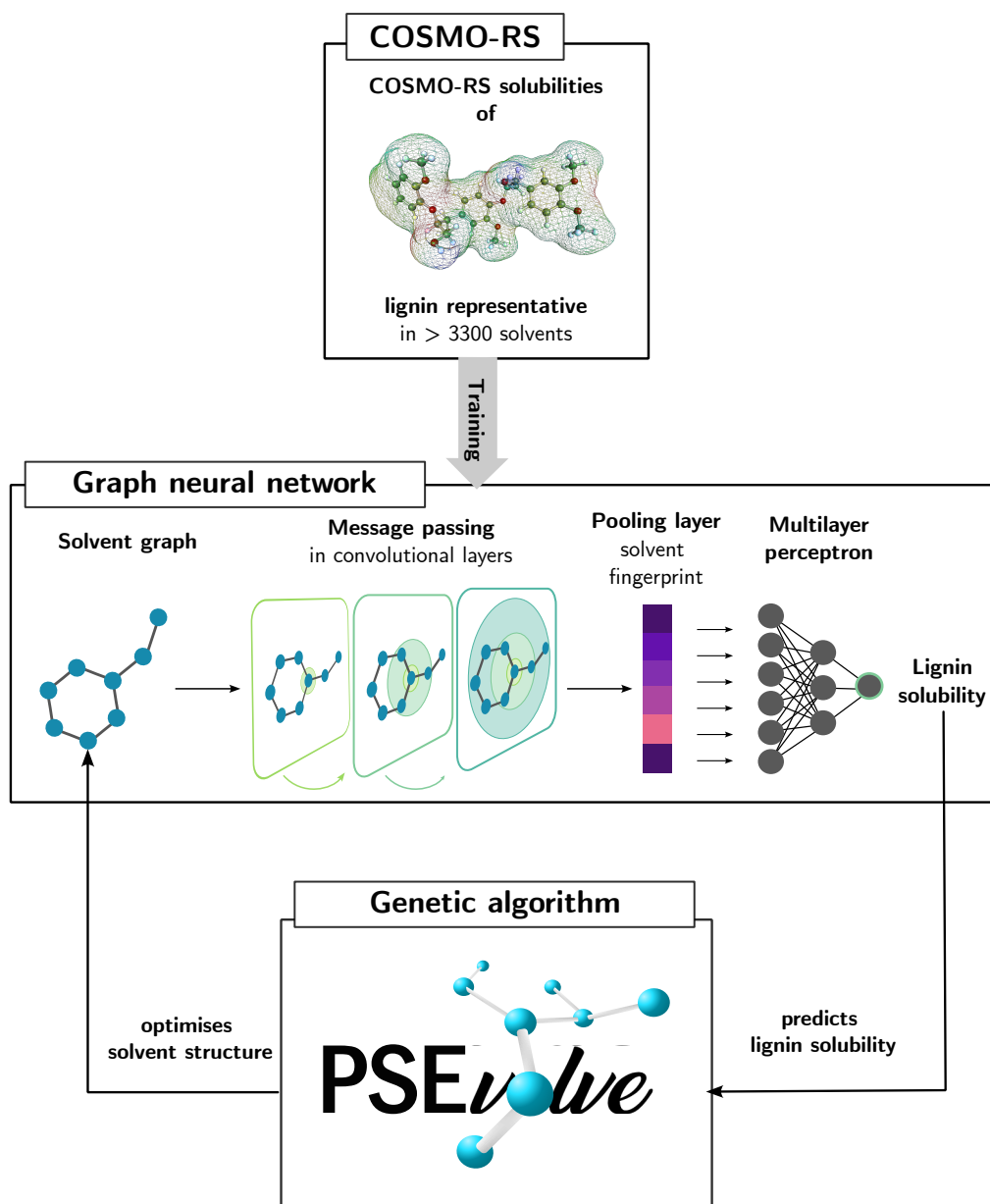
The objective of the solvent design is to maximise the lignin solubility in the solvent structures  $\mathbf{s}$ . The cellulose and hemicellulose solubilities were neglected. The high-throughput solvent screening showed that the cellulose solubility is generally low in most of the organic solvents (Section 6.1). Hemicelluloses readily depolymerise and solubilise under process conditions applied in AAF and organosolv pulping. Furthermore, the EHS properties of the solvents were not considered during the design stage but were manually assessed when specific solvent candidates were selected for experimental evaluation. During the solvent design for lignin dissolution and for AAF, the molecular weight of the solvents  $M_s$ , the synthetic accessibility score (SAS), and the atom types were restricted. Since solvents are generally rather small molecules,  $M_s \leq 200 \text{ g mol}^{-1}$  was required. The SAS is a useful metric how easily a compound can be chemically synthesised, ranging between one (easy) and ten (hard).<sup>441</sup> To spur the generation of molecules that can be easily synthesised, the SAS upper limit was 3.5. Furthermore, only C, H, N, S, O, and P-atoms were allowed in the design. Halogen atoms were excluded due to their generally high toxicity and reactivity. When applying the solvent design method to AAF, additionally the number of

functional groups that are prone to reactions with the process liquor or the biomass, also denoted as  $N_{\text{reactive}}$ , must be set to zero. In this manner, reactive functional groups were excluded, including primary and secondary amines, aldehydes, aromatic N-heterocycles, isocyanates, amides, esters, and hydrazides. Although ketones can undergo aldol condensation with aldehydes under acidic conditions, carbonyl groups were not excluded as they easily mutate to other functional groups, such as ether or C=C groups during the design.

To solve the inverse problem (Eqs. 6.6-6.10), the graph-based GA *PSEvolve* was combined with a GNN for lignin solubility predictions that was trained on COSMO-RS solubility data beforehand (Fig. 6.4). The algorithm was initiated with a population of hexane molecules and a population size of 1000 molecules. The solubility of lignin in n-hexane is generally low across different feedstocks.<sup>395,442</sup> During the iterative optimisation procedure, these suboptimal structures were designed to maximise the lignin solubility while meeting the imposed constraints. In each generation, 50 parents were chosen for cross-over to generate 50 offspring molecules. The mutation rate was set to 0.1. Molecules not meeting the constraints were directly eliminated during the cross-over and the mutation steps. The respective operation was performed until an eligible structure was generated. To evaluate the fitness of a solvent candidate, the lignin solubility was predicted by a GNN that acted as a surrogate model for COSMO-RS solubility predictions (Section 6.2.2). After 1000 generations, the GA was stopped and the resulting structures were analysed (Sections 6.2.3 and 6.2.4). Different sets of hyperparameters were tested. The hyperparameter test and the chosen set of parameters are provided in Appendix C.2.1.

## 6.2.2 Lignin solubility predictions using a graph neural network

In GNNs, chemical structures can be represented as undirected graphs  $G(V, E)$  where  $V$  is the set of vertices (also called nodes) representing the atoms, and  $E$  is the set of edges representing the bonds (Fig. 6.4). The nodes and edges are attributed by a feature vector, describing selected properties. In this work, the graph was attributed by several atom features (e.g. atom type, hybridisation, charge) and bond features (e.g. bond type, conjugation) which are summarised in Tab. C.5. To obtain additional information, the node features were subsequently updated by information about the neighbouring nodes and the connecting edges, also known as message passing step. By performing the message passing step multiple times, the node embeddings were successively enriched with information of their neighbourhood. Subsequently, the updated graph was passed through a pooling operation to yield the „molecular fingerprint“ of the solvent candidate. The obtained fingerprint acted as a tailor-made vectorial representation of the solvent candidate, and served as input to a multilayer perceptron which finally predicted the lignin solubility. Detailed information about the GNN architecture,



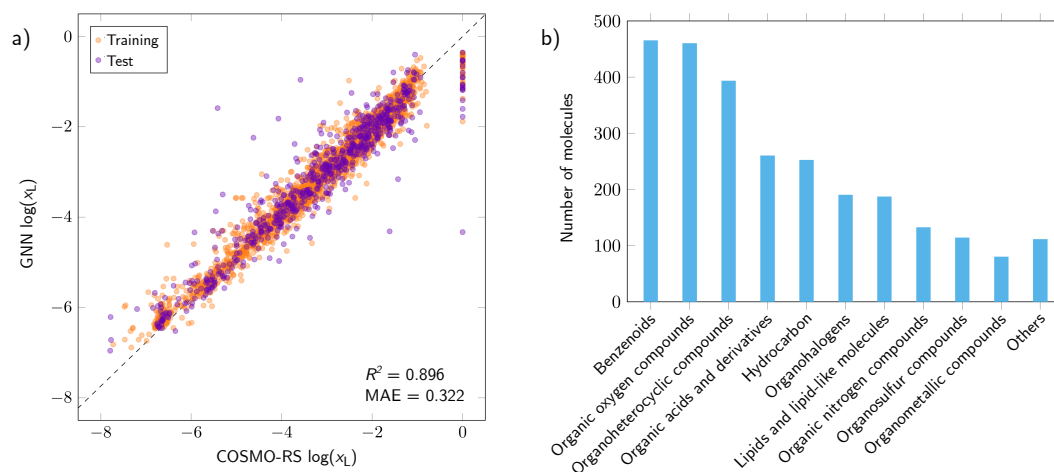
**Fig. 6.4** Connecting the molecular design algorithm *PSEvolve* with a GNN for lignin solubility predictions. The genetic algorithm *PSEvolve* optimises the structure of a molecule tailored towards high lignin solubilities as predicted by the GNN. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

hyperparameter tuning, and its applicability range are provided in Appendix C.2.1.1- C.2.1.3.

Since experimental lignin solubility data is scarce, the GNN was trained and tested on COSMO-RS solubility predictions. The COSMO-RS solubility predictions were performed using a representative lignin structure and around 3300 organic solvent candidates that were prior determined by the screening procedure ( $T = 70\text{ }^{\circ}\text{C}$ , see Section 6.1.3). ILs and DESs were excluded as their structure was unsuitable for the GNN architecture. For these 3300 solvent candidates, iterative solubility COSMO-RS lignin solubility predictions were performed to increase the accuracy (Section 5.2.3.1). To prevent time-consuming solubility calculations for all representative molecules, one lignin-representing structure was selected. A trimer of G units connected *via*  $\beta$ -O-4 bonds (GGG,  $M_{\text{GGG}} = 530.57\text{ g mol}^{-1}$ , see Tab. C.1 for chemical structure) was used for the refined COSMO-RS solubility predictions. In contrast to H- and S-units, only G-units are produced across various lignocellulose sources, including hardwood, softwood, and herbaceous biomass.<sup>60</sup> In grasses and hardwoods, G-units are less abundant compared to S- and H-units. However, G-units only differ by a methoxy group from H- and S-units and, therefore, represent intermediate structures. GGG was modelled on the quantum chemical level in accordance with Section 5.2.5.

The results of the training and testing show, that the GNN and COSMO-RS predictions are well in agreement ( $R^2 = 0.896$ , mean absolute error (MAE) = 0.322 for the test set), with slight deviations in the upper solubility ranges with  $\log_{10}(x_L) > -1$  (Fig. 6.5 a). To obtain accurate lignin solubility predictions for structurally different chemicals as generated by *PSEvolve*, the training set included structurally diverse chemicals (Fig. 6.5 b). The accuracy of the COSMO-RS solubility predictions is reduced for solubilities of  $\log_{10}(x_L) > -1$ ,<sup>412</sup> which might explain the disparity between the COSMO-RS and the GNN predictions in this region (Fig. 6.5 a). Since the objective function of the solvent design problem is to maximise the lignin solubility (Eq. 6.6), inaccuracies in the region of  $\log_{10}(x_L) > -1$  presumably hinder the identification of the optimal solvent structure. However, the main objective of this study is to identify a broad range of so-far unexplored solvent classes, rather than identifying a single optimal one. Therefore, the deviations are less impactful for the scope of this thesis and do not outweigh the advantages of using the GNN as a surrogate model of COSMO-RS. Indeed, the solvent design was only made possible by the low computational cost and suitable accuracy of the GNN.



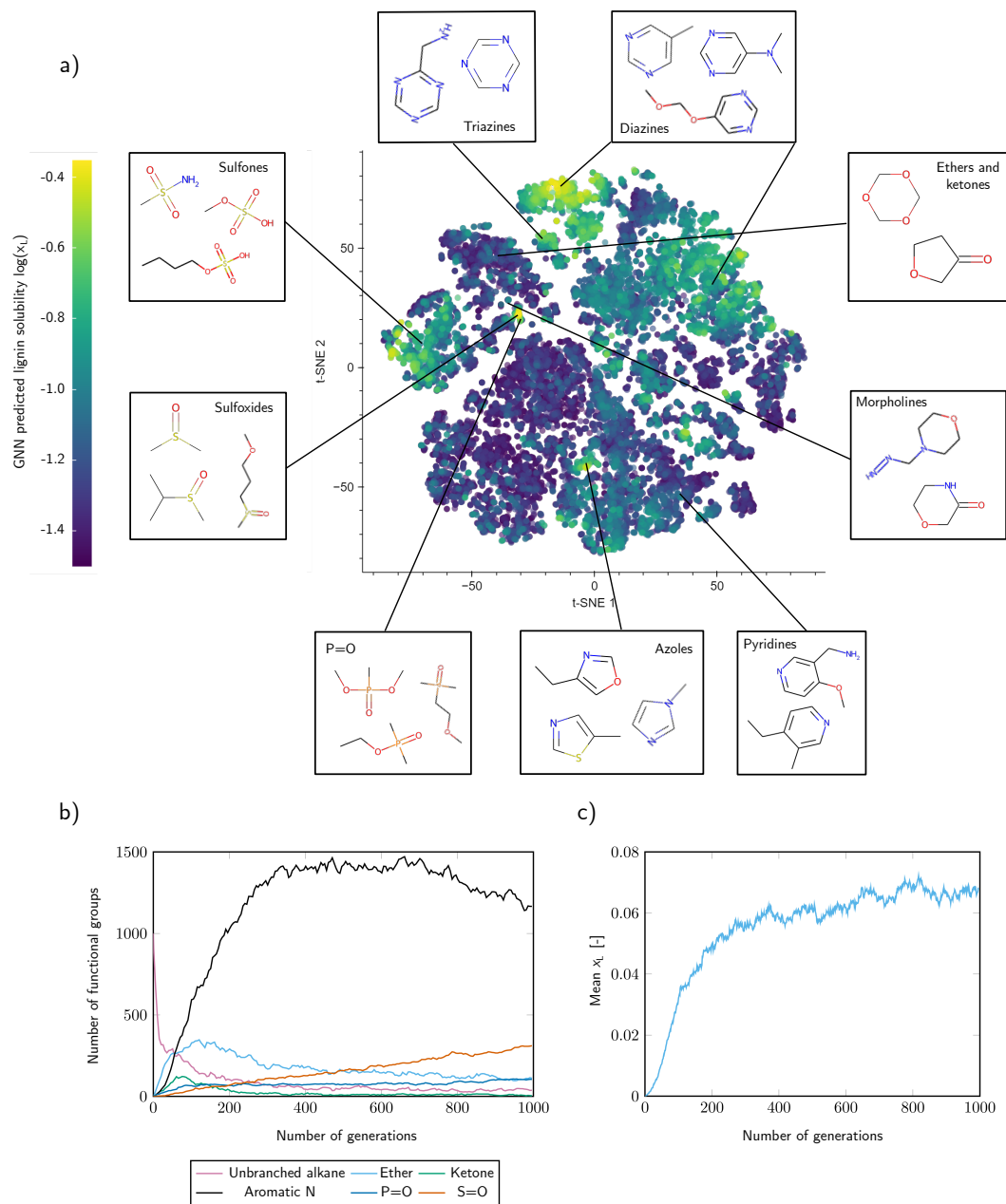


**Fig. 6.5** GNN-training with COSMO-RS solubility data for the GGG-trimer which serves as a lignin-representing structure. a) Parity plot for GNN vs. COSMO-RS predictions of the training and the test set. The coefficient of determination  $R^2$  and the MAE are given for the test set. b) Chemical classes of the training set as computed by the Classyfire toolbox. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

### 6.2.3 Solvent design for lignin dissolution

For analysing the most promising designed molecules for dissolution-based lignin upgrading, those with the highest solubilities were selected ( $\log_{10}(x_L) > -1.5$ ; around 21,000 molecules). In this manner, the relationship between molecular structure and the GNN predicted lignin solubilities was analysed by applying t-distributed stochastic neighbour embedding (t-SNE) to the GNN-generated solvent fingerprints. T-SNE reduced the complexity of the GNN fingerprint to a 2-dimensional space in which molecules with similar GNN fingerprints were located within proximity. Structurally similar molecules were predicted to have similar lignin solubilities (Fig. 6.6 a) Therefore, the GNN can be considered as a QSPR model that was trained „end-to-end“ from the molecular graph to the lignin solubility prediction. Additionally, the GNN was able to generate tailor-made molecular fingerprints optimised for lignin solubility predictions. Regions with particularly high lignin solubility predictions ( $\log_{10}(x_L) > -0.60$ ) corresponding to sulfoxides, compounds with P=O motif, sulfones, triazines, diazines, and azoles were discovered (Fig. 6.6 a). Other solvent classes promising for lignin upgrading were morpholines, cyclic ethers, and cyclic ketones. The overall fittest solvent was DMSO with  $\log_{10}(x_L) = -0.35$ , followed by 1,3,5-triazine ( $\log_{10}(x_L) = -0.36$ ), n,n-dimethylpyrimidin-5-amine ( $\log_{10}(x_L) = -0.37$ ), and dimethyl methylphosphonate (DMMP) ( $\log_{10}(x_L) = -0.38$ ). In addition to the already established lignin solvents DMSO and pyridine, commercially available azoles, such as thiazole or isoxazole, were discovered. Thiazole is only slightly toxic

(LD<sub>50</sub> oral rat: 938 mg kg<sup>-1</sup>, toxicity category III).<sup>156</sup> Toxicity data for isoxazole is currently lacking. Further aromatic N-heterocycles were designed, including triazines, diazines, pyridines, and bicyclic compounds. Common side chain motifs were methoxy-, alkyl-, and NH<sub>2</sub>-groups. Pyridines and many diazines have benign EHS properties<sup>156</sup> and are readily commercially available. Most triazines, but also sulfones and phosphonates are solid at room temperature, limiting their applicability in many areas of lignin upgrading. Cyclic ethers and ketones were associated with lower GNN-predicted lignin solubilities compared to the aforementioned solvents, however, they were predicted to have a higher lignin solubility compared to the usually applied 1,4-dioxane. During solvent design, functional groups associated with low lignin solubilities (e.g. alkanes) were gradually replaced by functional groups associated with higher lignin solubilities (e.g. aromatic N-atoms, Fig. 6.6 b) leading to a gradually increasing mean lignin solubility of the population (Fig. 6.6 c). All designed solvents with a lignin solubility of  $\log_{10}(x_L) > -1.5$  are summarised in Appendix C.2.

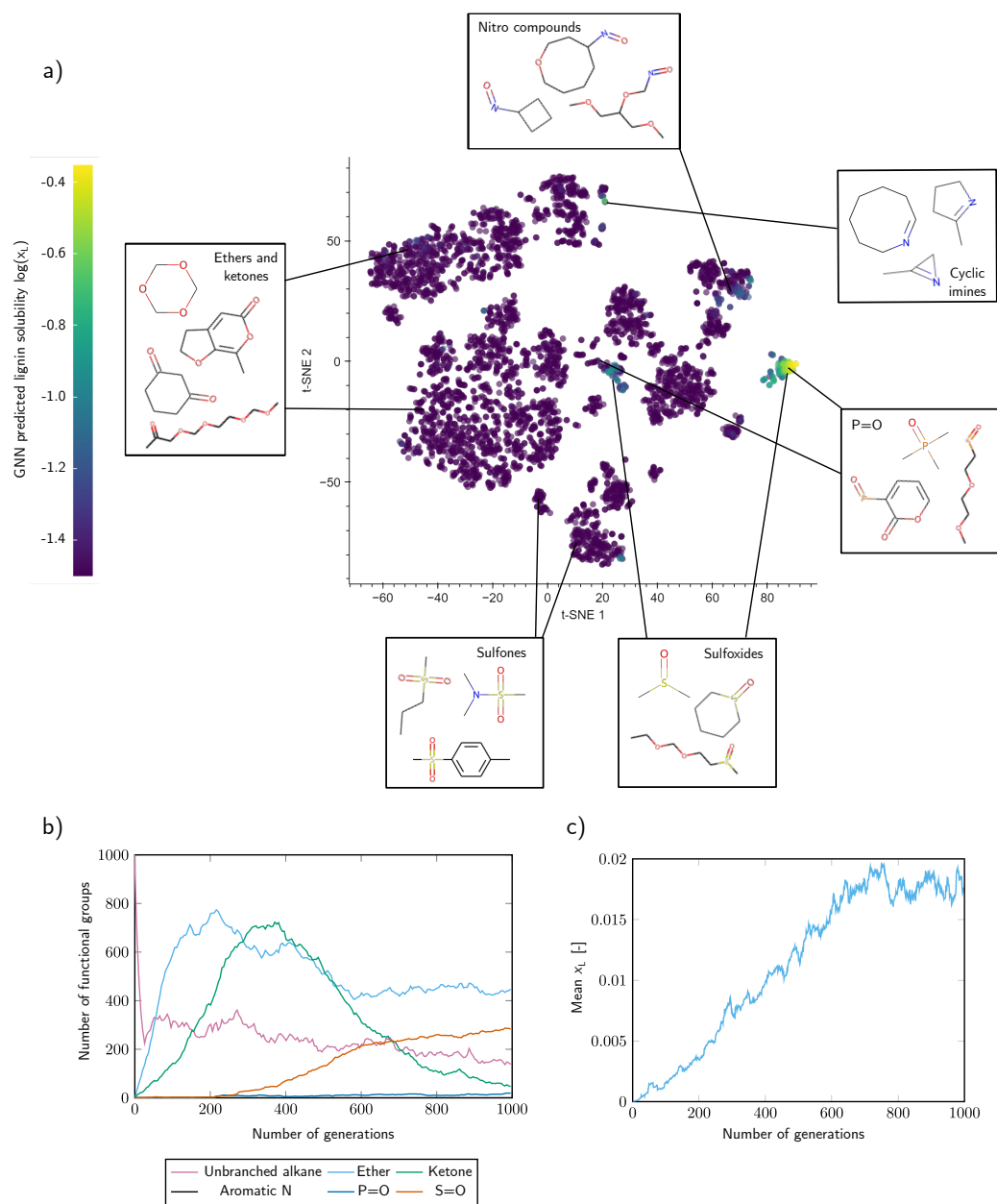


**Fig. 6.6** Application of the solvent design framework for lignin upgrading. a) T-SNE plot of the designed molecules with highlighted lignin solubility. b) Exploration of chemical space during molecule optimisation. c) Evolution of the molar lignin solubility during molecular optimisation. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

### 6.2.4 Solvent design for aldehyde-assisted fractionation

During the solvent design for AAF, the generated molecules were required to be non-reactive towards the process liquor and the biomass (Eq. 6.10). The algorithm generated sulfoxides, sulfones, ethers, ketones, phosphoryl compounds, and cyclic ethers and ketones as potential solvent candidates (Fig. 6.7 a). In addition, the functional group restrictions spurred the exploration of non-excluded nitrogen-containing patterns, such as nitro-groups or cyclic non-aromatic imines, with high GNN-predicted lignin solubilities. However, nitro-groups are explosive, and imines hydrolyse in aqueous, acidic environment, and ketones undergo aldol condensation when combined with aldehydes, rendering these compounds unsuitable for AAF.

Similar to the solvent design for lignin dissolution (Section 6.2.3), the number of alkane groups decreased rapidly within the first generations which were replaced by functional groups associated with higher lignin solubilities (Fig. 6.7 b). However, the search became more targeted, concentrating on solvent classes that are effective at dissolving lignin while taking into account the functional group restrictions. Among the 100 fittest solvent candidates, nearly 90% were sulfoxides, with DMSO being the overall fittest designed solvent ( $\log_{10}(x_L) = -0.35$ ). Due to the functional group constraints, the algorithm designed fewer solvents with high lignin solubilities compared to the run for lignin upgrading. As a consequence, the mean lignin solubility of the solvent population was lower (Fig. 6.7 c). A list of designed solvents is attached to Appendix C.2.



**Fig. 6.7** Application of the solvent design framework for AAF. a) T-SNE plot of the designed molecules with highlighted lignin solubility. b) Exploration of chemical space during molecule optimisation. c) Evolution of the molar lignin solubility during molecular optimisation. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

## 6.3 Experimental validation and investigating underlying structural patterns for lignin solubility

In the following sections, solvents identified by the solvent screening method and generated by the solvent design method are compared and selected for experimental investigations (Section 6.3.1). The selected solvents are used for solubility measurements for different types of lignin. The structural features of the selected solvents were analysed by applying attribution methods to the GNN, allowing to investigate structural patterns responsible for high lignin solubilities (Section 6.3.2). Finally, the most promising solvents were applied in AAF, demonstrating their potential for holistic utilisation of lignocellulosic biomass (Section 6.3.3).

### 6.3.1 Solvent selection

Both, the solvent screening and the solvent design method, identified sulfoxides, azines, oxazolines, and phosphonates as solvent candidates for effective lignin dissolution. However, the solvent design approach explored these solvent classes more in depth and explored a higher number of potential solvents in each class. Moreover, the solvent design approach identified highly promising solvent candidates that were not included in the database. These candidates mainly belonged to sulfones, azoles, morpholines, cyclic ethers, cyclic ketones, and further compounds containing phosphoryl groups. In the solvent design, linear ethers were predominant side-motifs.

From each class, solvent candidates were chosen for solubility measurements. The choice was guided by i) commercial availability ii) EHS properties as given in the SDS, and iii) price ( $\leq 100$  € per 10 g). Some of the structures designed by *PSEvolve* were not commercially available. In this case, solvents with similar chemical structures were chosen. As a control, the lignin solubility was measured in three control solvents: n-heptane, and dibutyl ether which have low lignin solubilities, and 2-MeTHF with mediocre lignin solubility. In total, 30 solvents were selected for lignin solubility measurements. Solvents for AAF were chosen based on the measured lignin solubility and reactivity towards the process liquor.

### 6.3.2 Lignin solubility measurements

Three different types of lignin were applied in the solubility measurements: Kraft lignin isolated from softwood species, FABIOLA<sup>TM</sup> organosolv lignin<sup>443</sup> isolated from hardwood, and mild acidolysis lignin (MAL) isolated from corn cob (see Appendix C.3.1 for 2D HSQC NMR of the lignocellulose samples and additional data points with lignin isolated from birchwood). The solubility measurements were performed according to Section 6.4.2 at

$T = 85\text{ }^{\circ}\text{C}$  (process temperature of AAF).

The measured lignin solubilities ranged between 20 - 60 wt.% in most of the computationally identified solvents (Fig. 6.8 a). The highest solubilities were measured for DMSO ( $\geq 60$  wt.%), and isoxazole ( $\geq 50$  wt.%), 2-picoline-n-oxide ( $\geq 49$  wt.%), 2,5-dimethylpyrazine ( $\geq 49$  wt.%), and thiazole ( $\geq 49$  wt.%). Note that for most solvents, the lignin saturation was not completely reached as the solutions became increasingly viscous with higher amounts of dissolved lignin. The high viscosity was challenging during the filtration step even with specialised filters designed for viscous samples. Deviations between predictions and experiments were observed for some of the selected ethers. Lignin solubilities of up to 51 wt.% were measured for diethylene glycol dimethyl ether (DEGDME) whereas the structurally similar diethylene glycol diethyl ether (DEGDDE) dissolved maximally 8.9 wt.% of lignin. The GNN was not able to discriminate the differences between the ether structures and predicted for both solvents nearly no lignin dissolution. Also COSMO-RS predictions could not resolve the difference (data not shown), suggesting that these deviations were not caused by the GNN training, but were more likely caused by the training data obtained from the COSMO-RS solubility predictions. In general, the experiments confirmed high solubilities for the computationally identified solvents. There were no remarkable differences in solubility between the lignin types, implying that the influence of the lignin source and extraction method was not decisive for the measured lignin solubilities. Furthermore, these results show that the selected GGG-trimer was representing all types of lignin quite well. The numerical solubility data is provided in Appendix C.3.2.

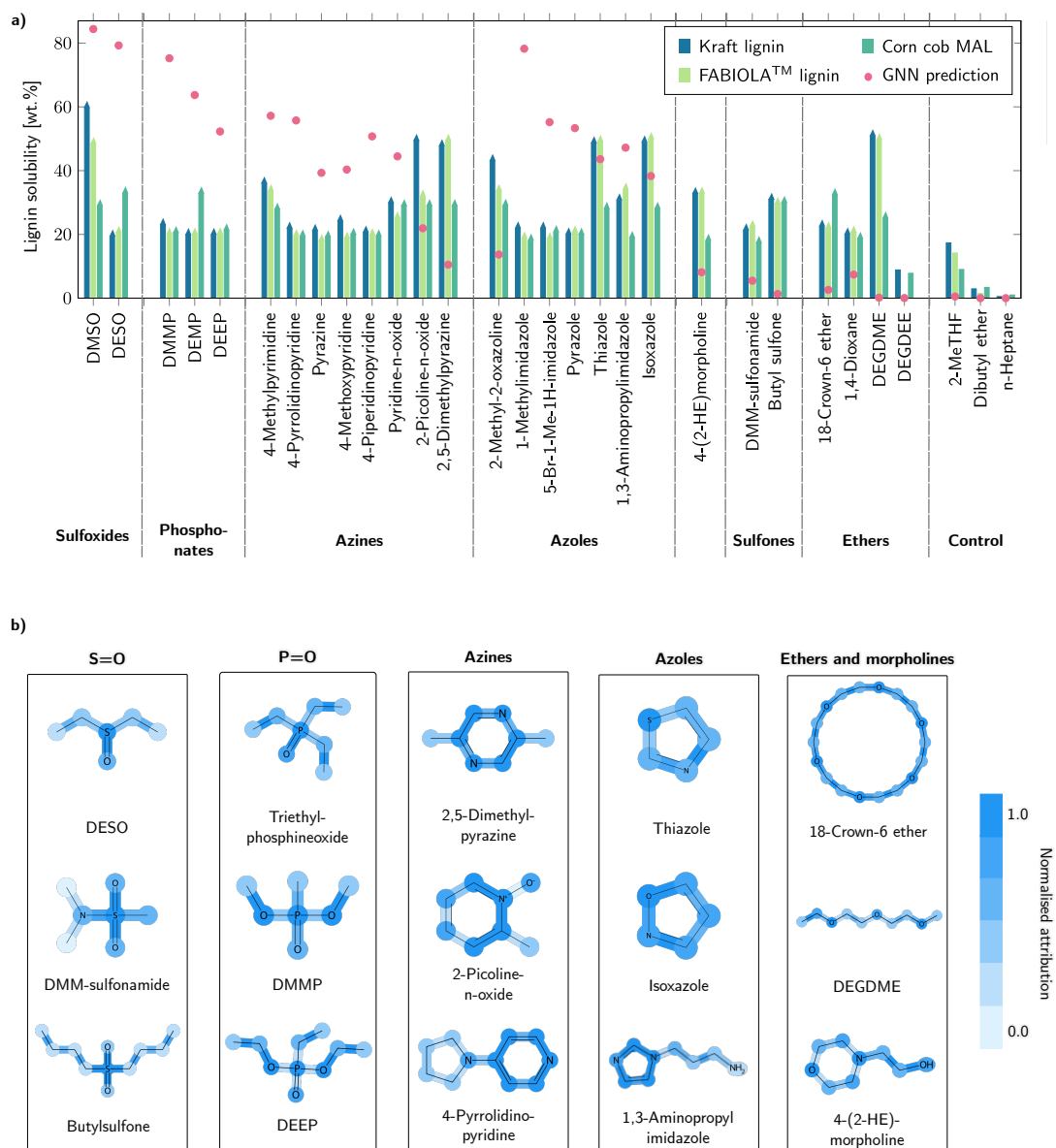
### 6.3.2.1 Analysing structural patterns using attribution methods on the GNN

To analyse the underlying structural patterns in the solvents with respect to the GNN predicted solubility values, the GNN was coupled with attribution techniques. Attribution techniques, such as the integrated gradient (IG) approach<sup>380,444</sup> were specifically developed to comply with the sensitivity and implementation invariance axioms. In the IG method, the integral of the gradients of the model's output with respect to its input is computed, while gradually changing the input values from a baseline to the actual input of interest (see Appendix C.2.1.4 for more details). This process effectively assigns importance scores to each input feature by attributing their contribution to the final prediction. The IG method identifies nodes and edges with the highest impact on a given prediction. Therefore, the contribution of each atom and bond within the solvent to the predicted lignin solubility can be visualised and potentially used as a guide in the explainability procedure. The IG method was employed to attribute the predicted lignin solubility to structural features of the solvent. To enable reliable interpretations of the results beyond theoretical predictions, the experimentally measured lignin solubilities served as a ground truth when analysing the

attributions.

First, the normalised attributions indicating the importance of nodes and bonds for the solubility predictions (Fig. 6.8 b) were computed. Higher attribution scores indicate a higher importance on the predictions. Note that the attribution scores were normalised, and therefore only allowed for a relative comparison within the same molecule. The high lignin solubility in sulfoxides and sulfones was mainly attributed to the presence of S-atoms and the adjacent double bonds in solvent structures. Similarly, the P=O motif in phosphine oxides was the most influential in promoting lignin solubility. Additionally, in phosphonates, P- and O-atoms received the highest attribution scores. The high lignin solubility in aromatic heterocycles could in general be attributed to the N-atoms and the neighbouring aromatic bonds. However, for specific classes like oxazoles and thiazoles, the S- and O-atoms within the aromatic rings were predicted to be more influential than the aromatic N. In contrast, N-atoms within side chains as well as alkyl chains and saturated rings found in compounds such as butyl sulfone, 4-pyrrolidinopyridine, and 1,3-aminopropyl imidazole, were predicted to have less impact on the solubility predictions. In the case of ethers, the O-atoms predicted to be more influential than the C-C bonds. In general, solvents containing S, N, P, O or aromatic bonds were associated with high solubility predictions and the attribution results were in line with the experiments. Indeed, lignin with numerous benzene rings and oxygen atoms potentially engages in  $\pi$ -interactions with solvent molecules and forms hydrogen bonds with these heteroatoms. However, as solubility predictions and experiments diverge for linear ethers, there likely exist additional structural features affecting solubility that could not be captured by the GNN. In line with the presented results, the computational Kamlet-Taft-parameter analysis of Sumer and van Lehn revealed that a solvent requires good hydrogen bond-accepting ability and intermediate to high polarity to effectively dissolve lignin.<sup>445</sup>





**Fig. 6.8** Experimental validation of the GNN predictions and attribution of structural features to the predicted lignin solubilities. a) Experimental lignin solubilities in the designed solvents for Kraft lignin, FABIOLA™ lignin, and MAL isolated from corn cob. Arrows indicate that lignin saturation was not yet reached, however, the high viscosity of the solution hindered measurements with higher lignin loadings. b) Normalised attributions for each discovered solvent class. A higher attribution score of the highlighted structural feature indicates higher importance for the lignin solubility prediction. Abbreviations: Dimethylsulfoxide (DMSO), diethylsulfoxide (DMSO), n,n-dimethylmethanesulfonamide (DMM-sulfonamide), dimethyl methylphosphonate (DMMP), diethyl methylphosphonate (DEMP), diethyl ethylphosphonate (DEEP), 5-bromo-1-methyl-1H-imidazole (5-Br-1-Me-1H-imidazole), 4-(2-hydroxyethyl)morpholine (4-(2-HE)morpholine), diethylene glycol dimethyl ether (DEGDME), diethylene glycol diethyl ether (DEGDEE), 2-methyltetrahydrofuran (2-MeTHF). Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

### 6.3.3 Propionaldehyde-assisted pretreatment

Based on the solubility measurements, the most promising solvent candidates were selected for analysing their performance in AAF. Due to their high solubilities and their potential stability under AAF conditions, sulfones, sulfoxides, phosphonates, cyclic ethers, and linear ethers were selected as solvent candidates. For AAF experiments, the selected solvent candidates, propionaldehyde, and an aqueous HCl-solution were added to birch wood chips (see Appendix C.3.3 for compositional analysis, and Appendix C.3.1 for 2D-HSQC NMR analysis). The performance of the selected solvents was assessed by i) observing potential reactivity between the solvent and other compounds in the reaction liquor (warming, fuming), ii) the yield of the pulp, iii) the yield of propionaldehyde-protected xylose dipropylxylose (DPX), iv) measuring the yield of lignin monomers after hydrogenolysis of the pretreatment liquor, and v) observing the appearance of the biomass after the pretreatment. These metrics can serve as indicators of the effectiveness of biomass depolymerisation and stability of the solvents.

The tested sulfoxides, sulfones (except the DMM-sulfonamide), phosphonates, and ethers were resistant to acidic conditions (0.4 M HCl) and elevated temperature ( $T = 85\text{ }^{\circ}\text{C}$ ), showing no visible signs of degradation. Upon completion of the pretreatment reaction, the biomass was disrupted to a powder-like residue with light colour in samples treated with butyl sulfone, DEGDEE, and DMM-sulfonamide (Fig. 6.9 a), constituting around 40 wt.% of the biomass. In contrast, in the samples containing DEGDME, 18-crown-6 ether, DMSO, and phosphonates, the biomass retained the form of wood chips. For these solvents, the filtrated pulp constituted up to 90 wt.% of the biomass, suggesting that the extraction of hemicellulose and lignin was not effective and that these components remained within the cellulose fibers. This observation is supported by the low DPX and lignin monomer yields for these solvents (Fig. 6.9 b and c). To disintegrate the cellulose sheets, the reaction liquor must be able to disrupt the hydrogen bonding between the cellulose chains, and furthermore, interfere with the hydrophobic stacking interactions between the nonpolar regions of cellulose sheets.<sup>446</sup> Presumably, these solvents were too polar to attack the hydrophobic domains. Surprisingly, despite having a higher lignin solubility, the DEGDME pretreatment resulted in a lower lignin monomer yield than DEGDEE, presumably due to more effective interactions between DEGDEE and the carbohydrates. This observation highlights that lignin isolation from biomass is not only influenced by the lignin solubility but also depends on the interactions between the carbohydrates and the pretreatment liquor. A lower monomer yield for DEGDME due to ineffective lignin stabilisation can be ruled out, since its 2D HSQC NMR spectrum (Appendix C.3.1) is not indicative of lignin condensation.

After separating the pulp by filtration, the resulting filtrate contained the extracted lignin with propionaldehyde-protected  $\beta$ -O-4 linkages as confirmed by the HSQC NMR

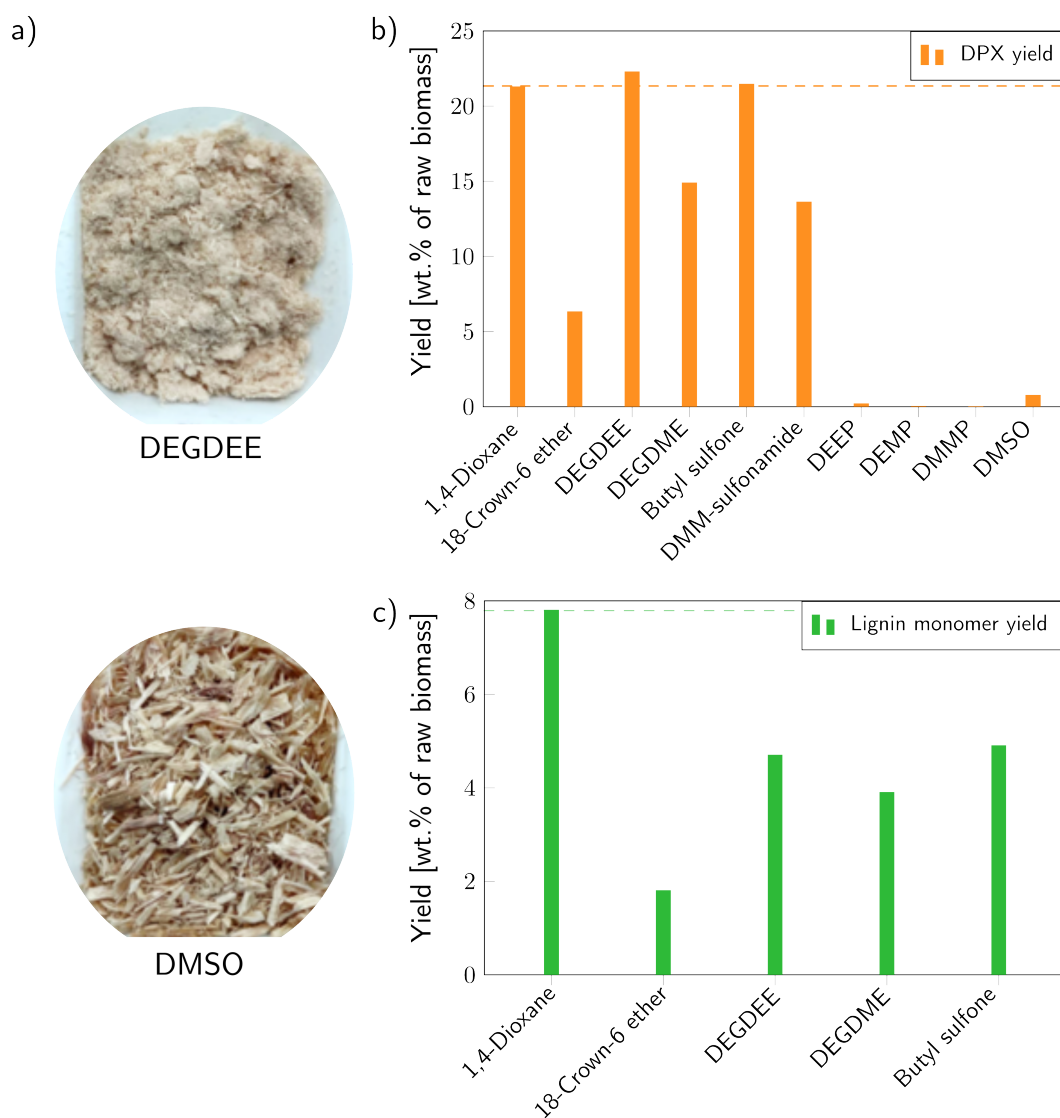
(Appendix C.3.1). Hydrogenolysis of such uncondensed lignin over a Ru/C catalyst at  $T = 250\text{ }^{\circ}\text{C}$  produces valuable aromatic lignin monomers. The quantification of the lignin monomers provides a measure for the effectiveness of lignin extraction and stabilisation. The benchmark solvent 1,4-dioxane allowed to produce near-theoretical 7.8 wt.% of monomers on a raw biomass basis, followed by DEGDEE and butyl sulfone, each providing ca. 5 wt.% of monomers, and DEGDME and 18-crown-6 ether, yielding less than 4 wt.% (numerical values of DPX and lignin monomer yields are given in Appendix C.3.4).

The yield of DPX exceeded 20 wt.% on raw biomass basis in the benchmark solvent 1,4-dioxane, and the designed solvents DEGDEE and butyl sulfone. The DPX yield for the 18-crown-6 ether, DEGDME and DMM-sulfonamide was lower than 15 wt.%. DMM-sulfonamide demonstrated signs of degradation during the pretreatment, while the other two solvents did not provide sufficient biomass disruption as mentioned above. Low amounts of DPX ( $< 1\text{ wt.}\%$ ) were produced during the pretreatment in phosphonates and sulfoxides. Interestingly, in control experiments with dibutyl ether and heptane, DPX yields comparable to 1,4-dioxane were obtained, suggesting that the interactions of non-polar solvents and the biomass enable sufficient contact between the reaction components (e.g. aldehyde) and xylan that closely interacts with cellulose in a plant cell. However, the lignin solubility in these solvents is low. Lignin forms a globule with reduced surface area in such highly nonpolar solvents<sup>442</sup>, potentially preventing aldehyde-protection, thus resulting in lignin condensation.

The pretreatment experiments showed that the designed glycol ethers, and sulfones are promising solvent candidates for the AAF procedure. Notably, DEGDEE and butyl sulfone provided effective fractionation of cellulose pulp, and successful extraction of PA-protected lignin and xylose in the pretreatment liquor with yields comparable to the carcinogenic benchmark solvent 1,4-dioxane. The solvent butyl sulfone is high boiling and solid at room temperature, and therefore requires adaptations of current procedures for the isolation of fractions from the liquor and for solvent recovery. Additionally, this solvent is currently produced in limited quantities and its toxicological profile is not well studied, opening opportunities for future research.

The solvents tested in AAF were specifically designed to target the lignin fraction of the biomass. Therefore, most candidates are rather polar since this characteristic is necessary to ensure lignin solubilisation in the liquor, thereby increasing its surface area exposed for a reaction with aldehyde in the mixture.<sup>442</sup> However, the overall quality and quantity of the extracted lignin were additionally influenced by other solvent properties. In addition to polar groups, the solvent requires sufficient non-polar domains that can disrupt the cellulose

microfibrils and facilitate the extraction of lignin and xylan which is in line with the experimental observations. Pretreatments with butyl sulfone which consists of both, a highly polar  $\text{SO}_2$  group and two aliphatic chains, resulted in high DPX and lignin monomer yield. Furthermore, the performance of DEGDEE in AAF experiments was remarkably better than that of the structurally similar solvent DEGDME which only differs from DEGDEE by the absence of the  $\text{CH}_3$ -end groups. Additionally, the stability of the solvent under acidic conditions at high temperatures provides a consistent and effective chemical environment throughout the process which is required for efficient extraction of hemicellulose sugars and lignin from the biomass, and for the protection reactions with the aldehyde.



**Fig. 6.9** Results of propionaldehyde-assisted pretreatment of birch wood ( $T = 85\text{ }^{\circ}\text{C}$ ,  $t = 3\text{ h}$ ). a) Cellulose-rich pulp after pretreatment with DEGDEE (above) and DMSO (below), b) DPX yield on raw biomass basis, c) lignin monomer yield. The yields are provided on a raw biomass basis. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier.

## 6.4 Experimental methods

### 6.4.1 Materials

#### 6.4.1.1 Solvents

The following solvents were purchased to measure the lignin solubilities: dimethyl sulfoxide (Sigma-Aldrich, > 99.5%), diethyl sulfoxide (Fluorochem, 98%), dimethyl methylphosphonate (STREM Chemicals, 97%), diethyl methylphosphonate (ACROS Organics, 96%), diethyl ethylphosphonate (Sigma-Aldrich, > 98%), 4-methylpyrimidine (ABCR, 98%), 4-pyrrolidinopyridine (ABCR, 95%), pyrazine (ACROS Organics, 99%), 4-methoxypyridine (ABCR, 97%), 4-piperidinopyridine (ABCR, 97%), 2-methyl-2-oxazoline (ABCR, 99%), 1-methylimidazole (Sigma-Aldrich, > 99%), 5-Bromo-1-methyl-1H-imidazole (ABCR, 95%), pyrazole (ThermoScientific, > 98%), n,n-dimethylmethanesulfonamide (ThermoScientific, > 98%), 18-crown-6 ether (Sigma-Aldrich, for synthesis), 1,4-dioxane (Carl Roth, > 99.5%), 2-methyltetrahydrofuran (Carl Roth, > 99%), dibutyl ether (ACROS Organics, > 99%), n-heptane (ABCR, > 96%), pyridine-n-oxide (Thermo Scientific Acros, 95%), 2-methylpyridine-n-oxide (Fluorochem, 98%), 2,5-dimethylpyrazine (Fluorochem, 98%), thiazole (Fluorochem, 99%), 1,3-aminopropyl imidazole (Acros Organics, 98%), isoxazole (Fluorochem, 98%), 4-(2-hydroxyethyl)morpholine (Acros Organics, 99%), butyl sulfone (Sigma-Aldrich, 99%), diethylene glycol dimethyl ether (Alfa Aesar, 99% stab. with 100 ppm BHT), diethylene glycol diethyl ether (Alfa Aesar, 99%).

#### 6.4.1.2 Preparation of lignin samples

The lignin samples were chosen to represent a broad diversity of feedstocks (hardwood, softwood, herbaceous biomass) and isolation methods (Kraft, FABIOLA™, mild acidolysis). Lignin from Rettenmaier beechwood (hardwood) was obtained from TNO (Netherlands) and was originally gained from the FABIOLA™ acetone organosolv process.<sup>443</sup> Kraft lignin was provided by Berner Fachhochschule and was originally isolated from softwood species by the Kraft process. Lignin from corn cobs and from birch wood were isolated by mild acidolysis.<sup>447</sup> Birch wood (*Betula pendula*, hardwood) was procured from M. Studer of the Bern University of Applied Sciences. The birch wood chips were sorted to remove residual bark and leaves. Corn cobs (herbaceous biomass) were procured from IP-Suisse in Lausanne, Switzerland. The corn cobs were sorted to remove residual leaves, stems, and corn. The sorted birch wood chips and the corn cobs were milled using a 6-mm screen and sieved with a 0.45-mm mesh. The fraction with a size of < 0.45 mm was used for lignin extraction by mild acidolysis.

### 6.4.1.3 Lignin extraction by mild acidolysis

Lignin from birch wood and corn cobs was isolated by mild acidolysis. The procedure for mild acidolysis was taken from Das et al.<sup>447</sup> with slight modifications. The mild acidolysis lignin extraction method allows to preserve the native structure of lignin to a large extent while extraction yields are small.<sup>447</sup> In brief, 10 g of ground biomass was dissolved in 120 ml of a dioxane/water mixture (9/1, v/v) containing 0.2 mol l<sup>-1</sup> HCl. The suspension was heated to 90-100 °C to reach refluxing and was stirred for 30 min for herbaceous biomass and for 45 min for hardwood. The cooled mixture was vacuum-filtered through a Whatman filter (paper grade 3). The residue was washed three times with 50 ml of the dioxane/water mixture (9/1, v/v). The pH of the resulting solution was adjusted to 3-4 using a saturated aqueous NaOH solution. Subsequently, the solution was concentrated to 50 ml by rotary-evaporation (45 °C) before any solid lignin residues appeared. The concentrated solution was added into a large volume of cold water (500 ml) to precipitate lignin. The precipitated lignin was washed with 100 ml of deionized water and dried in a desiccator. The 2D HSQC NMR spectra of the isolated lignins were measured to confirm lack of significant condensation (see Appendix C.3.1).

### 6.4.2 Lignin solubility measurements by gas chromatography

Commonly applied methods for lignin solubility measurements are based on solvent evaporation and gravimetric determination of the dissolved lignin. Due to the high  $T_b$  of several selected solvents, the lignin solubility was measured by gas chromatography (GC). First, lignin solvent suspensions were prepared. Here, dried lignin (0.1 g) was added to the solvent (0.4 g) in a glass vial with a magnetic stirring bar. The vials were sealed and placed in a pre-heated aluminum block holder ( $T = 85$  °C) under constant stirring at 400 rpm. The samples were kept under agitation until reaching equilibrium (2 h). Subsequently, the samples were filtered using a 1 ml syringe with an attached polytetrafluoroethylene filter (0.22  $\mu$ m pore size) to remove undissolved solid lignin. The saturated liquid phase (ca. 0.1 g) was diluted with dimethyl sulfoxide or acetone (ca. 1.5 g), and 1,3-dioxolane was added to the sample (ca. 0.1 g) as an internal standard. The samples were quantified by GC with flame-ionisation detection (FID). Calibration curves for each solvent with the internal standard were obtained using known amounts of solvent and 1,3-dioxolane dissolved in DMSO or acetone (if the peaks of DMSO and solvent were overlapping). All solubility tests were performed at least in duplicate. The lignin solubility was calculated using the following equation

$$\text{lignin solubility [wt.\%]} = \frac{m_L}{m_L + m_s} \cdot 100. \quad (6.11)$$

By measuring the mass of solvent  $m_s$  and the mass of the filtrate  $m_{\text{filtrate}}$ , the mass of lignin  $m_l$  was obtained using the following equations

$$m_s = \frac{RF \cdot \text{area}_s \cdot m_{\text{IS}}}{\text{area}_{\text{IS}}}, \quad (6.12)$$

$$m_L = m_{\text{filtrate}} - m_s, \quad (6.13)$$

where  $RF$  denotes the response factor and the subscript IS refers to the internal standard. A GC-FID system by Agilent Technologies (model no. 7890B) equipped with an HP-5 column was used for quantification. The injection temperature was 300 °C. The column temperature program was: 40 °C (3 min), 30 °C min<sup>-1</sup> to 100 °C, 40 °C min<sup>-1</sup> to 300 °C and 300 °C (5 min). The detection temperature was 300 °C. Response factors  $RF$  are given in Appendix C.3.2.

### 6.4.3 Lignin solubility measurements by evaporation

To validate the results of the GC-method, solubility measurements using the traditional gravimetric evaporation method for solvents with low  $T_b$ . For this purpose, the procedure from Dick et al.<sup>448</sup> was used. Here, after preparing a suspension of lignin and the solvent as described above, the filtered solution was dried to remove the solvent, and the mass of dissolved lignin was quantified. The vials containing the filtered solution were placed in a vacuum oven at  $T = 45$  °C and  $p = 20$  mbar, and dried overnight. Subsequently, the vials were re-tared to determine the mass of lignin. The solubility was determined using the following equation and final results were based on the average of two samples

$$\text{lignin solubility [wt.\%]} = \frac{m_{\text{vial with dry lignin}} - m_{\text{vial}}}{m_{\text{vial with lignin solution}} - m_{\text{vial}} - m_{\text{dry lignin}}}. \quad (6.14)$$

This procedure has been applied to six solvents with  $T_b < 150$  °C allowing for evaporation under the described conditions (data attached in Appendix C.3.2).

### 6.4.4 Propionaldehyde-assisted pretreatment

The propionaldehyde-assisted pretreatment of birch wood in the selected solvents was performed as described in detail in Talebi Amiri et al.<sup>155</sup> Birch wood (*Betula pendula*) was prepared as described in Section 6.4.1.2. Briefly, 4.5 g of milled birch wood, 4.8 ml of propionaldehyde, 0.85 ml of an aqueous HCl solution (37%, w/w), and 25 ml of solvent were added in a thick-walled glass reactor equipped with a stirring bar. The reactor was placed in an oil bath heated to 85 °C and the reaction proceeded for 3 h while stirring at 600 rpm. The reaction was cooled to room temperature and 20 ml of 1,4-dioxane was



added to the mixture to dilute any solidified solvent.

0.5 ml aliquot of the reaction liquor was taken, diluted in DMSO and analysed by GC-FID (Agilent Technologies 7890B) equipped with a HP-5 Column (Agilent) to determine DPX yield using a calibration curve obtained from authentic standards. The yield is provided on a raw biomass basis (non-dried, non-extracted birch wood) accounting for the weight of DPX derived from propionaldehyde. The reaction liquor obtained by AAF was filtered through a Nylon filter of 0.8  $\mu\text{m}$  to separate cellulose-rich pulp. The pulp was subsequently dried in a vacuum desiccator for 48 h and then weighed.

To quantify the yield of lignin monomers, first, the filtrate was neutralised by gradually adding 0.86 g of sodium bicarbonate until pH 6-7. The solution was diluted to 100 ml with 1,4-dioxane in a volumetric flask and centrifuged to remove precipitated salt. 20 ml of the 1,4-dioxane/lignin solution was taken for hydrogenolysis in a 50-ml Parr reactor (stainless steel) equipped with a magnetic stirrer and a K-type thermocouple. 100 mg of a 5% Ru/C catalyst was added to the solution and the reactor was pressurised to 40 bar of  $\text{H}_2$ . The reactor was heated to 250  $^\circ\text{C}$  for 3 h. Subsequently, the reactor was cooled to room temperature, depressurised, and the solution was filtered with 0.2  $\mu\text{m}$  PTFE syringe filter to remove the catalyst. 200  $\mu\text{l}$  of an internal standard solution of decane ( $0.05 \text{ g ml}^{-1}$ ) was added to the filtered solution and 1 ml sample was analysed by GC-FID to determine monomer yield. The quantification of monomers was performed using the Effective Carbon Number method described by Talebi Amiri et al.<sup>155</sup>

## 6.5 Summary and conclusions

In current approaches for lignocellulose fractionation, commonly used solvents either have a low to mediocre lignin solubility (e.g. EtOH or acetone) or have harmful EHS properties (e.g. 1,4-dioxane). To facilitate solvent selection, the developed computational solvent screening and design frameworks were applied to lignocellulose processing. Finally, the identified solvents were experimentally validated.

In the computational solvent screening, a database containing more than 8000 compounds was screened to identify solvent candidates for the dissolution of multiple lignocellulose fractions, and for selective lignin dissolution. After excluding unsuitable solvent candidates based on structural features and  $T_m/T_b$  limits, the solubilities of cellulose-, hemicellulose-, and lignin-representing structures were predicted using COSMO-RS. The EHS properties of the solvent candidates were predicted using VEGA models and were combined into the EHS score, expressing the benignity of a solvent's EHS properties on a scale from 0 (harmful) to 1 (benign). Based on the predicted solubilities and the EHS score,

the solvents were ranked according to three objectives: 1) joint dissolution of all biomass fractions and EHS score, 2) joint lignin and cellulose solubility, and 3) selective lignin solubility. For objectives 1 and 2, mainly ILs and 18-crown-6 ether were identified potential solvents. Literature data has shown that some of the identified ILs are able to completely dissolve wood chips, highlighting the applicability of the solvent screening approach. The identified ILs are interesting for applications involving cellulose dissolution. For objective 3, azines, sulfoxides, oxazolines, and phosphonates were predicted as promising solvent candidates for selective lignin dissolution. These solvents have high potential for application in dissolution-based lignin upgrading, such as in the fabrication of lignin nanoparticles, films, or fibres. However, some of the solvent classes are unstable in acidic environments or are reactive towards aldehydes (e.g. oxazolines and azines), limiting their application in organosolv pulping or AAF.

To explore a broader chemical space beyond the predefined database, the solvent design algorithm *PSEvolve* was applied. *PSEvolve* allowed for seamless integration of a GNN for lignin solubility predictions. Since COSMO-RS solubility predictions were not applicable due to their need for time-consuming quantum chemical calculations, the GNN was applied as a surrogate model for COSMO-RS. The GNN was trained and tested on a COSMO-RS solubility predictions of more than 3300 compounds. The use of a GNN enabled fast and reliable lignin solubility predictions enabling the design of tailored solvent structures. However, deviations between experimentally measured and GNN-predicted solubilities were observed for linear ethers. Beyond sulfoxides, azines, oxazolines, and phosphantes which were already identified in the solvent screening, promising sulfones, azoles, morpholines, cyclic ethers and ketones, and compounds containing phosphoryl groups were designed. Beyond the solvent screening method, *PSEvolve* explored promising solvent classes with more molecular detail.

From each solvent class, commercially available solvent candidates were evaluated with respect to their EHS properties and their price. 30 solvents were selected for lignin solubility measurements. Both, the solvent screening and the design approach, were able to identify non-intuitive solvent candidates with lignin solubilities between 20 – 60 wt.%. Particularly high solubilities were measured in DMSO, pyrazines, azoles, and DEGDM. These solvents have high potential for application in the fabrication of lignin-based films, nanoparticles, or resins. A common method to demonstrate the applicability of molecular design algorithm is the rediscovery target molecules with defined properties.<sup>449</sup> DMSO and pyridine were recently reported as the most effective solvents for lignin dissolution.<sup>100,101</sup> Both compounds were rediscovered by the presented solvent screening and design methods.

By applying an attribution method to the GNN, the importance of structural features of the experimentally evaluated solvent candidates with respect to the GNN predicted lignin solubilities was analysed. In line with experimental results, solvents containing sulfinyl, sulfonyl or phosphoryl groups were connected to high lignin solubilities. Furthermore, aromatic nitrogen and neighboring aromatic bonds, and the presence of O-atoms were linked to high lignin solubilities. Therefore, the presented solvent design framework not only facilitated the exploration of promising solvents but also provided valuable insights into the specific molecular characteristics essential for achieving high lignin solubility. Furthermore, the presented solvent design approach could serve as a blueprint for other types of solutes, e.g. for cellulose for which only few effective solvents were reported.

Solvent selection for AAF is a complex task. In addition to effectively dissolving lignin, a suitable solvent for AAF must remain unreactive towards the process liquor. Additionally, benign EHS properties are imperative for sustainable biorefinery processes, however, immensely narrow down the feasible molecular search space. The solvent screening showed, that out of 8000 potential solvents contained in the database only few seemed applicable in AAF. The solvent design framework generated structures tailored towards AAF, including sulfones, sulfoxides, phosphonates, and cyclic and linear ethers. Solvents from these groups showed acid stability under AAF conditions and provided performance nearly on par with 1,4-dioxane in terms of DPX yield. Slightly lower lignin monomer yields were obtained compared to 1,4-dioxane. However, a major advantage of the tested solvents are their preferable EHS properties. Especially DEGDEE, DEGDME, and butyl sulfone are interesting solvent candidates for AAF. Furthermore, the experiments revealed that solvents with a higher lignin solubility do not necessarily lead to more effective lignin isolation from biomass. In the native biomass, lignin is embedded in the lignin-carbohydrate-complex where the biomass fractions are connected *via* various chemical bonds. Especially solvents with low or mediocre DPX yields (e.g. 18-crown-6-ether or DEGDME), suggesting insufficient interaction with the carbohydrates, led to lower lignin monomer yields, despite their high lignin solubilities. Furthermore, the AAF experiments suggest that stable aprotic solvents possessing heteroatoms (e.g. O, N, S), and nonpolar domains (e.g. hydrocarbon motifs) could be suitable for application in AAF and in lignocellulose processing in general. The challenge for finding the right balance of these characteristics remains open for future research.

The newly identified solvents present alternatives to state-of-the-art solvents applied in organosolv and AAF processing, as well as for dissolution-based lignin upgrading. However, most of the identified chemicals are not frequently employed or have very specific practical application in industry or laboratory, and are not commonly used as solvents. Before

these solvents can be applied in such processes, further research is required to establish procedures for solvent recovery, recycling, cost, and product separation. Furthermore, a life cycle assessment is necessary to assess the sustainability and the resulting global warming potential depending on the selected solvent. However, solvent-specific data relevant for life cycle assessments is currently lacking.

In general, computational modelling provided a more systematic, high-throughput approach to solvent selection compared to currently established methods, e.g. experimental solvent selection guided by Hansen solubility parameters. Both methods, the solvent screening and the solvent design approach, represent a major leap in solvent selection for dissolution-based lignin upgrading and lignocellulose processing.

## 7 | Solvent screening for the fractionation of wet microalgal biomass

Microalgae are an attractive feedstock for producing food, feed, chemicals, fuels, pigments, and other high-value products in biorefineries.<sup>163,164,450–452</sup> However, with current technologies, obtaining clean streams of proteins, carbohydrates, lipids, and pigments remains challenging (see Section 3.2.6.3 for further details) and are not yet adopted for application on the industrial scale. Lab-scale fractionation methods rely heavily on the use of toxic<sup>281,284</sup> or expensive solvents,<sup>288–290</sup> or require costly equipment.<sup>180,276,283</sup> Moreover, the energy-intensive biomass drying step was identified as the bottleneck of the downstream process that limits the economic feasibility of the overall process.<sup>226</sup> To eliminate the energy demand of the drying step, recent research efforts investigated wet algal paste as obtained after harvest - still having a moisture content of about 85 wt.% - as a feedstock for biorefineries.<sup>226,247,248,261</sup> However, solvents that were commonly applied to extract lipophilic compounds from dried microalgae (e.g. alkanes), resulted in remarkably lower yields when applied to wet, untreated biomass,<sup>247,248</sup> aggravating the difficulties for biorefining. Various cell disruption methods were explored as a way to facilitate the contact between the solvent and the target compounds, thus increasing the yield.<sup>247,261,453,454</sup> In addition to cell disruption methods, systematic solvent selection has the potential to not only increase the yield of target compounds but also to develop innovative separation strategies required in biorefineries. Solvents can separate the biomass fractions by i) either selectively extracting the target compounds in a sequence of extractions, or ii) by extracting multiple target compounds simultaneously and subsequently separating the mixture by further separation steps, such as by LLE- or SLE-based methods. To accurately consider these thermodynamic aspects, the high moisture content of the biomass must be taken into account. Additionally, EHS properties represent important selection criteria. In the following sections, the developed solvent screening approach (Section 5.2) is leveraged to identify benign solvents enabling the fractionation of wet microalgal biomass. *P. tricornutum* (Section 3.1) was selected as a model feedstock. In contrast to lignocellulose, no industrially established biorefinery process for *P. tricornutum* exists to date. Therefore, potential biorefinery products of a *P. tricornutum*-based biorefinery and the most important

separation steps are determined (Section 7.1). Subsequently, the solvent screening approach is applied to identify solvent candidates for biomass fractionation (Section 7.2). The selected solvents are subsequently assessed in extraction experiments (Section 7.3). Based on the computational and experimental results, a lab-scale biomass fractionation process is developed for wet *P. tricornutum* biomass (Section 7.4). All experimental methods describing the cultivation, harvest, and extraction procedure, and further analysis of the extracts are provided in Section 7.5.

*This chapter is based on and includes direct as well as modified excerpts from the publications of König-Mattern et al.<sup>394</sup> and König-Mattern et al.<sup>455</sup> used under Creative Commons license CC BY 4.0.*

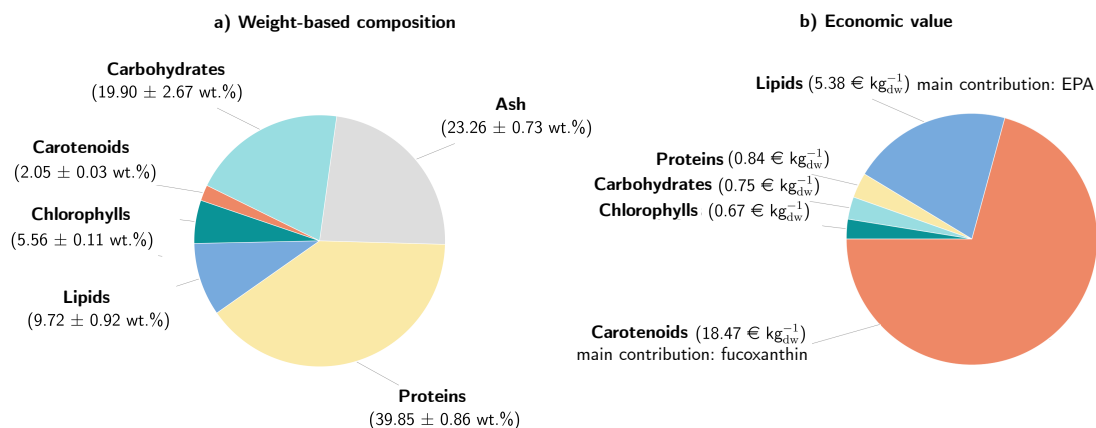
## 7.1 Biomass composition and potential biorefinery products of *P. tricornutum*

*P. tricornutum* was cultivated in a flat-panel photobioreactor (Section 7.5.1) under artificial light irradiation and was harvested at a cell density of  $4.81 \text{ g l}^{-1}$ . Experimental analysis of the biomass composition (Sections 7.5.2-7.5.10) quantified 39.9 wt.%<sub>dry</sub> proteins, 19.9 wt.%<sub>dry</sub> carbohydrates, 9.7 wt.%<sub>dry</sub> lipids, 5.5 wt.%<sub>dry</sub> chlorophylls, and 2.1 wt.%<sub>dry</sub> carotenoids (7.1). A detailed composition of fatty acids, chlorophylls, and carotenoids is summarised in Appendix D.3.1. For each biomass fraction, potential end-products and their economic value were reviewed. To analyse the economic potential of each biomass fraction, the corresponding products were allocated to the most profitable application (see Appendix D.5). In line with other studies, the red pigment Fx was identified as the most abundant carotenoid (1.5 wt.%<sub>dry</sub>). Carotenoids can be marketed in the health industry or as nutritional supplements due to their antioxidant properties at prices of  $900 \text{ € kg}_{\text{prod}}^{-1}$ .<sup>165</sup> Despite their low abundance on a weight basis (Fig. 7.1 a), the carotenoids comprise about 70% of the overall economic biomass value, highlighting their importance for the economic viability of the overall biorefinery (Fig. 7.1 b). Furthermore, the chlorophylls can be applied as natural colourants (e.g. as E140 or E141).<sup>456</sup> However, their economic importance for the biorefinery is low compared to that of the carotenoids. The lipid fraction contributes to around 20% of the overall economic biomass value and spans a diverse range of potential applications. EPA (2.6 wt.%<sub>dry</sub>) is a PUFA has the highest economic importance among the lipids with a value of  $200 \text{ € kg}_{\text{prod}}^{-1}$ . EPA is frequently applied as food additive or a nutritional supplement.<sup>457</sup> For the lipids of lower value, the largest economic potential lies in the conversion to biolubricants, with an estimated sales price that is twice as high as that of biodiesel. The water-soluble storage carbohydrate chrysolaminarin (2.78 wt.%<sub>dry</sub>)

is considered as a natural plant-protection agent,<sup>183</sup> and is known for its antioxidant properties.<sup>458</sup> The selling price of chrysolaminarin is  $20 \text{ € kg}_{\text{prod}}^{-1}$  which is considerably lower than that of EPA and the carotenoids. Chrysolaminarin is the carbohydrate with the highest economic value. However, for the economic profitability of the overall biorefinery process, chrysolaminarin valorisation plays a rather subordinate role. Other water-soluble carbohydrates (11.78 wt.%<sub>dry</sub>) could be marketed as food or feed. The water-insoluble proteins (21.72 wt.%<sub>dry</sub>) and carbohydrates (5.34 wt.%<sub>dry</sub>) could serve as a feedstock for biopolymer production which play an important role in the defossilisation of the polymer industry. Water-soluble proteins (18.13 wt.%<sub>dry</sub>) have emulsifying properties with potential application in the health and food industry.<sup>228</sup> The protein fraction could also be converted to biofuels. However, the economic value of biofuels obtained from proteins is nearly ten-fold lower than that of protein-based biopolymers or functional proteins. Compared with other species, the ash content of *P. tricornutum* is high (23.3 wt.%<sub>dry</sub>). The ash fraction has not been commercially used to date. Depending on the cultivation conditions, *P. tricornutum* ash contains biosilica which could be applied in biosensing or energy applications.<sup>459</sup> However, silica isolation from biomass requires high temperatures, oxidising agents, or strong acids, compromising the integrity of other biomass compounds.<sup>459,460</sup> Alternatively, ash could be applied as an additive for biochar production.<sup>461</sup> In an ideal biorefinery scenario, assuming loss-free separation of the highest value products analysed above, *P. tricornutum* biomass has an overall value of  $26.11 \text{ € kg}_{\text{dry}}^{-1}$  matter which is comparable to other species.<sup>165</sup> Fx and EPA were identified as most valuable compounds, comprising about 90 % of the overall economic value of the biomass. Therefore, a profitable *P. tricornutum* biorefinery requires efficient Fx and EPA extraction and separation. However, both high-value compounds account for only 4.1 wt.%<sub>dry</sub> of the overall biomass. Therefore, these compounds must be extracted at high yields and the remaining biomass fractions should be isolated using inexpensive extraction and separation techniques to increase resource-efficiency while respecting economic feasibility.

## 7.2 Solvent screening

Systematic solvent selection requires a mechanistic understanding of the wet extraction process to deduce meaningful criteria guiding the solvent selection. Therefore, Section 7.2.1 reviews the current knowledge of microalgal wet extraction. Based on this review and the derived solvent selection criteria, the mathematical optimisation problem is formulated (Section 7.2.2), and representative molecules are defined for each biomass fraction (Section 7.2.3). Subsequently, solvent candidates were identified using the developed computational screening approach (Section 7.2.4).



**Fig. 7.1** Biomass composition of moisture-free *P. tricornutum* on a) weight basis and b) based on economic value. The high-value components, EPA and fucoxanthin, render the lipid and pigment fraction the most valuable biomass components. Reprinted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

## 7.2.1 Mechanism of extracting wet microalgal biomass and solvent selection criteria

To select solvents for a biorefinery process that uses wet microalgal biomass, the mechanisms underlying the wet extraction process must be considered to allow for a systematic decision on selection criteria. Solvent-based extraction of wet microalgal paste is a solid-liquid extraction process influenced by interactions between the solvent and the moisture, the cell wall, and the target compounds.<sup>247,257,462</sup> First, the solvent is brought in contact with the algal paste by mixing. Subsequently, the solvent passes the cell wall and the surrounding moisture by diffusion. The *P. tricornutum* cell wall is composed of sulfated  $\alpha$ -mannan with glucuronic acid residues, proteins, and long-chain polyamines.<sup>204</sup> Upon solvent contact, the cell wall polymers may swell and change their conformation depending on the applied solvent.<sup>463</sup> Therefore, the selected solvent likely influences the ease of transport through the cell wall depending on its interactions with cell wall compounds and the moisture. After passing the cell wall, the solvent diffuses to the organelles containing the target compounds and moisture. Some compounds are chemically bound to the target organelles, e.g. the pigments are bound to Fx-chlorophyll-protein complexes *via* hydrogen bonds, located within the thylakoid membranes of the chloroplast.<sup>464</sup> Hence, for efficient extraction of such chemically bound molecules, the solvent i) requires access to the target compounds, ii) must be able to disintegrate the compounds from its storage location, and iii) a high solubility of the target compound is desirable. These points are particularly important when considering wet biomass, since the cells contain moisture and are enveloped by it. Due to the high polarity of water, the moisture likely influences all steps of the extraction process. A detailed mechanistic



explanation of interactions between solvent, biomass, and moisture is still an open field of research. In general, the obtained extract is a mixture of the solvent, multiple extracted target compounds, and water originating from the moisture of the biomass. Therefore, the moisture should be taken into account when designing subsequent separation steps as well. Common extraction techniques for wet microalgal biomass focus on a maximum degree of cell wall disruption, e.g. by sonication or milling, to release the target compounds and to facilitate the transport of the solvent through the cell wall.<sup>247,261,453,454</sup> These approaches consider the cell wall and the moisture as a barrier that must be overcome to maximize the yield, usually aiming at the lipids as the only target biomass fraction. In contrast, this thesis views the moisture and the cell wall as additional degrees of freedom that might be leveraged for extracting and separating all biomass fractions in a biorefinery process. Therefore, the computational solvent screening incorporates the effect of water on the solubility of multiple target compounds. Additionally, the water originating from the moisture can potentially be exploited for inducing the formation of a second liquid phase to separate the extracted target compounds. The use of water for phase formation eliminates the need for additional organic solvents which, in contrast to water, do frequently not align with the principles of green chemistry. Considering the effect of water on the solubilities and the separation of the target compounds, the screening focuses on solvents with three types of miscibility behaviours to be defined at extraction conditions:

- Partially water-miscible (PWM) solvents are miscible with water until the maximum water solubility is reached. After exceeding this solubility limit, two liquid phases are formed. Below the water-solubility limit, PWM solvents allow for extraction under monophasic conditions, decreasing the risk of emulsion formation. For separating the extracted compounds, phase formation can be triggered by adding excess water. The solvent composition of the aqueous and the organic phase commonly represents a mixture, where the aqueous phase is characterised by a higher fraction of water and the organic phase by a higher fraction of organic solvent. Since the phases can usually not be considered as highly pure streams of water and solvent, additional steps for solvent recycling might be required.
- Water-miscible (WM) solvents are miscible with water in all ratios without forming two liquid phases. For efficient separation of the biomass fractions, WM solvents that selectively extract single target compounds can be applied in sequential extraction steps. After extraction, the solvent must be effectively separated from water.
- Non water-miscible (NWM) solvents, also called practically water-immiscible solvents, have negligible water-solubility and immediately form two liquid phases upon contact with water, comparable to mixing water and oil. While wet extraction using NWM solvents, such as alkanes, resulted in low yields of lipophilic compounds, NWM solvents

are generally preferred for separating the mixture of target compounds by liquid-liquid extraction. In contrast to PWM solvents, the solvent composition of the aqueous and the organic phase is generally highly pure, facilitating solvent recycle.

In the following sections, the developed solvent screening approach is applied to wet biomass of the model alga *P. tricornutum* to identify solvents with desired EHS properties, target solubilities and LLE of the solvent/water mixture for application in a biorefinery process.

## 7.2.2 Optimisation problem and screening procedure

Based on the formulation of the general optimisation problem stated in Eqs. 5.1-7.8, the solvent screening problem to identify WM, PWM and NWM solvents is formulated as

$$\min_{\mathbf{s} \in \mathcal{S}_{\text{database}}} F(\mathbf{t}(\mathbf{b}, \mathbf{s}), \mathbf{e}(\mathbf{s}), \mathbf{s}) \quad (7.1)$$

$$\text{s.t.} \quad T_m(\mathbf{s}) \leq 25 \text{ }^\circ\text{C} \quad (7.2)$$

$$40 \text{ }^\circ\text{C} \leq T_b(\mathbf{s}) \leq 120 \text{ }^\circ\text{C} \quad (7.3)$$

$$Q(\mathbf{s}) = 0 \quad (7.4)$$

$$\text{EHS score}(\mathbf{s}) > \text{EHS score}_{\text{hexane}} \quad (7.5)$$

$$x_{b,1} \vee x_{b,i} \vee x_{b,I} \geq x_{b,\text{ref}} \quad (7.6)$$

$$x_i \leq x_{\text{H}_2\text{O}}^{\text{org}} \leq x_j \quad (7.7)$$

$$x_s^{\text{aq}} < 0.1 \quad (7.8)$$

Since *P. tricornutum* contains heat-labile compounds, such as carotenoids, extractions are commonly applied at room temperature. Therefore, the solvent should be liquid at these conditions as given by its  $T_m$  and  $T_b$  limits (Eq. 7.2-7.3, (atmospheric pressure)). The fragility of the carotenoids additionally requires a mild solvent recovery. Therefore, a  $T_b$  of 120 °C is chosen as an upper limit. All solvent candidates were required to have a charge  $Q$  of net zero. Only charge-neutral mixtures of ions as found in ILs were eligible solvent candidates. The EHS score serves as an exclusion criterium to focus the search on benign solvent candidates. For this purpose, the toxic solvent n-hexane, which is commonly applied in microalgal processing, provides a reference EHS score of 0.79 (Eq. 7.5). Furthermore, solubility limits are defined. For each biomass fraction, a reference solvent with a high solubility for the respective fraction is assigned (Tab. 7.1) and the COSMO-RS solubilities of representative molecules  $\mathbf{b}$  in the corresponding reference solvents is predicted ( $T = 25 \text{ }^\circ\text{C}$ ). The reference solvents and representative molecules for proteins, carbohydrates, NLs, PLs, and carotenoid are summarised in Tab. 7.1. The predicted solubilities in the reference solvents serve as a lower solubility boundary. All solvent candidates exhibiting a higher or equal solubility of at least one of the biomass fractions passed the solubility screening step (Eq. 7.6).

Subsequently, the water-miscibility of the candidates is determined. Here, COSMO-RS LLE predictions of a mixture containing the solvent candidate and water, mimicking the moisture of the biomass, are performed ( $T = 25\text{ }^{\circ}\text{C}$ , atmospheric pressure). Based on the LLE predictions, the solvents are assigned to being WM, PWM, or NWM (Eq. 7.7). For solvents considered fully water-miscible, no LLE is predicted. In the context of this thesis, the molar fraction of water in the organic phase is defined as  $0.1 < x_{\text{H}_2\text{O}}^{\text{org}} < 0.9$  for PWM solvents, and  $x_{\text{H}_2\text{O}}^{\text{org}} \leq 0.1$  for NWM solvents. For both, PWM and NWM solvents, the solvent concentration in the aqueous phase should be low to reduce the efforts for additional solvent recovery. Therefore, the molar fraction of solvent in the aqueous phase is limited to  $x_{\text{solv}}^{\text{aq}} < 0.1$ . All solvents within these boundaries pass this screening step. Since the LLE defines the maximum water-solubility of the solvent, additional predictions of the solubility of the representative molecules in the corresponding solvent/water mixture are subsequently performed for selected mixtures. Based on the defined optimisation problem, the database  $\mathcal{S}_{\text{database}}$  (Section 5.2.1) is screened to identify PWM solvents (Section 7.2.4.1), NWM solvents (Section 7.2.4.2), and WM solvents (Section 7.2.4.3). In the described solvent screening problem, the objective function  $F(\cdot)$  is not defined since the extraction mechanisms influencing the yields of the target compounds  $\mathbf{b}$  when extracting wet microalgal biomass are not yet fully understood. Therefore, the solvent screening aims to eliminate unsuitable solvents from the database by evaluating the constraints of the optimisation problem (Eq. 7.2-7.8), to yield a list of PWM, WM, and NWM solvent candidates for further evaluation in experiments. The identified solvent candidates are manually evaluated for commercial availability, suitable price ( $\leq 100\text{ }\text{€}$  per 10 g), EHS properties (according to the solvents' SDSs), and water miscibility.

### 7.2.3 Representative microalgal molecules and reference solvent system

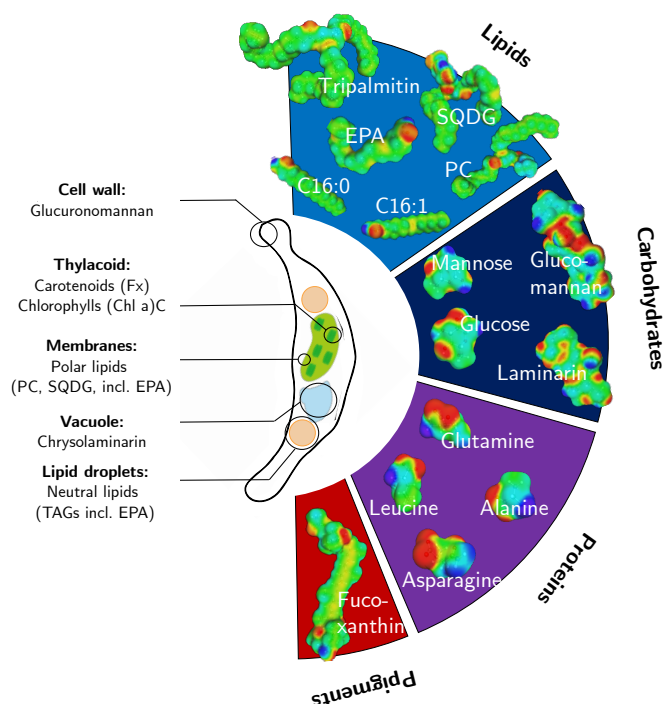
C16:0, C16:1 and EPA were the most abundant fatty acids in the cultivated *P. tricornutum* biomass (Appendix D.3.1). Therefore, these three fatty acids were chosen as representative molecules for SFAs, MUFAs, and PUFAs. Often, these fatty acids are incorporated in TAGs.<sup>198</sup> The TAG tripalmitin consists of three C16:0 fatty acids chains and was selected as a representative molecule for the NLs. PLs were represented by a PC and a SQDG. The PC was composed of C18:0 and C16:0 connected *via* a glycerol group to the phosphatidylcholine functionality ((3-hexadecanoyloxy-2-octadecanoyloxypropyl) 2-(trimethylazaniumyl)ethyl phosphate). The SQDG is composed of glycoside of sulfoquinovose and a diacylglycerol with C16:0 and EPA ([[(2S,3S,4S,5R,6S)-6-[2-hexadecanoyloxy-3-[(5E,8E,11E,14E,17E)-icosa-5,8,11,14,17-pentaenoyl]oxypropoxy]-3,4,5-trihydroxyoxan-2-yl]methanesulfonic acid). Fx, the most abundant carotenoid of *P. tricornutum* (see Appendix D.3.1), was used to represent the pigment fraction. Although chlorophylls are even more abundant in *P. tricornutum*, Fx was selected because it is located in both, the chloroplast and the lipid droplets. Furthermore, Fx is of higher economic interest compared to the chlorophylls (Section 7.1). The

carbohydrate fraction consists of mono- and polysaccharides. The main monosaccharides are D-mannose and D-glucose were selected as monosaccharide-representatives.<sup>174,465</sup> To reduce the computational time required for the quantum chemical calculations, the polysaccharides were represented by simplified structures. Chrysolaminarin was represented by a  $\beta$ -1,3-glycosidically linked polysaccharide made of three glucose units.  $\beta$ -1,6-glycosidically linked branches were neglected. Additionally *P. tricornutum* produces an immunostimulatory, water-soluble polysaccharide that is rich in glucose and mannose.<sup>205,206,466</sup> A simplified model carbohydrate composed of in total four glucose and mannose units connected via  $\beta$ -1,4- glycosidic bonds was selected as a representative molecule. Proteins are the largest polymers in the algal cells and, due to their size and various intramolecular interactions, the most complex fraction of the cell to model. Andersson et al. modelled an entire protein using Turbomole.<sup>467</sup> However, such computations are extremely time-consuming. In this thesis, the most abundant amino acids L-glutamine, L-asparagine, L-alanine, and L-leucine of *P. tricornutum* were used as representative molecules<sup>174,215</sup>. The selected amino acids exhibit different thermodynamic behaviour: L-glutamine and L-asparagine are more hydrophilic than L-alanine and L-leucine.<sup>468</sup> At neutral pH, as given for most of the solvents in the database, the zwitterionic form of amino acids is present<sup>469</sup> which was used for the QC calculations. Water-insoluble proteins and carbohydrates were not considered in the screening as they can be recovered as insoluble cell debris at the end of the extraction process for further valorisation. Therefore, unless otherwise specified, "proteins" and "carbohydrates" refer to the water-soluble part of these fractions. The representative molecules from each biomass fraction were selected (Fig. 7.2) and modelled individually on a quantum chemical level to reduce the complexity (see Section 5.2.5 for computational details).

**Tab. 7.1** Representative molecules according to their fraction in the algal biomass of *P. tricornutum* and their reference solvents. Later in the screening, the solubilities of representative molecules in the solvent candidates is compared to the reference solvents. A solvent candidate will be eliminated in case of lower solubility than the reference.

Fraction	Representative molecules	Reference solvent
Carbohydrates	D-glucose D-mannose Laminarin Glucomannan	Water
Proteins	L-leucine L-alanine L-asparagine L-glutamine	Water
Carotenoids	Fucoxanthin (Fx)	Ethanol
Neutral lipids	Glycerol tripalmitate	N-hexane
Polar lipids	Phosphatidyl choline (PC) Sulfoquinovosyl diacyl glycerol (SQDG) Palmitic acid (C16:0) Palmitoleic acid (C16:1) Eicosapentaenoic acid (EPA)	Ethanol

To identify solvents with high COSMO-RS solubilities of the representative molecules, a



**Fig. 7.2** Representative algae molecules of the diatom *P. tricornutum* with their  $\sigma$ -surfaces. Adapted from König-Mattern et al.<sup>394</sup> with permission from the Royal Society of Chemistry.

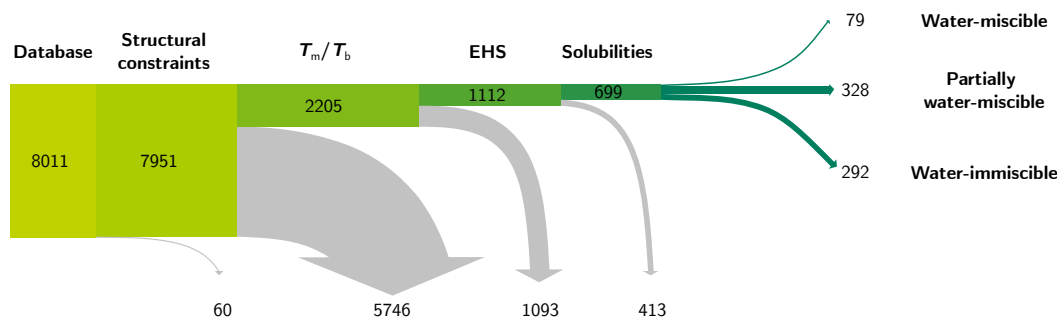
reference solvent system was chosen for each biomass fraction to define threshold COSMO-RS solubilities (Eq. 7.6). Commonly, the NL fraction is extracted with n-hexane. For Fx and PL extraction, typically EtOH is applied.<sup>255,276</sup> A high fraction of the *P. tricornutum* carbohydrates and proteins are water-soluble (see Appendix D.3.1), therefore, water was selected as a reference solvent. All considered biomass fractions, their representative molecules and the assigned reference solvent are summarised in Tab. 7.1. Furthermore, the  $\sigma$ -profiles of the representative molecules, and a comparison of COSMO-RS predicted solubilities and experimental data to validate the chosen representative molecules are provided in Appendix D.1.1.

#### 7.2.4 Screening results

In the first screening steps, 28% of the solvent candidates met the structural constraints and were in the desired  $T_m/T_b$  ranges, as illustrated in Fig. 7.3. In particular, a high number of DESs was excluded during this step due to their high BPs as predicted by COSMO-RS. BPs for all ILs were uncertain as stated in a warning of COSMOtherm, hence, these candidates were kept in the screening to prevent false exclusion. Half of the identified

solvent candidates passed the EHS property constraint (Eq. 7.5) since they were predicted to have a higher EHS score than the reference solvent n-hexane. Mostly, VEGA models were well in line with the corresponding information obtained from safety datasheets of the solvents. Only rare cases, the EHS score was predicted too low. To prevent false exclusion, these solvents were manually re-added to the screening (see list of added in Appendix D.1.2). After this step, around 28% of the solvent candidates were eligible. The thus identified 1112 solvent candidates were subsequently subject to COSMO-RS solubility predictions ( $T = 25\text{ }^{\circ}\text{C}$ ) using the representative molecules defined in Tab. 7.1. If a solvent was predicted to have a higher solubility for one of the considered biomass fractions than the assigned reference solvent (Eq. 7.6, Tab. 7.1), it passed the screening step. After this step, the screening was divided into a screening to identify solvents for the rather hydrophilic fractions (carbohydrates and proteins) and a screening step to find suitable solvents for the lipophilic fractions (NLs, PLs, and pigments). The screening approach identified 16 solvent candidates for the carbohydrate fraction and 18 for the protein fraction (see Appendix D.1.3 for a list of the identified candidates). For these potential solvents, the SDSs and the prices were manually evaluated. The solvent candidate with the most benign EHS properties and the lowest price was water. Other solvent candidates were either expensive (in particular ILs) or no information about the EHS properties was available. Furthermore, several organic acids were among the candidates, leading to degradation of the biomass. Although no other promising solvent could be identified, this result highlights that the screening method was able to re-discover water as the most promising solvent.

For the lipophilic fractions, 699 solvent candidates passed the solubility screening step (Fig. 7.3) which are summarised in Appendix D.1.3. To identify water-miscible, partially water-miscible and practically water-immiscible solvents, the 699 identified candidates were subject to solvent/water COSMO-RS LLE predictions ( $T = 25\text{ }^{\circ}\text{C}$ ). 79 water miscible solvents, 328 partially water-miscible solvents, and 292 water immiscible solvents (as defined in Section 7.2.2) were identified. After evaluating their price and the EHS properties manually, the most promising WM, PWM and NWM solvents were determined. Each solvent class shows a distinct solubility profile (Fig 7.4). All selected solvents were predicted to have low protein carbohydrate solubilities. High NL solubilities were only predicted for the NWM solvents. While WM and PWM solvents tend to have particularly high PL and pigment solubilities, the solubility of these fractions is comparably low in NWM solvents. However, these solubilities were mainly used to discriminate between solvents for lipophilic and hydrophilic representative molecules and do not yet take into account the presence of moisture of the biomass. In the following section, the selected solvent candidates are discussed in detail.



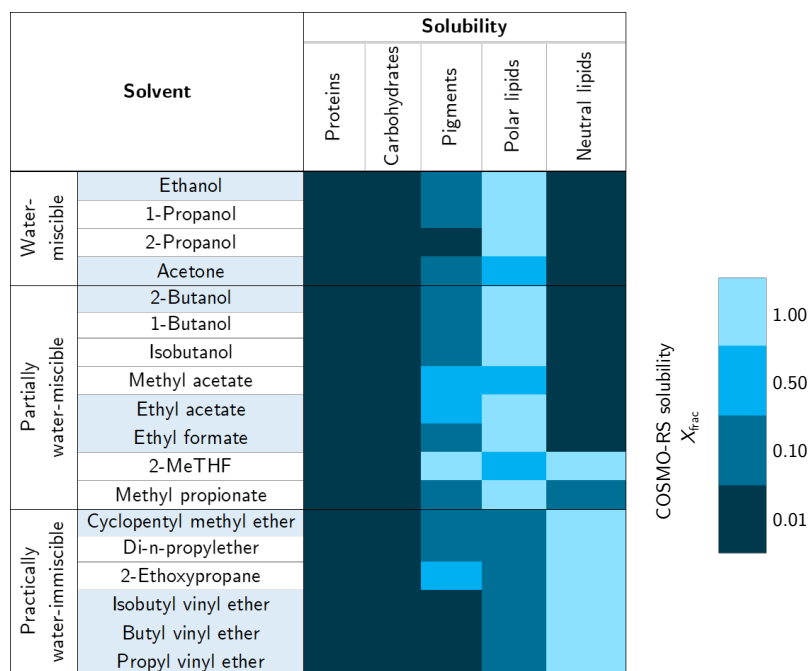
**Fig. 7.3** Screening progress. As a first step, solvents meeting the structural constraints,  $T_m$  and  $T_b$  limits, and EHS criteria were identified. Subsequently, the solubilities of the biomass fractions in the identified solvents were predicted. Solvents with high solubilities for the lipophilic fractions (NLs, PLs, and carotenoids) were selected for predictions of their LLE with water ( $T = 25\text{ }^\circ\text{C}$ ). In this way, the identified solvents were classified as being WM, PWM, or NWM.

#### 7.2.4.1 Partially water-miscible solvents

Before PWM solvent candidates for further experimental evaluation were selected, the influence of the water content in the solvent on the COSMO-RS-predicted solubilities and the partition coefficients of the representative molecules was investigated. For this purpose, the solubilities of the representative molecules in the water-saturated solvents with the phase composition  $x_{\text{H}_2\text{O}}^{\text{org}}$  were predicted using COSMO-RS (Eq. 5.6). The presence of water in the organic solvent strongly affects the solubilities of the representative molecules (Fig. 7.5 a). Solvent/water mixtures with a higher water content have in general a higher solubility of the lipophilic compounds than solvent/water mixtures with a lower water content. Compared to the pure solvent, the solubilities of all lipophilic compounds decreased up to several log units with increasing water content while the solubility of the hydrophilic fractions increased. E.g. for 2-butanol (2-BuOH), one of the solvents with the highest predicted water content in the organic phase (COSMO-RS predicted  $x_{\text{H}_2\text{O}}^{\text{org}} = 0.74$ ), the solubility of NLs decreased from  $\log_{10}(x_{\text{NL}}) = -2.25$  in the pure solvent to  $\log_{10}(x_{\text{NL}}) = -14.87$ . In contrast, the solubility of the carbohydrates increased from  $\log_{10}(x_{\text{carb}}) = -7.51$  in the pure solvent to  $\log_{10}(x_{\text{carb}}) = -2.24$ .

After extracting the target molecules, a convenient method for their separation would be liquid-liquid extraction. Since water is already in the system due to the moisture of the microalgal paste, phase formation can be triggered by further addition of water beyond the miscibility limit. In this way, more hydrophilic molecules tend to move to the aqueous phase, while more less hydrophilic compounds move to the organic phase. Furthermore, the moisture within the biomass can potentially be utilised for the phase separation instead of

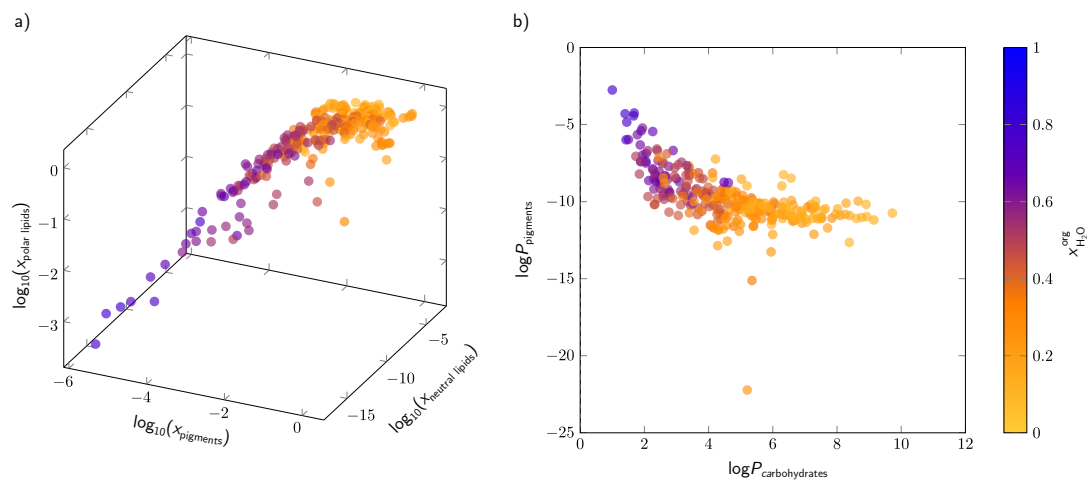




**Fig. 7.4** COSMO-RS predicted solubilities of NLs, PLs, and carotenoids in the selected solvents.

being evaporated in the energy demanding biomass drying step. To evaluate the partitioning behaviour of the representative molecules between organic and the aqueous phase, the partition coefficients were predicted using COSMO-RS (Eq. 5.8). The dependence of the partition coefficients  $\log P$  on  $x_{\text{H}_2\text{O}}^{\text{org}}$  is exemplarily presented for carbohydrates and pigments in in Fig. 7.5 b. A high water content in the organic phase is disadvantageous for the separation of hydrophilic and lipophilic molecules as the partition coefficients approach zero.

The remaining 328 partially water-miscible solvent candidates were manually analysed with respect to commercial availability, price, and EHS properties. Among these candidates were mainly  $\text{C}_4$ -alcohols, and esters derived from carboxylic acids (Tab. 7.2). One DES (decanoic acid/dodecanoic acid,  $x_{\text{decanoic acid}} = 0.67$ ) was identified as a potential candidate. This DES is completely synthesised from fatty acids. Furthermore, the screening procedure identified two trihexyltetradecylphosphonium-based ILs. Both ILs were predicted to have a high solubility of lipophilic compounds, as well as proteins and carbohydrates. The two identified ILs are known as Cyphos IL 103 and 104. However, the recovery of fragile molecules (e.g. pigments) from ILs and DESs is difficult. Advanced separation strategies are necessary for mild product recovery and are currently being researched.<sup>470,471</sup> 2-BuOH, ethyl acetate, and ethyl formate were selected for experiments due to their varying degrees of water-solubility which increases in the order of ethyl formate  $\leq$  ethyl acetate  $\leq$  2-BuOH.




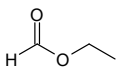


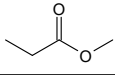
**Fig. 7.5** Effect of the water content in the organic phases ( $x_{H_2O}^{org}$ ) of the identified solvent on COSMO-RS predicted a) solubilities and b) partition coefficients. The solubility in water saturated solvents decreases for all lipophilic fraction with increasing water content, especially for neutral lipids. Partition coefficients are approaching zero with increasing water content. Adapted and reprinted from König-Mattern et al.<sup>394</sup> with permission from the Royal Society of Chemistry.

**Tab. 7.2** Overview of the PWM solvents identified using the computational screening approach. The  $T_b$  was taken from the NIST Chemistry WebBook.<sup>440</sup> The solubilities  $\log_{10}(x_{frac})$  of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem.<sup>414</sup>

Solvent	$T_b$ [°C]	$x_{H_2O}^{org}$	$\log_{10}(x_{frac})$			EHS properties	Structure
			NL	PL	Car.		
2-Butanol	99.5	0.71 <sup>472</sup>	-2.25	0.00	-1.80	flammable, irritant	
1-Butanol	117.7	0.49 <sup>473</sup>	-2.15	0.00	-1.66	flammable, irritant, corrosive	
Isobutanol	108	0.16 (40 °C) <sup>474</sup>	-2.10	-0.01	-1.70	flammable, irritant, corrosive	

Continued on next page


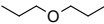

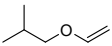

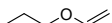
Tab. 7.2 – Continued from previous page

Solvent	$T_b$ [°C]	$x_{H_2O}^{org}$	$\log_{10}(x_{frac})$			EHS properties	Structure
			NL	PL	Car.		
Methyl acetate	57	0.26 <sup>475</sup>	-4.13	-0.31	-0.99	flammable, irritant	
Ethyl formate	54.4	0.17 <sup>476</sup>	-3.13	-0.15	-1.10	flammable, irritant	
2-MeTHF	78	0.16 <sup>477</sup> (19.2 °C)	0.00	-0.68	-0.13	flammable, irritant, corrosive	
Ethyl acetate	77.1	0.13 <sup>475</sup>	-2.03	-0.17	-0.53	flammable, irritant	
Methyl propionate	79.8	0.12 <sup>478</sup> (50 °C)	-1.90	-0.23	-0.87	flammable, irritant	

#### 7.2.4.2 Non-water miscible solvents

In the screening procedure, 292 NWM solvent candidates were identified, among which were dialkylethers and vinyl ethers (Tab. 7.3). Cyclopentyl methyl ether (CPME) was selected for experiments as a dialkyl ether which has a suitable price, as well as desirable LLE, and EHS properties. Di-n-propylether was not selected due to its suitable price. Other Alkylether with longer alkyl chains, such as di-n-butylether, have a more suitable price, however, exceed the desired  $T_b$  limits. Furthermore, propyl vinyl ether (PVE), butyl vinyl ether (BVE), and isobutyl vinyl ether (IBVE) were selected for experiments as these vinyl ethers were not explored for wet extraction for microalgae to date. However, the vinyl ethers must be handled carefully as they pose a risk for explosion and are prone to spontaneous polymerisation reactions.<sup>479</sup> Nevertheless, they are interesting solvent candidates for the substitution of n-hexane due to their otherwise benign EHS properties.

**Tab. 7.3** Overview of the NWM solvents identified using the computational screening approach. The  $T_b$  was taken from the NIST Chemistry WebBook.<sup>440</sup> The solubilities  $\log_{10}(x_{\text{frac}})$  of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem.<sup>414</sup>

Solvent	$T_b$ [°C]	$x_{\text{H}_2\text{O}}^{\text{org}}$	$\log_{10}(x_{\text{frac}})$			EHS properties	Structure
			NL	PL	Car.		
CPME	107	0.01 <sup>480</sup>	0.00	-1.23	-1.24	flammable, irritant	
Di-n-propyl-ether	89	0.04 <sup>481</sup>	0.00	-1.72	-1.94	flammable, irritant	
2-Ethoxy-2-methyl-propane	54	n.a.	0.00	-1.14	-0.53	flammable, irritant	
IBVE	83	n.a.	0.00	-1.37	-2.46	flammable, irritant	
BVE	94	n.a.	0.00	-1.21	-2.22	flammable, irritant	
PVE	65	0.01 <sup>482</sup>	n.a.	n.a.	n.a.	flammable, irritant	

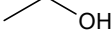
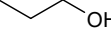
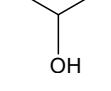
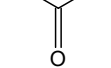
#### 7.2.4.3 Water-miscible solvents

79 WM solvent candidates were identified using the screening approach. Among the identified solvent candidates were  $C_2$ - and  $C_3$ -alcohols, as well as acetone. The reference solvent EtOH was re-discovered by the screening approach, highlighting the applicability of the approach. To ensure structural diversity in the extraction experiments, acetone was chosen as a ketone, and EtOH was chosen as an alcohol. Both solvents are inexpensive, and have benign EHS properties.

### 7.3 Experimental validation

The following sections provide experimental investigations for the selected WM, PWM, and NWM solvents from identified in the Sections 7.2.4.1-7.2.4.3. In the Sections 7.3.1-7.3.3,

**Tab. 7.4** Overview of the WM solvents identified using the computational screening approach. The  $T_b$  was taken from the NIST Chemistry WebBook.<sup>440</sup> The solubilities  $\log_{10}(x_{\text{frac}})$  of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem.<sup>414</sup>

Solvent	$T_b$ [°C]	$\log_{10}(x_{\text{frac}})$			EHS properties	Structure
		NL	PL	Car.		
Ethanol	78	-4.25	-0.22	-1.98	flammable	
1-Propanol	97	-2.86	-0.06	-1.67	flammable, irritant, corrosive	
2-Propanol	82	-3.35	-0.06	-2.04	flammable, irritant	
Acetone	56	-4.83	-0.39	-1.39	flammable, irritant	

the computationally identified solvents are applied in lipid and pigment extractions of wet and freeze-dried *P. tricornutum* biomass to investigate their potential for application in a biomass fractionation process. Based on the computational and experimental results, a novel biorefinery approach for *P. tricornutum* is developed (Section 7.4). The experimental procedures, describing the cultivation, harvest, and extraction procedure, as well as further analysis of the extract composition, are provided in Section 7.5.

### 7.3.1 Lipid extraction from wet *P. tricornutum* biomass

The solvents identified using the computational screening approach (Section 7.2.4) were tested in lipid extraction experiments on wet *P. tricornutum* biomass (moisture content  $MC = 81 - 85$  wt.%). Additionally, the influence of a cell disruption step on the lipid yield was investigated. For this purpose, the samples were sonicated for 2 min using a sonication probe. The resulting lipid extract was separated into its NL and PL fractions using silica columns. The fatty acid composition of the NLs and PLs was determined using GC-FID after conversion to FAMES (see Section 7.5.5). The experimental methods are described in detail in the Sections 7.5.1-7.5.5.

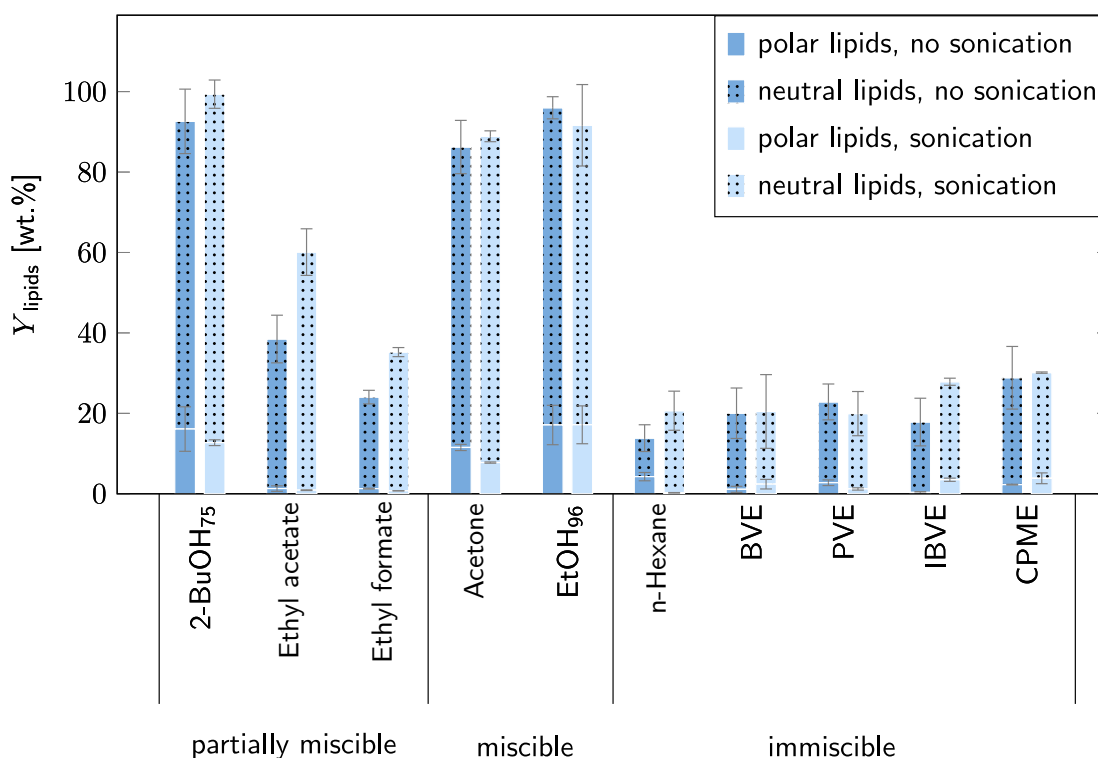
Although the solubility of NLs is typically higher in the nonpolar, NWM solvents (Fig. 7.4), the lipid yield was higher in the WM and PWM solvents (Fig. 7.6). Even without sonication, 96 vol.% ethanol (EtOH<sub>96 vol.%</sub>) extracted 96.0 wt.% of the lipids, followed by 75 vol.% 2-butanol (2-BuOH<sub>75 vol.%</sub>) and acetone, which led to lipid yield of 92.6 wt% and

86.2 wt.%, respectively. With sonication, extraction with 2-BuOH<sub>75 vol.%</sub> resulted in a lipid yield of 99.4 wt.%. These results suggest that the target compound solubility is not entirely the governing factor for the efficient extraction of lipids from wet microalgal biomass. Furthermore, the lipid yields of the untreated and sonicated samples after EtOH<sub>96 vol.%</sub>, 2-BuOH<sub>75 vol.%</sub>, and acetone extraction were comparably high. Presumably, these solvents could efficiently diffuse through the untreated cell walls leading to yields comparable to the sonicated samples. High lipid yields for WM solvents were also reported in the literature. Liu et al.<sup>483</sup> applied the WM solvent 1,2-dimethoxyethane for the extraction of lipids from wet *Botryococcus braunii* resulting in a close to total lipid yield (determined using n-hexane on dry biomass). However, the present regulations limit the use of 1,2-dimethoxyethane on industrial scale due to its health hazard and safety risks.<sup>240</sup> Among the PWM solvents, the highest lipid yield was obtained by 2-BuOH<sub>75 vol.%</sub>. Notably, the obtained lipid yields using the PWM solvents decreased in line with decreasing water-miscibility (water miscibility of 2-BuOH > ethyl acetate > ethyl formate). Extractions employing the NWM solvents resulted in a maximum lipid yield of 33.1 wt.% for CPME and could be barely increased by sonication. All investigated NWM solvents as well as the solvents with low water miscibility (ethyl acetate, ethyl formate) selectively extracted NLs, however, also the overall NL yield was low. Other studies reported comparably low lipid yields for the wet extraction of *Nannochloropsis sp.* and *Chlorella pyrenoidosa* biomass when ethyl acetate, n-hexane, and CPME were used as solvents.<sup>247,248</sup> In these studies, the lipid yields ranged between 20 and 40 wt.% and could be improved to around 60 wt.% by adding a more polar co-solvent such as methanol. Derwenskus et al. performed PLE on wet *Chlorella vulgaris* and *P. tricornutum* biomass.<sup>180</sup> For *P. tricornutum*, increasing yields were obtained for n-hexane < ethyl acetate < EtOH. Similarly, wet *C. vulgaris* was least efficiently extracted by hexane, however, the use of ethyl acetate led to higher lipid yields than EtOH.

### 7.3.2 Carotenoid and chlorophyll extraction from wet microalgal biomass

Subsequently, the chlorophyll and carotenoid yields after extracting wet *P. tricornutum* biomass were determined using the most promising solvents from Section 7.3.1. The PWM solvent 2-BuOH<sub>75 vol.%</sub>, and the WM solvents EtOH<sub>96 vol.%</sub> and acetone were selected, since these solvents obtained the highest lipid yields and are promising for further investigations. These solvents were compared against the NWM solvents CPME and n-hexane. The vinyl ethers were excluded as they were ineffective for the extraction of lipids. Furthermore, white particles with a yellow hue were observed in the solvent containers of the vinyl ethers, presumably caused by spontaneous polymerisation reactions. The experimental procedures are described in the Sections 7.5.4 and 7.5.6.

Similar to the lipid extraction experiments, EtOH<sub>96 vol.%</sub>, acetone, and 2-BuOH<sub>75 vol.%</sub> most effectively extracted chlorophylls and carotenoids. Without sonication, a maximum

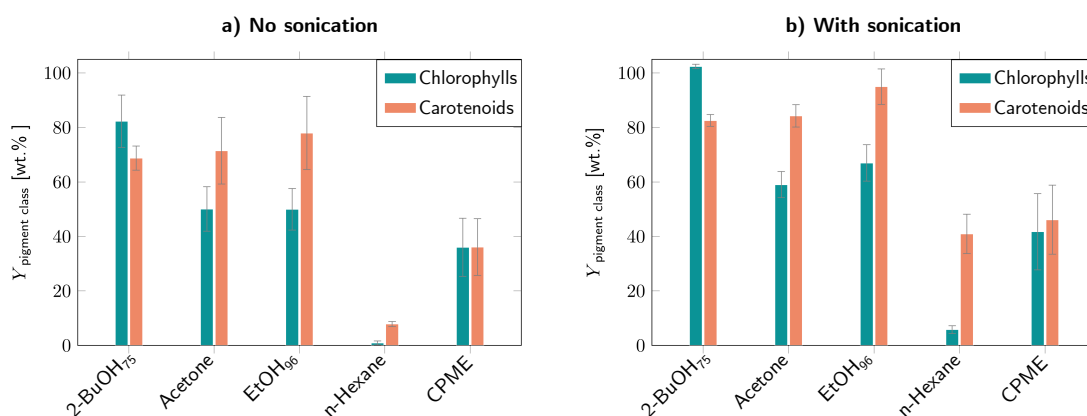


**Fig. 7.6** Lipid extraction ( $t = 90$  min) from wet *P. tricornutum* paste, with and without sonication treatment. Even without sonication, 2-BuOH<sub>75</sub> vol.%, EtOH<sub>96</sub> vol.% and acetone extracted more than 86 wt.% of the lipids. n-Hexane and all other practically water-immiscible solvents in contrast, attained yields below 33 wt.% with sonication treatment. Adapted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

carotenoid yield of 77.8 wt.% was obtained by extraction with EtOH<sub>96</sub> vol.%. 2-BuOH<sub>75</sub> vol.% extractions resulted in the maximum chlorophyll yield with 82.3 wt.% (Fig. 7.7 a). Surprisingly, among the WM and the PWM solvents, the carotenoid yields were lower than that of the lipids (Fig. 7.6), although, conversely, the carotenoid solubility was predicted to be higher than the NL solubility (see Appendix, Fig. D.5). A potential cause for the reverse solubility-yield relationship could be the accessibility of the solvent to the target organelles. The pigments are connected to the FCP complexes located in the thylakoid membranes *via* hydrogen bonds. In contrast, most of the NLs are stored in lipid globules that float inside the cells. Since the lipid globules are presumably easier accessible to the solvent than the FCP complexes, the lipid yield might be higher despite the lower lipid solubility.

A sonication step increased both, the carotenoid and chlorophyll yield, for all tested solvents (Fig. 7.7 b). EtOH<sub>96</sub> vol.% combined with sonication led to the highest carotenoid yield of 95.0 wt.%, followed by acetone. Extraction with 2-BuOH<sub>75</sub> vol.% combined with sonication resulted in complete chlorophyll extraction and a carotenoid yield of 82.6 wt.%,

indicating that the solvent was efficiently transported to the FCPs, however, chlorophylls were more effectively extracted. The NWM solvents hexane and CPME led to the lowest pigment yields. For the NWM solvents, a stronger effect of sonication treatment on the carotenoid yields compared to the lipid yield was observed. For n-hexane, the sonication step increased the carotenoid yield from 7.9 to 40.8 wt.%, whereas the increase in lipid yield was less prominent. While hexane extractions were more selective to the carotenoids than to the chlorophylls, CPME extracted both pigments with comparable efficiency. According to the  $\sigma$ -profiles (Appendix D.1.3.1), CPME exhibits slight HBA behaviour whereas hexane is completely nonpolar. These characteristics might explain the differences in chlorophyll yield between CPME and hexane. Derwenskus et al. reported increasing carotenoid yields for PLE of wet *P. tricornutum* biomass, increasing from hexane < ethyl acetate < EtOH<sup>180</sup> which is in line with the obtained experimental results in this work.



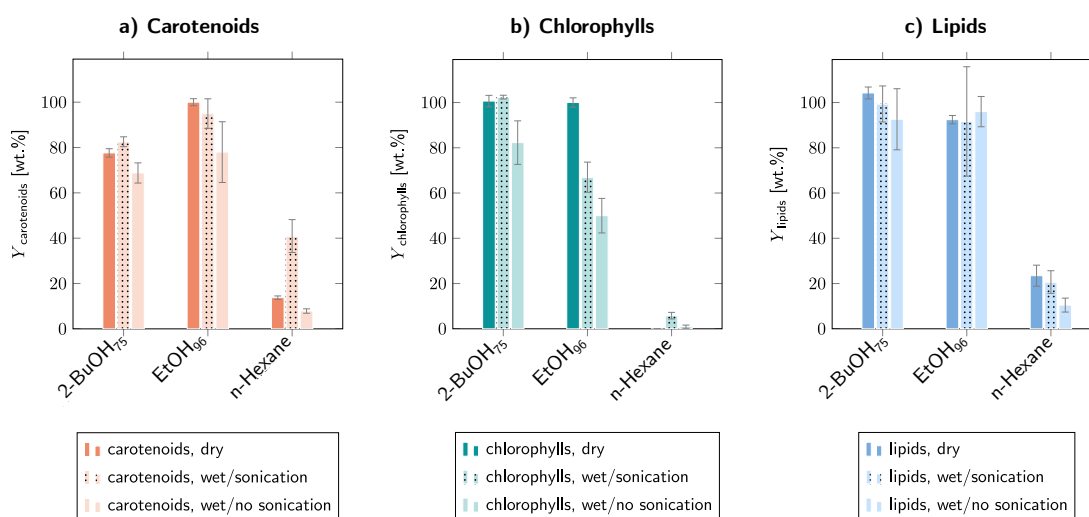
**Fig. 7.7** Chlorophyll and carotenoid extraction ( $t = 90$  min) from wet *P. tricornutum* paste ( $MC = 81 - 85$  wt.%), a) with sonication ( $t = 2$  min) and b) without sonication. Adapted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

### 7.3.3 Lipid and pigment extraction from dried microalgal biomass

To compare the yields obtained from wet and dry biomass, extractions were performed on lyophilised biomass ( $MC = 8.5$  wt.%) and compared to the yields of the wet extractions. For the dry extractions, 2-BuOH<sub>75</sub> vol.% and EtOH<sub>96</sub> vol.% were chosen since they were the best-performing PWM and WM solvents, respectively. The NWM solvent n-hexane is commonly used as a solvent for lipid and pigment extraction on dry biomass and was used as a benchmark. The experimental details are explained in detail in the Sections 7.5.4- 7.5.6. The highest carotenoid yield was obtained for dry extractions using EtOH<sub>96</sub> vol.% (corresponding to reference method for pigment extraction, therefore, 100 wt.% carotenoid yield) (Fig. 7.8 a). Additionally, absolute EtOH was tested for carotenoid extraction on dry biomass, however, resulting in a lower yield than that obtained with EtOH<sub>96</sub> vol.% (data not shown). Wet



extraction with EtOH<sub>96</sub> vol.% and sonication was slightly less efficient (within standard deviation), resulting in a carotenoid yield of 95.0 wt.%. 2-BuOH<sub>75</sub> vol.% (wet + sonication and dry route) and EtOH<sub>96</sub> vol.% (dry route) led to complete chlorophyll extraction (Fig. 7.8 b). The most effective solvent for the dry extraction of lipids was 2-BuOH<sub>75</sub> vol.%, resulting in a slightly higher lipid yield than the reference method (chloroform/methanol, 1/1, v/v) (Fig. 7.8 c). The experiments showed that the extraction of wet, undisrupted biomass was the least effective method for all analysed biomass fractions. When combined with a sonication step, lipid and pigment yields comparable to the dry route could be obtained for EtOH<sub>96</sub> vol.% and 2-BuOH<sub>75</sub> vol.%. In particular, for 2-BuOH<sub>75</sub> vol.% and n-hexane, the sonication step boosted the carotenoid yields beyond dry route levels. An increase of carotenoid yields was also reported for wet extraction of disrupted *Chlorella thermophila* by Sarkar et al.<sup>484</sup> Since the cell disruption step is less energy-intensive than a drying step, these findings are highly important for the design of microalgal biorefineries.



**Fig. 7.8** Dry extractions ( $t = 90$  min) of freeze-dried *P. tricornutum* biomass (moisture content = 8.5 wt.%) compared to wet *P. tricornutum* paste (moisture content = 81 - 85 wt.%). The ratio of algal dry matter to solvent was held constant. The yields of carotenoids, chlorophylls and lipids are shown in a), b), and c), respectively. Adapted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

## 7.4 Lab-scale fractionation process for wet *P. tricornutum* biomass

The economic analysis of potential *P. tricornutum* biorefinery products revealed that the carotenoids and the lipids have the highest economic values (see Section 7.1). Therefore,

separating lipids and carotenoids is crucial for the economic viability a *P. tricornutum* biorefinery process. In the literature, two strategies for this separation were reported: anti-solvent or temperature-induced precipitation of Fx with subsequent filtration,<sup>276</sup> or liquid-liquid extraction followed by chromatography.<sup>485</sup> The experiments performed in Section 7.3.1 showed cell disruption affects pigment and carotenoid extraction differently. Cell disruption is not necessary for efficient lipid extraction from wet *P. tricornutum* when solvents with high water miscibility are employed. In contrast, pigment extraction without cell disruption was less effective (Section 7.3.2). For an incubation time of  $t = 90$  min, 2-BuOH<sub>75</sub> vol.% extracted 92.6 wt.% of the lipids but only 68.8 wt.% of the pigments from wet, undisturbed biomass. Tailoring the operational parameters towards high lipid and low carotenoid yields potentially leads to a more selective extraction of lipids, whereas the carotenoids preferably would remain in the biomass. A second extraction stage could extract the residual pigments. In this manner, the carotenoid-lipid separation would be integrated in the extraction stage. In the following section, experiments investigating the potential for integrated extraction and separation of lipids and carotenoids are performed. Based on these experiments, a lab-scale biorefinery process is developed.

#### 7.4.1 Integrated extraction and separation of lipids and carotenoids from wet *P. tricornutum* biomass

To study the effect of incubation time and water content  $w_{\text{water,extr}}$ , 2-BuOH was selected as a solvent. 2-BuOH has a large window of water-miscibility in which the polarity of the solvent/water mixture can be conveniently tuned. After extraction, adding water beyond the miscibility window triggers phase formation for separating co-extracted carbohydrates and proteins, which is particularly interesting for designing a biorefinery process. In the experiments with the incubation time and the water content  $w_{\text{water,extr}}$  were independently varied to systematically study their effect on carotenoid and pigment yields. Lipid and pigment extraction experiments on wet *P. tricornutum* biomass were performed as described in Section 7.5.4 with varying water amount at constant incubation time  $t = 90$  min. Furthermore, experiments were performed at varying incubation time at constant  $w_{\text{water,extr}} = 32$  wt.% (corresponding to 2-BuOH<sub>75</sub> vol.%).

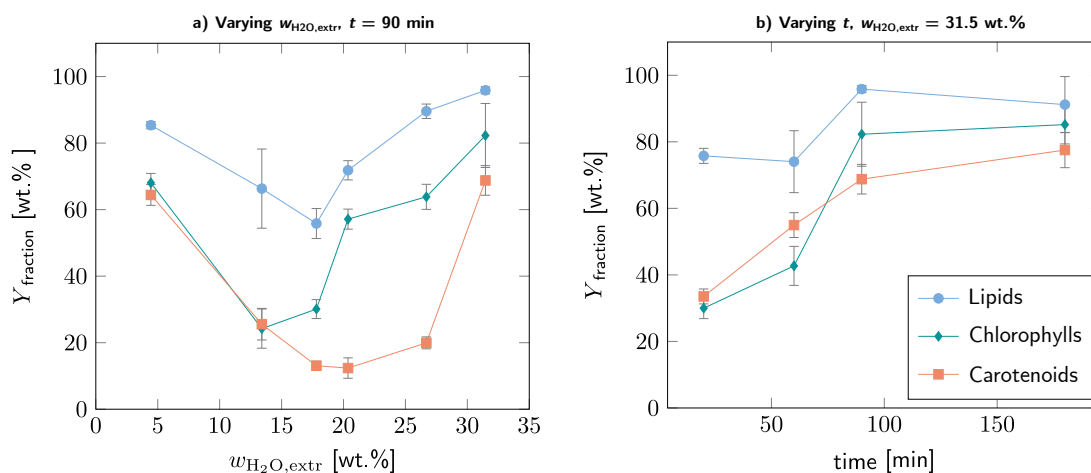
For the polarity modification of the 2-BuOH/water mixture, first, the miscibility range was assessed. The LLE of 2-BuOH and water ( $T = 20$  °C) was experimentally determined by Lladosa et al.<sup>472</sup> The mass fraction of water in the organic phase at LLE was  $w_{\text{H}_2\text{O}}^{\text{org}} = 36$  wt.%. Thus, to prevent phase separation during extraction, the water content in the 2-BuOH/water-mixture must be held below  $w_{\text{water,extr}} < 36$  wt.%. The water content during extraction  $w_{\text{water,extr}}$  was varied between  $5$  wt.%  $< w_{\text{water,extr}} < 32$  wt.% to stay closely within the miscibility window. A water content of  $w_{\text{water,extr}} = 32$  wt.% corresponds to a 2-BuOH content of 75 vol.% (2-BuOH<sub>75</sub> vol.%, neglecting the excess volume of the real

mixture). Please, note the difference between weight- and volume-based expressions. The water content in the system  $w_{\text{water,extr}}$  also accounts for the mass of water originating from the moisture of the biomass  $m_{\text{moisture biomass}}$  and is defined as

$$w_{\text{water,extr}} = \frac{m_{\text{moisture biomass}}}{m_{\text{moisture biomass}} + m_{\text{H}_2\text{O,s}} + m_s}, \quad (7.9)$$

where the mass of solvent is given as  $m_s$  and the mass of water additionally to the solvent is denoted as  $m_{\text{H}_2\text{O,s}}$ .

Experiments with varying  $w_{\text{water,extr}}$  were performed on wet *P. tricornutum* paste for  $t = 90$  min (no sonication). A parabolic relationship between yields and  $w_{\text{water,extr}}$  was observed. A minimum pigment and lipid yield was observed between  $17 \leq w_{\text{water,extr}} \leq 21$  wt.% (Fig. 7.9 a). The highest experimentally tested water content led to the highest lipid (95.9 wt.%), carotenoid (68.7 wt.%) and chlorophyll yield (82.3 wt.%). This parabolic yield- $w_{\text{water,extr}}$  relationship is surprising, since, according to the “like dissolves like” principle, a higher polarity of the solvent is associated with lower solubility of nonpolar biomass compounds, which was confirmed by COSMO-RS solubility predictions (Appendix, Fig. D.5). Therefore, the experimental results cannot be explained by the solubility of the target compounds. Interestingly, also Ren et al. reported increased lipid yields when a water treatment was performed between extraction stages with organic solvents.<sup>486</sup> The water treatment caused cell wall alterations that were attributed to the increase in lipid yield. Presumably, the combined effects of solvent-cell wall interactions, solvent-target organelle interactions, and target compound solubility contribute to the observed parabolic yield profile. However, more research is required to quantify the extent to which each of the phenomena influence the yield. For efficient carotenoid-lipid separation, maximum selectivity towards lipids is desirable. The carotenoid-lipid selectivity can be described by the difference between carotenoid and lipid yield which reached a maximum between  $20 \text{ wt.\%} \leq w_{\text{water,extr}} \leq 30 \text{ wt.\%}$  where about 80 wt.% of the lipids were extracted and 80 wt.% of the carotenoids remained in the residual biomass. Additionally, the dependence of the yields on the incubation time, at  $w_{\text{water,extr}} = 32$  wt.% was experimentally investigated to study the extraction rates of carotenoids and lipids as potential means for their separation. The lipids were more rapidly extracted than the pigments (Fig. 7.9 b). After an incubation time of  $t = 20$  min, already 75.8 wt.% of the lipids were extracted while 66.5 wt.% of the carotenoids remained in the biomass. The observed differences in extraction rates offer an additional option to separate lipids and carotenoids.



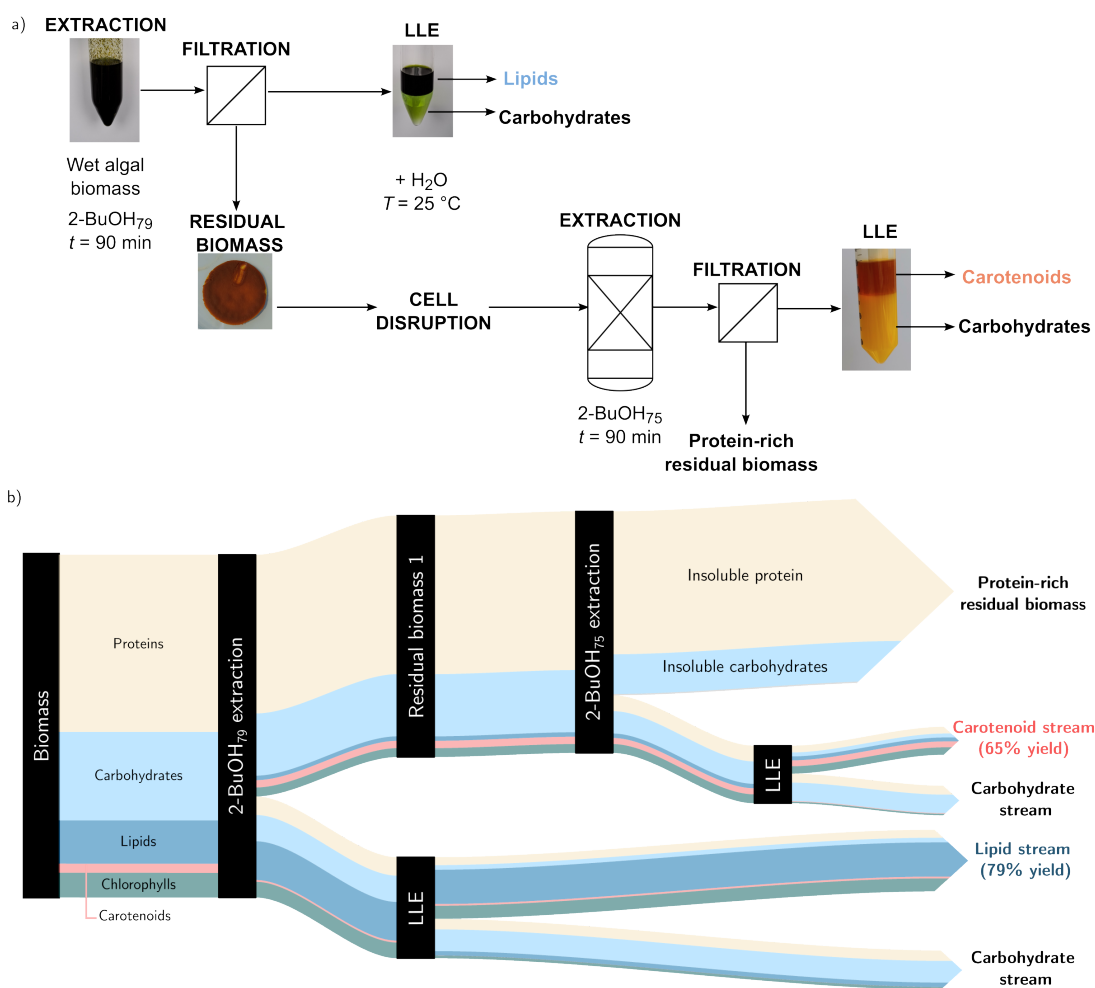
**Fig. 7.9** Influence of the water content  $w_{water,extr}$  and the incubation time  $t$  on lipid, carotenoid, and chlorophyll yields during extraction of wet *P. tricornutum* biomass ( $MC = 81 - 85$  wt.%, no sonication). a) Varying  $w_{water,extr}$  at constant incubation time ( $t = 90$  min), b) varying incubation time  $t$  at constant water content ( $w_{water,extr} = 32$  wt.%). Adapted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

#### 7.4.2 Lab-scale 2-butanol/water-based biomass fractionation process for wet *P. tricornutum* biomass

Based on the observed differences in carotenoid and lipid yields for varying  $w_{water,extr}$  and incubation time, a lab-scale biomass fractionation process was developed. The novel biorefinery process consists of two extraction units (Fig. 7.10 a) that selectively extract lipids and carotenoids, respectively. The extract streams are subject to separation units in which water addition triggers phase separation of the 2-BuOH/water mixture, allowing to further fractionate the extracted compounds. The yields of extracted lipids, pigments, proteins, and carbohydrates were experimentally determined. Furthermore, the amount of these fractions contained in the aqueous and organic phases was measured. The applied experimental procedures are described in the Sections 7.5.4-7.5.8, and 7.5.11. Based on the yields and the partition coefficients, the mass balance for the developed process was modelled (Appendix D.6) to calculate the overall yields for each fraction and to provide an overview of the mass flows to assess potentials for process optimisation.

In the biorefinery process (Fig 7.10 a), wet *P. tricornutum* biomass ( $MC = 81 - 85$  wt.%) was first extracted with 79 vol.% 2-butanol (2-BuOH<sub>79 vol.%</sub>) ( $t = 90$  min) which corresponds to  $w_{water,extr} = 26$  wt.% where a maximum difference in carotenoid and lipid yields was observed (Fig. 7.9 a). In this step, 89.6 wt.% of the lipids were extracted and 80.0 wt.% of the pigments remained in the residual biomass. The lipid-rich extract was further purified by phase separation which was triggered by adding water. In this way, the lipids partitioned

to the organic phase while the co-extracted carbohydrates and proteins partitioned to the aqueous phase. The residual biomass showed an orange hue due to the remaining pigments (Fig. 7.10 a). After a cell disruption step, the second extraction step using 2-BuOH<sub>75 vol.%</sub> ( $t = 90$  min) aimed to extract the residual pigments at maximum yield. To separate the pigments from the co-extracted carbohydrates and proteins, water was added to induce phase formation. The carotenoids were accumulated in the organic phase, while the carbohydrates and proteins partitioned to the aqueous phase. Based on the determined mass flows, the final carotenoid-rich stream (Fig. 7.10 b) contained 65 wt.% of the carotenoids that were initially present in the biomass. The final lipid yield was 79 wt.%. Carbohydrates were the predominant component in the aqueous phases of both phase separation steps. Proteins were the main compound of the residual biomass remaining after the second extraction stage. Therefore, the developed biorefinery process effectively fractionated the biomass into lipids, carotenoids, carbohydrates and proteins by employing only 2-BuOH and water as benign solvents at ambient conditions. In contrast to PLE, the developed process does not require costly equipment, high temperatures, and high pressures. Furthermore, the novel biorefinery process eliminates the energy-intensive biomass drying step that is commonly applied in microalgal biomass processing. This process is still in the early-stage development and several options remain for further improvements. The carotenoid/lipid-separation, representing the compounds with the highest economic value, should be further optimised. Their separation could be improved by applying design of experiments-approaches, taking into account the extraction rates in addition to the water content  $w_{\text{water,extr}}$ . Mixing rates and biomass loading should be investigated as additional degrees of freedom for increasing the lipid selectivity of the first extraction since extraction rates are highly dependent on solvent contact with the target compounds. Considering their high economic value, losses stemming from co-extracted carotenoids in the first extraction stage and unextracted carotenoids in the second extraction stage should be minimised. In further studies, the approach should be tested and adopted for other microalgal species. Envisioning an industrial scale biorefinery process, process units for energy-efficient solvent recycling should be designed and optimised. The organic and aqueous phases originating from the separation units represent azeotropic 2-BuOH/water-mixtures that can be separated by pervaporation.<sup>487</sup> Since the process employs 2-BuOH/water mixtures for both extraction steps, there is no need to recover highly pure solvents, reducing the energy demand for solvent recycling. Overall, the novel biorefinery approach represents a simple and inexpensive method based on only two benign solvents to fractionate *P. tricornutum* biomass at high moisture contents and operates at ambient conditions. For a potential scale-up, carotenoid losses should be minimised and effective solvent recycling strategies should be implemented.



**Fig. 7.10** Overview of the developed biorefinery process for wet *P. tricornutum* biomass. a) Process flow diagram. b) Corresponding mass flows modelled on the basis of experimental measurements. The final yields are based on the weight of the fractions in the initial biomass. Reprinted from König-Mattern et al.<sup>455</sup> with permission from Elsevier.

## 7.5 Experimental methods

### 7.5.1 Cultivation of *P. tricornutum*

*P. tricornutum* (strain 1090-1b, Culture Collection of Algae, University Göttingen, SAG) was cultivated in modified Mann and Myers medium<sup>276,488</sup> in a 28 l flat panel airlift photobioreactor (Subitec, Germany). To prepare the medium, 10.0 g l<sup>-1</sup> NaCl, 2.8 g l<sup>-1</sup> MgSO<sub>4</sub> · 7H<sub>2</sub>O, 2.8 g l<sup>-1</sup> MgCl<sub>2</sub> · 6H<sub>2</sub>O, 1.2 g l<sup>-1</sup> CaCl<sub>2</sub> · 2H<sub>2</sub>O. After sterile-filtrating the medium, 2 ml of a sterile trace element solution consisting of 3 g l<sup>-1</sup> H<sub>3</sub>BO<sub>3</sub>, 0.7 g l<sup>-1</sup> MnCl<sub>2</sub> · 4H<sub>2</sub>O, 0.17 g l<sup>-1</sup> ZnSO<sub>4</sub> · 7H<sub>2</sub>O, 2.86 mg l<sup>-1</sup> CoCl<sub>2</sub> · 6H<sub>2</sub>O and 0.13 g l<sup>-1</sup> Na<sub>2</sub>MoO<sub>4</sub> · 2H<sub>2</sub>O was added. Iron, ammonium, nitrate, and phosphate were added separately as stock solutions to reach ion concentrations of 1 mg l<sup>-1</sup> Fe<sub>3</sub><sup>+</sup>, 50 mg l<sup>-1</sup> PO<sub>4</sub><sup>3-</sup>, 75 mg l<sup>-1</sup> NH<sub>4</sub><sup>+</sup>, and 500 mg l<sup>-1</sup> NO<sub>3</sub><sup>-</sup> in the medium. Fe-(III)-citrate · H<sub>2</sub>O, K<sub>2</sub>HPO<sub>4</sub>, KH<sub>2</sub>PO<sub>4</sub>, NH<sub>4</sub>HCO<sub>3</sub>, and KNO<sub>3</sub> were used to prepare the respective stock solutions. The ratio of K<sub>2</sub>HPO<sub>4</sub> to KH<sub>2</sub>PO<sub>4</sub> in the PO<sub>4</sub><sup>3-</sup> stock solution was 1.27/1 (w/w). NO<sub>3</sub><sup>-</sup> and PO<sub>4</sub><sup>3-</sup> concentrations in the medium were daily measured by ion exchange chromatography (Metrohm, Compact IC Flex oven/SeS/PP/Deg) equipped with an anion column (Metrosepp A Supp 5/150/4.0). The medium was supplemented by nitrate and phosphate stock solutions to the desired concentrations. The reactor was equipped with warm-white LEDs (Sanlight e.U., P4W). The light intensity was measured using a PAR-sensor (LI-COR, LI19OR-DWC) and adjusted with increasing culture density. The pH was adjusted to 7.4 by automatic addition of NaOH and CO<sub>2</sub>. The temperature was maintained at 20 °C as measured with a temperature sensor. Dissolved oxygen was measured by an optical pO<sub>2</sub> electrode (Hamilton, Visiform DD ARC 12 H0). At the end of the exponential growth phase, cells were harvested by centrifugation and washed with MilliQ water. The resulting algal paste (MC = 81 – 85 wt.%) was aliquoted in 50 ml Falcon tubes wrapped with aluminium foil, and stored at -20 °C until further use.

### 7.5.2 Determination of the moisture content

1 g of wet biomass was placed to a dried and pre-weighed porcelain crucible and oven-dried overnight until a constant weight was reached. The weight of the biomass-filled crucibles was recorded before and after drying. The amount of water in the biomass was determined (Eq. 7.10). The moisture content *MC* was determined for each aliquot of algal paste (Section 7.5.1).

$$MC [\text{wt.}\%] = \frac{m_{\text{wet biomass}} - m_{\text{dry biomass}}}{m_{\text{dry biomass}}} \cdot 100\% \quad (7.10)$$

### 7.5.3 Lyophilisation

For experiments on freeze-dried biomass, the algal paste as obtained in Section 7.5.1, was subjected to lyophilisation (Vaco2, Zirbus Technology) overnight. The resulting dried biomass pellets were ground with mortar and pestle to obtain a fine powder ( $MC = 8.5$  wt.%).

### 7.5.4 Extraction procedure for wet and dry biomass

500 mg of wet algal paste (see Section 7.5.1) was weighed into 50 ml Pyrex tubes for wet extraction. For cell disruption, 2 ml of the solvent or solvent mixture was added and the samples were sonicated for 2 min (amplitude = 80%, cycle = 0.6, UIS250V combined with a LS24d7-L2 probe, Hielscher). After sonication, further 8 ml of solvent or solvent mixture was added. In experiments without sonication, a stirring bar and 10 ml of the solvent system were directly added to the algal paste. For the extraction of dry biomass, 100 mg of lyophilised biomass ( $MC = 8.5$  wt.%, corresponding to the same mass of dry matter as in the wet extraction experiments), were used and 10 ml solvent or solvent mixture were added. The samples were incubated for 90 min on a magnetic stirrer (250 rpm). Subsequently, the samples were filtered through 0.2  $\mu\text{m}$  PTFE syringe filters and 10 mg butylhydroxytoluol (BHT) was added. The extracts were stored at  $-20$  °C until further analysis. The samples were protected from light during all steps using aluminium foil. Microscopic images of the wet *P. tricornutum* biomass are provided in Appendix D.4.

### 7.5.5 Lipid analysis

The extracts were thawed and the solvent was evaporated under a gentle stream of  $\text{N}_2$ . The lipid extract was separated into neutral and polar lipids using silica columns as described by Breuer et al.<sup>489</sup> with minor modifications. Glyceryl trinonadecanoate (C19-TAG, Sigma Aldrich) served as the internal standard for NLs, and 1,2-dipentadecanoyl-sn-glycero-3-phosphocholine (C15-PL, Avanti Polar Lipids) as the as internal standard for PLs. Solid phase extraction columns (Sep Pak Vac, 6 cc, 1 g silica, Waters) were equilibrated using 10 ml n-hexane. The lipid extract was dissolved in chloroform/methanol and was applied to the equilibrated columns. Neutral lipids were eluted using 10 ml n-hexane/diethyl ether solution (7/1, v/v). PLs were eluted with 10 ml acetone/methanol/n-hexane (2/2/1, v/v/v) followed by 10 ml methanol. The eluents were evaporated under  $\text{N}_2$ . In case no separation of NLs and PLs was performed, only C15-PL was added as the internal standard to the lipid extract.

The lipid extracts containing the internal standard were methylated and analysed by GC-FID as described by Breuer et al.<sup>490</sup> with minor modifications. For methylation, 3 ml



of methanol containing 5 vol.% concentrated  $\text{H}_2\text{SO}_4$  was added to the lipid extracts and heated to 70 °C for 3 h in a shaker (Thermo Mixer C, Eppendorf). After methylation, the samples were cooled to room temperature. Subsequently, 8 ml of 5% NaCl/water solution (w/v) and 3 ml of heptane were added for phase separation. During methylation, the lipids were converted to fatty acid methyl esters (FAMES), which were collected in the heptane phase. 1  $\mu\text{l}$  of the heptane phase was injected into the GC system (6890 N Agilent) which was equipped with a HP88 column (60 m, 0.25 mm diameter). The flow rate was 0.9  $\text{ml min}^{-1}$ . The FAMES were detected using a FID. The amount of lipid was calculated based on calibrations for each FAME (C12:0, C14:0, C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:5 (EPA)) with respect to the internal standards as described by Breuer et al.<sup>490</sup>

The total lipid content was determined by extracting 100 mg of lyophilised biomass with 10 ml chloroform/methanol (1/1, v/v). The samples were methylated and FAMES were quantified using GC as described above. The lipid content was given as wt.% of algal dry matter.

### 7.5.6 Pigment analysis

The pigments were quantified according to Zapata et al.<sup>252</sup> with minor modifications. The extraction solvent was evaporated under  $\text{N}_2$  in amber Eppendorf tubes in the dark. The dried pigment extract was redissolved in 1 ml 95% aq. methanol solution (v/v) containing 0.1 wt.% BHT. 833  $\mu\text{l}$  methanol solution containing the sample were mixed with 167  $\mu\text{l}$  MilliQ water for injection into a high-pressure liquid chromatography (HPLC) system (1290 Infinity, Agilent). The injection volume was 40  $\mu\text{l}$ . The HPLC system was equipped with an Eclipse XDB-C8 Zorbax column (4.6 mm x 150 mm x 3.5  $\mu\text{l}$ ). Chlorophylls and carotenoids were eluted using methanol/acetonitrile/aqueous pyridine solution (2/1/1, v/v/v, eluent A) and methanol/acetonitrile/acetone (1/3/1, v/v/v, eluent B). The aqueous pyridine solution was adjusted to pH = 5 with acetic acid and the pyridine concentration was 0.25 M. The eluents dosage was adjusted according to the following ramp: 0 min: 100% eluent A, 22 min: 60% eluent A, 28 min: 5% eluent A, 38 min: 5% eluent A, 40 min: 100% eluent A. The flow rate was 1  $\text{ml min}^{-1}$ . The pigments were detected by diode-array spectroscopy (350 to 750 nm wavelength) and the chlorophylls were additionally identified by a fluorescence detector (excitation: 440 nm, emission: 650 nm). The amount of chlorophylls (chlorophyllide a, chlorophyll c, pheophorbide a, chlorophyll a) and carotenoids (Fx, Ddx, Dx,  $\beta$ -Car) was calculated using calibration with authentic standards. All standards were obtained from DHI Laboratory Products Denmark, except the fucoxanthin standard was purchased from Sigma Aldrich.

The amount of carotenoid was calculated as the sum of all carotenoids used in the calibration. Chlorophylls are very fragile and prone to degradation. Chl a is degraded to chlorophyllide a Chld a which is further degraded to pheophorbide a Phbd a.<sup>491,492</sup> To account for the change of the total chlorophyll amount resulting from the degradation reactions, the detected amount of Chld a and Phbd a is converted into Chl a-equivalents using their molar weights:

$$\text{Chl a eq.} = \left( m_{\text{Phbd a}} \frac{M_{\text{Chld a}}}{M_{\text{Phbd a}}} \right) \frac{M_{\text{Chl a}}}{M_{\text{Chld a}}} \quad (7.11)$$

Total chlorophylls were calculated as the sum of Chl c, Chl a, and Chl a equivalents. Total carotenoids are calculated as the sum of Fx, Ddx, Dx, and  $\beta$ -Car amounts. The total amount of chlorophylls and carotenoids was determined by extracting 100 mg of lyophilised biomass with 10 ml of EtOH<sub>96 vol.%</sub>. The amount of chlorophylls and carotenoids was given as wt.% of algal dry matter.

### 7.5.7 Protein analysis

After extraction (Section 7.5.4), the solvent was evaporated under N<sub>2</sub>, the extracts were redissolved in MilliQ water, and subjected to an ultrasound bath (Sonorex RK102H S, Bandelin) for 15 min. The suspension was filtered through 0.2  $\mu\text{m}$  PTFE filters, to separate the undissolved lipids from the liquid filtrate containing the dissolved protein. The amount of protein was determined using the method of Lowry et al.<sup>493</sup> Absorption of the resulting blue solution was read at 750 nm using UVvis (Specord S600, Analytik Jena). Calibration was performed using bovine serum albumin as a reference protein. The total protein content of the biomass was determined by treating a suspension containing 0.3 mg ml<sup>-1</sup> algal dry matter with the Lowry reagents. The protein content was given as wt.% of algal dry matter.

Water-soluble protein was determined by extracting 1000 mg of wet algal paste with MilliQ water. First, 5 ml of water was added to the biomass and treated with a sonication probe (amplitude = 80%, cycle = 0.6, UIS250V combined with a LS24d7-L2 probe, Hielscher) for 5 min. The samples were extracted for further 85 min on a magnetic stirrer at room temperature. The samples were filtered through 0.2  $\mu\text{m}$  PTFE filters and directly analysed according to the Lowry method.

### 7.5.8 Carbohydrate analysis

To determine the amount of carbohydrates, after extraction (Section 7.5.4), the solvent was evaporated under a stream of N<sub>2</sub>. The dried extract was redissolved in MilliQ water and subjected to an ultrasound bath for 15 min, followed by filtration through 0.2  $\mu\text{m}$  PTFE syringe filters. The amount of carbohydrates was determined by the phenol-sulfuric

acid method.<sup>494</sup> The absorption of the resulting golden-brown liquid was measured at a wavelength of 488 nm using an UVvis photo spectrometer (Specord S600, Analytik Jena). Calibration was performed using D-glucose. The total carbohydrate content of the biomass was determined on a suspension of 0.3 mg ml<sup>-1</sup> algal dry matter. The suspension was directly treated with H<sub>2</sub>SO<sub>4</sub> and phenol. The carbohydrate content was given as wt.% of algal dry matter.

Water-soluble carbohydrates were determined by extracting 1000 mg of wet algal paste with MilliQ water. First, 5 ml of water was added to the biomass and treated with a sonication probe (amplitude = 80%, cycle = 0.6, UIS250V combined with a LS24d7-L2 probe, Hielscher) for 5 min. The samples were extracted for further 85 min on a magnetic stirrer at room temperature. The samples were filtered through 0.2 µm PTFE syringe filters and directly analysed according to the phenol-sulfuric acid method.

The laminarin content was determined using an enzymatic assay (yeast β-glucan assay kit, K-EBHLG, Megazyme).

### 7.5.9 Yield calculation

The yield  $Y_{\text{fraction}}$  of the extracted biomass fractions is defined as

$$Y_{\text{fraction}} [\%] = \frac{m_{\text{fraction,extracted}}}{m_{\text{fraction,total}}} \cdot 100, \quad (7.12)$$

where  $m_{\text{fraction,extracted}}$  denotes the mass of the fraction determined by the extraction experiments and  $m_{\text{fraction,total}}$  represents the total mass of the fraction in the biomass.

### 7.5.10 Ash content

The ash content was determined by weighing 50 mg of dried algal biomass into dried and pre-weighed porcelain crucibles. The crucibles were subjected to a muffle furnace for 12 h at 450 °C and the weight was noted. The ash content is given in wt.% of algal dry matter.

### 7.5.11 Quantification of microalgal compounds after phase separation

After 2-BuOH extraction, the extract was filtered through a 0.22 µm PTFE filter. The extract was transferred to a graduated glass measuring cylinder and the volume was noted. MilliQ water was added to trigger phase separation. The cylinder was wrapped into aluminium foil to protect the sample from light. After 2 h, the volume of the organic and aqueous phase was noted. Samples from both phases were carefully taken with a syringe and were transferred to amber tubes. 0.1 wt.% BHT was added to the organic phase samples to prevent degradation

of pigments and lipids during storage. The samples were stored at  $-20\text{ }^{\circ}\text{C}$  until further analysis. The lipids, pigments, proteins, and carbohydrates were quantified as described in the Sections 7.5.5-7.5.8.

### 7.5.12 Error analysis of experiments

At least two independent biological replicates were performed. The error is given as the standard deviation. The error for quantifying the total amount of lipids and pigments equals the sum of the standard deviations of the single pigments and fatty acids, respectively.

### 7.5.13 Solvents

The following solvents were used for extractions: n-hexane (Merck,  $\geq 98.0\%$ ), ethanol (Merck,  $\geq 99.9\%$ ), ethyl acetate (Merck,  $\geq 99.8\%$ ), 2-butanol (Merck,  $\geq 99.5\%$ ), chloroform (Merck,  $\geq 99.8\%$ ), methanol (Merck,  $\geq 99.8\%$ ), and ethyl formate (Sigma Aldrich,  $\geq 97.0\%$ ), acetone (Sigma Aldrich,  $\geq 99.5\%$ ), propyl vinyl ether (Sigma Aldrich,  $\geq 99.0\%$ ), butyl vinyl ether (Sigma Aldrich,  $\geq 98.0\%$ ), isobutyl vinyl ether (Sigma Aldrich,  $\geq 99.0\%$ ), cyclopentyl methyl ether (Sigma Aldrich,  $\geq 99.9\%$ ), acetonitrile (Roth,  $\geq 99.9\%$ ), pyridine (Sigma Aldrich,  $\geq 99.9\%$ ), n-heptane (Merck,  $\geq 99.0\%$ ).

## 7.6 Summary and conclusions

Wet microalgal biomass is a promising, yet challenging feedstock for biorefineries due to the high moisture content and difficulties in extracting the target compounds from the biomass and separating in the extracted biomass fractions. Despite these challenges, wet extraction of microalgae has enormous potential to bypass the energy-intensive biomass drying step which was identified as an economic bottleneck in microalgal processing. This Chapter leveraged the computational screening approach that was developed in this thesis to systematically select solvents with benign EHS properties for fractionating wet microalgal biomass. The identified solvent candidates were experimentally validated. Based on the computational and experimental results, a novel biomass fractionation process that operates at ambient conditions using benign solvents was developed.

First, the computational solvent screening framework was applied to identify WM, PWM, and NWM solvents, exemplified for a *P. tricornutum* biorefinery. A solvent database containing more than 8000 potential solvents was screened for suitable structural features, desirable  $T_m$  and  $T_b$  ranges, and benign EHS properties. Subsequently, the solubilities of model molecules representing the biomass fractions were predicted using COSMO-RS. The COSMO-RS solubility predictions were in qualitative agreement with solubility data obtained from the literature. Therefore, COSMO-RS is a suitable model for solubility predictions in microalgal biorefineries.

The identified solvent candidates were employed in extraction experiments on wet and dried *P. tricornutum* biomass. In general, the PWM and WM solvents achieved higher lipid and pigment yields than the NWM solvents, despite having a lower lipid solubility. The results suggest, that not only solubility, but also other factors, such as solvent-cell wall and solvent-target organelle interactions significantly influence the yield. PWM and WM solvents outperformed NWM solvents in terms of lipid and pigment yields also on the dry route. Furthermore, the experiments showed that carotenoid and lipid extraction from wet microalgal biomass can result in yields comparable to the dry route if the solvent is carefully selected.

Subsequently, the PWM solvent 2-BuOH was investigated as a candidate for designing a lab-scale biorefinery process. 2-BuOH was of particular interest for biorefinery design as it allows for monophasic conditions during extraction and it is able to form two liquid phases when water beyond the solubility limit is added to separate the extracted compounds. Experiments revealed that the water content  $w_{\text{water,extr}}$  during extraction and differences in extraction rates can be exploited to separate the two most valuable fractions - lipids and carotenoids - during the extraction stage. Based on these findings, a novel biorefinery process that fractionates lipids, carotenoids, carbohydrates and proteins was developed. This fractionation approach employs two sequential extraction stages at different 2-BuOH/water-ratios that extract lipids and pigments highly selectively. Co-extracted compounds, mainly carbohydrates, are subsequently separate from the lipid- and carotenoid-rich streams by inducing phase formation triggered by water addition. Proteins are the main component in the residual biomass. In this manner, the whole biomass is fractionated by applying only two benign solvents at ambient conditions. Further studies are required to optimise the carotenoid and lipid yields and to establish efficient solvent recycling strategies.

In summary, the screening approach successfully identified promising, non-intuitive solvent candidates applicable for wet *P. tricornutum* biomass. A close combination of computational and experimental investigations unveiled promising separation strategies applicable in a biorefinery concept. Moreover, the results break with the common notion that water originating from the moisture represents a barrier that must be overcome. In contrast, the presence of water was taken into account early on in the solvent selection processes. Instead of fighting the „water barrier“, water was exploited as an additional degree of freedom, finally enabling the biomass fractionation.



## 8 | Conclusions and outlook

Biorefineries are drivers of a circular economy within planetary boundaries by converting bio-based, renewable feedstocks into a broad product spectrum, including pharmaceuticals, chemicals, food ingredients, feed, and fuels. In a biorefinery process, the biomass is split up into its macromolecular fractions by a sequence of separation steps, commonly employing organic solvents. Although the solvent choice is a molecular-level decision, it has a process-level impact on the biorefinery, influencing the efficiency of the biomass fractionation and its sustainability. Thus, solvent selection is a key decision that must be made in early stages of the biorefinery design. Experimental solvent selection is time- and resource-consuming. In contrast, computational methods represent a faster, and more systematic approach, thus serving as a high-throughput guide for targeted experimental tests. However, due to the lack of computer-guided solvent selection methods applicable to biorefinery processes, the potential of solvent selection to improve and develop biorefinery processes cannot be fully harnessed.

To fill this gap, in the present thesis a computational solvent screening approach and a solvent design method were specifically developed for solvent selection in biorefineries. The methods were experimentally validated for the two most abundant sources of biomass - lignocellulose and microalgae. The high-throughput screening approach allows to search a database containing more than 8000 potential solvents, including also ILs and DESs. The database is screened for potential solvents with desired structural features,  $T_m$  and  $T_b$  ranges, EHS properties, solubilities, and LLEs. COSMO-RS is applied as the gold-standard method for solubility predictions of representative biomolecules, allowing for a qualitative comparison of the solvent candidates. Although a search space of 8000 molecules appears to be already large, the size of the chemical space is estimated between  $10^{18}$  and  $10^{200}$  organic molecules. The graph-based GA *PSEvolve* developed in this thesis enables the exploration of the solvent space beyond the pre-defined database, tailoring solvent structures towards desired properties. Since, unlike other GAs, *PSEvolve* performs the structural alterations on the molecular graph, *PSEvolve* generates only structurally feasible molecules, efficiently exploring the chemical space. In addition to the design of solvents, *PSEvolve* can potentially be applied in other areas of molecular design (e.g. drug

design), provided that suitable models are available to predict the target properties.

Lignocellulose biorefineries aim at holistic valorisation of the cellulose, hemicellulose and lignin fractions. Lignin represents one of the few abundant natural sources of renewable aromatic hydrocarbon. However, harsh conditions commonly applied in lignocellulose processing lead to lignin condensation, impeding efficient lignin depolymerisation. Commonly, condensed lignins as obtained from Kraft or organosolv pulping, are processed to thermoplastic, nanoparticles, lignin-based films, or fibres. To harness lignins' potential as a source of aromatic carbon for the chemical industry, lignin-first biorefinery approaches, such as AAF, have been developed. In AAF, lignocellulosic biomass is treated with a solvent, acid, and an aldehyde, providing active stabilisation of lignins'  $\beta$ -O-4 bonds through acetal formation on the diol side-chains. The  $\beta$ -O-4 bonds are effectively cleaved *via* hydrogenolysis, thus resulting in near theoretical yields of aromatic monomers. In AAF and many areas of lignin upgrading, lignin dissolution is a crucial step. The solvent screening approach developed in this thesis provided an overview of the solvent landscape considering the cellulose, hemicellulose, and lignin solubilities, as well as EHS properties. For the joint dissolution of the carbohydrate fraction and lignin, the screening proposed several so-far unexplored ILs. Furthermore, azines, sulfoxides, oxazolines, and phosphonates were identified as promising solvent candidates for selective lignin dissolution. The solvents have benign EHS properties and have potential application in dissolution-based lignin upgrading. However, the use of oxazolines and azines is limited for AAF as they are instable towards the process liquor. To expand the solvent search beyond the fixed database, *PSEvolve* was applied for tailored solvent design. *PSEvolve* was combined with a GNN trained on COSMO-RS lignin solubility predictions. Thus, the GNN served as a surrogate model for COSMO-RS, allowing *PSEvolve* to generate optimised solvent structures with high lignin solubility. In addition to the solvent classes identified in the solvent screening, *PSEvolve* designed sulfones, azoles, morpholines, cyclic ethers and ketones, and compounds containing phosphoryl groups as highly promising structures for lignin dissolution. To generate tailor-made solvents for AAF, acid- and aldehyde-reactive groups were excluded from the design space. For AAF, sulfones, sulfoxides, phosphoryl compounds, cyclic and linear ethers were designed. The computationally identified solvents were subsequently tested in solubility experiments. The experimentally measured lignin solubilities ranged between 20 and 60 wt.% ( $T = 85\text{ }^{\circ}\text{C}$ ). While the solubility of NaCl in water is already considered high with 27 wt.% under similar conditions,<sup>495</sup> the designed solvents surpass this value by far, in particular DMSO, pyrazines, azoles, and DEGDME. To examine the relation of the solvent structure and the lignin solubility, an attribution method was applied to the GNN, allowing to quantify the importance of structural features to the obtained predictions. Sulfinyl, sulfonyl, and phosphoryl groups, the presence of O-atoms,



as well as aromatic nitrogen and its neighbouring aromatic bonds are linked to high lignin solubilities, providing a general guideline for solvent selection. Therefore, *PSEvolve* and the GNN do not only facilitate the exploration of the chemical space but also provide insightful details into the structure-solubility relationship. Furthermore, the designed solvents were applied in AAF experiments with birch wood. The generated ethers and butyl sulfone are able to separate the biomass into three separate streams containing cellulose, aldehyde-protected hemicellulose sugars, and aldehyde-stabilised lignin. The stabilised lignin has a high  $\beta$ -O-4 content allowing for efficient conversion to aromatic monomers. Especially DEGDEE, a glycol ether, was identified as a promising solvent and outperformed the toxic and carcinogenic benchmark 1,4-dioxane in terms of DPX yield and EHS properties, with slightly lower lignin monomer yield as a trade-off, while being available at a comparable price.

Microalgae are producers of high-value carotenoids and lipids, as well as proteins and carbohydrates. The development of efficient fractionation strategies for microalgae is still in early development stages since energy-intensive biomass drying and the use of toxic solvents pose considerable challenges. To circumvent the drying step, wet microalgal biomass as obtained after harvest, still having a moisture content of around 85 wt.%, has been frequently proposed as a feedstock. The screening approach was applied to identify solvents applicable for the fractionation of wet microalgal biomass. *P. tricornutum* was selected as a model species due to its balanced biomass composition. Furthermore, the diatom produces Fx, and EPA as value-added products. The economic analysis of potential *P. tricornutum* biorefinery products showed, that EPA and Fx comprise around 90% of the overall economic value of the biomass, highlighting the need for efficient extraction and separation techniques particularly for these fractions. The solvent screening method identified several solvents that outperformed the commonly applied, toxic benchmark solvent n-hexane in terms of lipid and carotenoid yields, as well as EHS properties in experiments. Furthermore, the PWM solvent 2-BuOH was selected for detailed experiments to establish a novel biorefinery approach for wet *P. tricornutum* biomass. 2-BuOH has a large miscibility window with water which was exploited to investigate the effect of different 2-BuOH/water ratios on the yields. Two extraction steps with different 2-BuOH/water ratios enables sequential extraction lipids and carotenoids, effectively separating both high-value fractions. In a later process stage, water addition beyond 2-butanol's miscibility window triggered the formation of two liquid phases, which further separated carbohydrates from the carotenoid and lipid streams. The major component of the residual biomass were proteins. In this manner, the 2-BuOH/water biorefinery process effectively fractionated the biomass at ambient conditions without the need for prior biomass drying.

For both feedstocks, lignocellulose and microalgae, the extraction experiments showed,

that the solubility of the target compounds is not the only factor influencing the yield. The interplay of the target compounds with other surrounding biomass structures (e.g. cell walls or other biomass fractions), should be taken into account as well. In the case of lignocellulose, the experiments showed that solvents with higher lignin solubilities may result in lower lignin monomer yields compared to solvents with lower solubilities. Since lignin is embedded in a lignin-carbohydrate complex, the solvent must not only have sufficient lignin solubility, but should also be able to disrupt the bonds to the carbohydrate fraction to increase its accessibility to lignin. Extraction experiments on *P. tricornutum* biomass showed, that a high target solubility is not the dominating factor to obtain high yields. Solvent interactions with the cell wall, the target organelles, and moisture likely play a significant role for the extraction kinetics.

A paradigm in process optimisation states that optimising only one process unit alone does not lead to the best performance of the overall process. To optimise a biorefinery as a whole system, computational methods that not only optimise the solvent selection as a standalone problem, but rather consider the overall biorefinery process should be developed in the future. To link solvent selection with the process-level, models predicting the yield of the biomass compounds are required. However, a model-based understanding between the effects of solvents on the yields is still lacking and fundamental research investigating the diverse interactions between solvent and feedstock is required. Moreover, the sustainability of the biorefinery processes can only be assessed on the process level. Although evaluating the EHS properties of a solvent is crucial for safe process operation, they do not predict the effect on the planetary boundaries (e.g. GHG emissions or deforestation). In particular, the use of organic solvents contributes to GHG emissions and environmental pollution. These aspects are commonly investigated in a life cycle assessment. However, for the rather unconventional solvents that were identified by the developed computational methods, compound specific data for a life cycle assessment is still lacking. A step towards integrating solvent selection with economic and environmental aspects into the biorefinery design was made by Kopton et al., who developed a multiobjective superstructure optimisation approach based on the results of this thesis.<sup>496</sup> However, more research in this area is required to allow for the development of „benign-by-design“ biorefineries.

Overall, in this thesis, computational methods applicable to diverse types of biomass were developed. These computational approaches were guiding the solvent selection to perform targeted experiments, and were able to identify highly efficient, so-far unexplored solvents for lignocellulose-based and microalgal biorefineries. These methods allowed to find replacements for toxic solvents and to develop novel biomass fractionation approaches.

”

*The walls between art and engineering  
exist only in our minds.*

---

Theo Jansen

“



## A | Phase equilibria

Phase equilibria are determined by the Gibbs free energy  $G(T, p, n)$ . The total differential of  $G$  is defined as<sup>341</sup>

$$dG(n_c, \dots, n_C) = \left(\frac{\partial G}{\partial T}\right)_{p, n_c} dT + \left(\frac{\partial G}{\partial p}\right)_{T, n_c} dp + \sum_{i=0}^C \left(\frac{\partial G}{\partial n_c}\right)_{T, p, n_{i \neq c}} dn_c. \quad (\text{A.1})$$

At thermodynamic equilibrium,  $p$  and temperature  $T$  are constant, and the total differential of  $G$  simplifies to

$$dG(n_c, \dots, n_C) = \sum_{c=0}^C \left(\frac{\partial G}{\partial n_c}\right)_{T, p, n_{i \neq c}} dn_c := \sum_{c=0}^C \mu_c dn_c, \quad (\text{A.2})$$

where  $n_c$  denotes the amount of substance for the chemical species  $c \in \mathbb{N}$ . The term  $\mu_c$  describes the chemical potential. Considering multiple phases  $j$ , the following relationship holds

$$dG(n_c, \dots, n_C) = \sum_{j=0}^J dG^j(n_c^j, \dots, n_C^j) = \sum_{j=0}^J \sum_{c=0}^C \mu_c^j dn_c^j. \quad (\text{A.3})$$

Furthermore, at thermodynamic equilibrium,  $dG = 0$ ,

$$0 = \sum_{j=0}^J \sum_{c=0}^C \mu_c^j dn_c^j. \quad (\text{A.4})$$

Considering two phases  $j = 0$  and  $j = 1$ , it becomes clear that  $dn_c^{j=0} = -dn_c^{j=1}$ , since  $n_c = n_c^{j=0} + n_c^{j=1} = \text{constant}$ . Therefore,

$$0 = \sum_{c=0}^C (\mu_c^{j=0} - \mu_c^{j=1}) dn_c. \quad (\text{A.5})$$

Eq. A.5 is only true, when

$$\mu_c^{j=1} = \mu_c^{j=2}. \quad (\text{A.6})$$

Hence, at phase equilibrium, the chemical potentials in each phase must be equal for the chemical species  $c$ .

## A.1 Solubility of solids in liquids

According to Eq. 4.6, the chemical potentials of the components  $c$  are equal in the liquid phase  $l$  and the solid phase  $s$ . Choosing the pure component in the solid state as a reference leads to

$$\mu_c^{0,s} + RT \ln(\gamma_c^s x_c^s) = \mu_c^{0,l} + RT \ln(\gamma_c^l x_c^l). \quad (\text{A.7})$$

Assuming that the solute  $c$  is a pure component leads to  $\gamma_c^s x_c^s = 1$ . Under that assumption, Eq. A.7 simplifies to

$$x_c^l = \exp\left(\frac{\mu_c^{0,s} - \mu_c^{0,l}}{RT}\right) \frac{1}{\gamma_c^l}. \quad (\text{A.8})$$

At system temperature  $T$ , the pure component  $c$  exists as a solid with a chemical potential  $\mu_c^{0,s}$ , while the chemical potential of the liquid phase  $\mu_c^{0,l}$  can be regarded as that of a subcooled liquid. By heating the solid to its  $T_{m,c}$ , melting it at constant temperature  $T_{m,c}$ , and by subsequently cooling it back to system temperature  $T$  without solidification occurring, the pure component  $c$  transitions to a subcooled liquid. The difference of chemical potentials between the final subcooled state and the initial solid state corresponds to a change in the free enthalpy  $\Delta G_{\text{trans}}^{s \rightarrow l} = \mu_c^{0,l} - \mu_c^{0,s} = \Delta G_{\text{fus}}^{342}$ , leading to

$$x_c^l = \exp\left(\frac{-\Delta G_{\text{fus}}}{RT}\right) \frac{1}{\gamma_c^l}. \quad (\text{A.9})$$

To solve Eq. A.9,  $\Delta G_{\text{fus}}$  and  $\gamma_c^l$  must be determined. For this purpose, COSMO-RS (COSMOthermX19, Biovia 3DS) was applied in this thesis. In COSMO-RS, the chemical potential is obtained as a so-called pseudochemical potential which is derived from the  $\sigma$ -profiles, the chemical potential of a surface segment with screening charge density  $\sigma$ , and a combinatorial contribution (see Ben-Naim for a detailed derivation<sup>497</sup>). The pseudo-chemical potential is defined as

$$\mu_c^* = \mu_c^{*,0} + RT \ln(x_c \gamma_c) - RT \ln x_c = \mu_c^{*,0} + RT \ln \gamma_c. \quad (\text{A.10})$$

where  $\mu_c^{*,0}$  is the reference state for the calculation of  $\mu_c^*$ . The activity coefficient  $\gamma_c^l$  can be obtained from Eq. A.10 and inserted into Eq. A.9 such that the solubility of  $c$  in the solvent can be calculated as

$$\ln x_c^l = \frac{\mu_c^{*,0}(T, p, x_c^l = 1) - \mu_c^{*,l}(T, p, x_c^l) - \Delta G_{\text{fus}}(T)}{RT}. \quad (\text{A.11})$$

In COSMO-RS, the solubility  $x_c^l$  is expressed on a logarithmic scale

$$\log_{10} x_c^l = \frac{\mu_c^{*,0}(T, p, x_c^l = 1) - \mu_c^{*,l}(T, p, x_c^l) - \Delta G_{\text{fus}}(T)}{RT \ln(10)}. \quad (\text{A.12})$$

The free enthalpy of fusion  $\Delta G_{\text{fus}}$  can be calculated from the Schröder-van Laar equation if the enthalpy of fusion  $\Delta H_{\text{fus}}$ , the entropy of fusion  $\Delta S_{\text{fus}}$ , and the heat capacity of fusion  $\Delta C_{p,\text{fus}}$  are known

$$\Delta G_{\text{fus}}(T) = \Delta H_{\text{fus}}(T) - T \Delta S_{\text{fus}}, \quad (\text{A.13})$$

$$= \Delta H_{\text{fus}} \left( 1 - \frac{T}{T_m} \right) - \Delta C_{p,\text{fus}}(T_m - T) + \Delta C_{p,\text{fus}} T \ln \frac{T_m}{T}. \quad (\text{A.14})$$

Here, the heat capacity of fusion  $\Delta C_{p,\text{fus}}$  is considered temperature-independent. For the transition of a solid solute to the subcooled liquid state,  $\Delta G_{\text{fus}} < 0$ . If the solute is liquid, no additional transition to the liquid state has to be considered, hence,  $\Delta G_{\text{fus}} = 0$ . However, in general,  $\Delta H_{\text{fus}}$ ,  $\Delta S_{\text{fus}}$ , and  $\Delta C_{p,\text{fus}}$  are unknown. Therefore,  $\Delta G_{\text{fus}}$  was estimated using a QSPR approach implemented in COSMOthermX19<sup>412</sup>

$$-\Delta G_{\text{fus}}(T = 298 \text{ K}) = c_1 \mu_c(\text{H}_2\text{O}) + c_2 N_{c,\text{ring}} + c_3 V_i + c_4 + [c_5 N_{c,\text{amino}}]. \quad (\text{A.15})$$

where  $c_1$  to  $c_5$  are the parameters of the QSPR model<sup>412</sup>,  $\mu_c(\text{H}_2\text{O})$  is the chemical potential of solute  $c$  in  $\text{H}_2\text{O}$ ,  $N_{c,\text{ring}}$  is the number of ring atoms, and  $V_i$  corresponds to the molecular volume of  $c$ . The Walden's rule, applicable to nonsymmetric organic molecules, was applied to estimate  $\Delta G_{\text{fus}}$  also for other temperatures

$$\Delta S_{\text{fus}} = 0.0135 \text{ kcal mol}^{-1} \text{ K}^{-1}. \quad (\text{A.16})$$

Applying the  $\Delta G_{\text{fus}}$ -estimate and an initial guess for  $\mu_c^{*,l}(T, p, x_c^l)$  to Eq. A.12, a rough approximation for the solubility  $x_c^l$  can be obtained. This approximation was considered sufficient in the computational solvent screening approach (Section 5.2). However, for training the GNN for lignin solubility predictions for the solvent design (Section 6.2), iterative calculations were applied, until the difference of  $x_c^l$  between two consecutive iterations was below  $10^{-5} \log(x_c^l)$  units.

For large chemical potential differences, the accuracy of COSMO-RS is around  $1.5 \text{ kJ mol}^{-1}$  for solubilities in water which corresponds to around 0.27 log units<sup>498</sup>.





## **B | Database**

The database comprises 8011 potential solvent candidates including organic solvent candidates, ILs, and DESs. A list of all solvent candidates is included in the electronic supplements (`database.xlsx`).



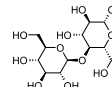
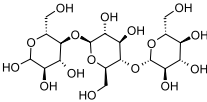
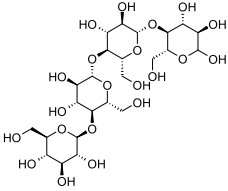
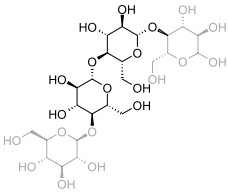
## C | Lignocellulose

### C.1 Solvent screening

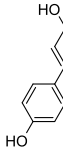
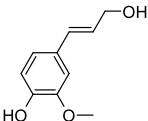
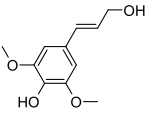
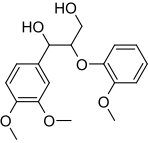
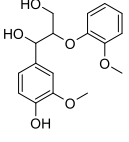
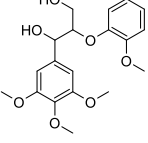
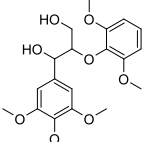
#### C.1.1 Representative biomolecules

The chemical structures of the representative molecules for lignocellulose are summarised in Tab. C.1

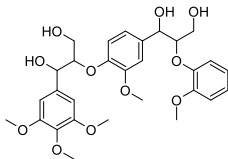
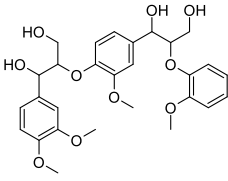
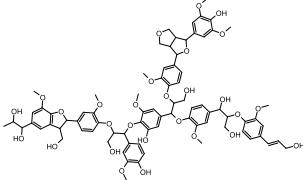
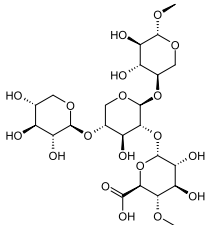
**Tab. C.1** Chemical structures of the representative molecules evaluated for the computational screening.

Fraction	Representative molecule	Structure
Cellulose	Cellobiose	
	Cellotriose	
	Cellotetraose	
	Cellotetraose (capped)	
Continued on next page		

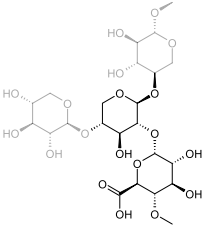
Tab. C.1 – Continued from previous page

Fraction	Representative molecule	Structure
Lignin	<i>p</i> -Coumaryl alcohol	
	Coniferyl alcohol	
	Sinapyl alcohol	
	GG	
	GG phenolic	
	SG	
	SS	
Continued on next page		

Tab. C.1 – Continued from previous page

Fraction	Representative molecule	Structure
	SGG	
	GGG	
	Lignin (1,500 g mol <sup>-1</sup> )	
Hemicellulose	Glucuronoxylan	
Continued on next page		

Tab. C.1 – Continued from previous page

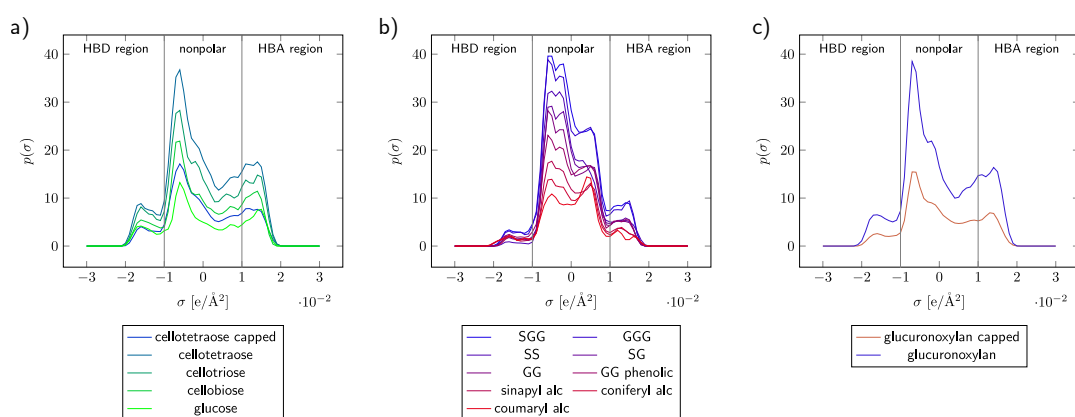
Fraction	Representative molecule	Structure
	Glucuronoxylan (capped)	

### C.1.1.1 $\sigma$ -profile analysis

The  $\sigma$ -profiles of the representative biomolecules give first insights about interactions with potential solvents. In the following, the  $\sigma$ -profiles for all representative molecules are analysed. The  $\sigma$ -profiles of all cellulose representatives show strong HBA- and HBD-behavior, as visible by the peaks at  $0.017$  and  $-0.018$  e  $\text{\AA}^{-2}$ , respectively (Fig. C.1 a). These peaks suggest the possibility for hydrogen bonding within the cellulose-representing structures and with potential solvent candidates. Furthermore, the carbon atoms of the sugar rings cause a peak in the nonpolar region at  $-0.007$  e  $\text{\AA}^{-2}$ . With increasing polymer chain length, the height of all peaks increases, especially in the nonpolar region. As a comparison, glucose which exhibits a higher water-solubility than all considered cellulose-representing molecules was included in the  $\sigma$ -profile analysis. Glucose showed the lowest peak height in the nonpolar region among the cellulose representatives. Hence, a higher peak in the nonpolar region indicates decreasing water-solubility for the cellulose oligomers. The  $\sigma$ -profile of the capped cellotetraose molecule, which only considers cellobiose as repeating unit, resembles that of glucose in the HBA/HBD region, and cellobiose in the neutral region. Overall, a solvent capable of cellulose dissolution is likely to engage in hydrogen bonding but also to have domains with neutral screening charges.

All lignin representatives (Fig. C.1 b) show two distinct peaks in the nonpolar region at  $-0.005$  and  $-0.0025$  e  $\text{\AA}^{-2}$  caused by the slightly electropositive hydrogen atoms of the aromatic rings and its carbon atoms, respectively. The  $\pi$ -face of the aromatic ring causes a peak at  $0.008$  e  $\text{\AA}^{-2}$ , which is consequently the highest for the trimeric lignin representatives. The oxygen atoms of the  $\beta$ -O-4 bonds and the free hydroxyl groups cause a peak in the HBA region. Additionally, a small peak in the HBD region is visible, originating from the

hydrogen atoms of the free hydroxyl groups. Therefore, inter- and intramolecular hydrogen bonding is possible, but with less intensity compared to the cellulose and hemicellulose representatives. Compared to the  $\sigma$ -profiles of the cellulose representatives, the peaks in the nonpolar region of the lignin-representing molecules are broader. This peak behaviour indicates different dissolution mechanisms and the possibility for solvent-based separation of lignin from cellulose and hemicellulose. The  $\sigma$ -profiles of the hemicellulose representatives (Fig. C.1 c) have high similarities to that of cellulose leading potential hydrogen bond formation.



**Fig. C.1**  $\sigma$ -profiles for representative molecules of a) cellulose, b) lignin, and c) hemicellulose. Reprinted from<sup>395</sup> with permission from Elsevier.

### C.1.1.2 Correlation with experimental data

To evaluate the performance of COSMO-RS solubility predictions using the described representative structures, experimental data collected from literature and the predicted solubilities were compared. The accuracy of the predictions was analysed by linear regression. In the analysis, the hemicellulose fraction was not considered as this fraction depolymerises and readily dissolves under commonly applied process conditions (see Section 2.2).

Sameni et al. reported the solubilities of four different types of lignin: alkaline lignin isolated from an industrial mix of hardwood and non-wood species, Kraft lignin from eucalyptus, Indulin AT (commercial softwood Kraft lignin), and protobind (commercial non-wood soda lignin) in different solvents ( $T = 25\text{ }^{\circ}\text{C}$ )<sup>101</sup>. The chemical structure of lignin the mentioned is highly dependent the biological source and the applied extraction conditions. The studied lignins showed distinct structural features, potentially leading to a different solubility characteristics<sup>101</sup>. The data points for the lignin solubilities in DMSO and pyridine were removed from the dataset, as lignin saturation was not reached

in the experiments. For the remaining nine data points, the COSMO-RS predicted and the experimental solubilities were converted to wt.% [ $g_{\text{solute}} (g_{\text{solute}} + g_{\text{solvent}})^{-1}$ ] and linear regression was performed. Three compilations of COSMO-RS predictions were used: averaged COSMO-RS solubilities for monolignols, dimers, trimers; an average of all monolignols, dimers and trimers altogether; and a single conformer of the  $1,500 \text{ g mol}^{-1}$  lignin fragment.

The most robust and overall highest correlations were observed for the averaged solubilities of monolignols, dimers and trimers denoted as  $\text{average}_{\text{m,d,t}}$  (see correlation coefficients  $R^2$  in Tab.C.2). Correlation coefficients varied strongly among the wood species and the representative molecules. For alkaline lignin from hardwood/non-wood species, correlation coefficients of  $R^2$  of about 0.7 were obtained for trimers, the  $1,500 \text{ g mol}^{-1}$  lignin fragment, and  $\text{average}_{\text{m,d,t}}$ . Interestingly, the correlation coefficients for monolignols was higher than for the dimers, which could be explained by the polydispersity of the alkaline lignin<sup>499</sup>. It must also be noted, that impurities in form of cellulosic materials were observed in this type of lignin, potentially impacting the solubility of the samples<sup>499</sup>. For The Eucalyptus Kraft lignin and protobind lignin, higher correlations were observed for monomers than for trimers and the  $1,500 \text{ g mol}^{-1}$  lignin fragment. This behaviour might be attributed to the low molecular weight of these technical lignins compared to the other samples. Especially for the Eucalyptus Kraft lignin, lignin-like impurities with low molecular weight were observed<sup>499</sup>, which might explain the high correlation coefficient with the monomers. In general, the correlation coefficients for trimers was similar to or higher than for the  $1,500 \text{ g mol}^{-1}$  lignin fragment, indicating that such large representative structures do not necessarily lead to more reliable results. Also from a thermodynamic perspective, COSMO-RS predictions with larger structures are not automatically more accurate. Lignin is an amorphous polymer which changes its conformation based on the solvent environment. Especially in unfavourable solvents, lignin forms globules in order to minimize solvent exposure<sup>442</sup>. As a result, only certain parts of the surface area of such a fragment, and, consequently, respective parts of the  $\sigma$ -surface would be accessible to the solvent. Additionally, the  $\sigma$ -surface itself would be influenced by the described folding behaviour. The polymer folding is not captured during the QC calculations performed for the screening, since the geometry optimisation is performed in a perfect conductor and not in the solvent environment. Based on this analysis, the average solubilities of monomers, dimers, and trimers were used in this study. However, it should be noted that the analysis only focused on technical lignins extracted under alkaline conditions. Under such condition, the native lignin structure undergoes a series of depolymerisation and repolymerisation reactions, causing lignin condensation (see Section 2.1.3). Furthermore, only 9 datapoints were considered. Hence, this preliminary analysis is not sufficient to validate the accuracy



of COSMO-RS solubility predictions using the described lignin-representating structures for native lignin, or lignins isolated under acidic conditions. A detailed experimental validation follows in Section 6.3.

**Tab. C.2** Correlation coefficients  $R^2$  from linear regressions between solubilities for lignin from different sources and COSMO-RS predictions for lignin monomers, dimers, trimers, the  $1,500 \text{ g mol}^{-1}$  lignin fragment and the average of all monomers, dimers, and trimers. Experimental data was taken from<sup>101</sup>.

Lignin type	Mono-lignols	Dimers	Trimers	Lignin fragment	Average <sub>m,d,t</sub>
Alkaline lignin (hardwood/non-wood)	0.25	0.07	0.67	0.68	0.71
Kraft lignin (Eucalyptus)	0.69	0.09	0.38	0.24	0.69
Indulin AT	0.06	0.56	0.77	0.55	0.47
Protobind	0.66	0.78	0.50	0.33	0.78

To date, only few solvents with the ability to dissolve cellulose are known, e.g. several ILs<sup>435,500</sup> and NMMO which is applied in the lyocell process<sup>438</sup>. To investigate the suitability of the different cellulose representatives, COSMO-RS solubility predictions and experimental solubility data of cellulose in ILs taken from literature<sup>435,500</sup> were compared. Similar to the analysis of the lignin representatives, also here the experimental solubility data and the COSMO-RS solubility predictions were converted to wt.% and performed linear regression. All ILs containing  $\text{Cl}^-$ -ions were excluded from the analysis, as there are known to be systematic deviations in COSMO-RS predictions for this class of anions<sup>306,307</sup>. As a result, the capped cellotetraose molecule led to the highest correlation coefficient ( $R^2 = 0.71$ ) which was remarkably higher than that of the other cellulose representing structures ( $R^2 \leq 0.37$ ). Therefore, in this work, the capped cellotetraose molecules was used as a representative molecule for the cellulose fraction. In analogy with the cellulose fraction, hemicellulose is represented by the capped glucuronoxylan molecule.

### C.1.2 Solvent ranking

Lists of the most promising solvents for each process objective is provided in the electronic supplements (lignocellulose\_screening.zip).

**Tab. C.3** Hyperparameter test for *PSEvolve*. Selected values are marked as bold numbers.

Parameter	Value	Fittest candidate ( $\log_{10}(\chi_L)$ )	Mean $\log_{10}(\chi_L)$
Population size	10	1-methyl imidazole (-0.45)	-2.5
	100	Dimethyl methyl phosphonate (-0.38)	-0.5
	<b>1000</b>	DMSO (-0.35)	-1.6
	5000	DMSO (-0.35)	-3
Mutation rate	0.05	5-methylpyrimidine (-0.39)	-2.6
	<b>0.1</b>	DMSO (-0.35)	-1.6
	0.5	DMSO (-0.35)	-1.7
	0.9	DMSO (-0.35)	-1.7
Number of parents/ children	10/20	1-methyl imidazole (-0.45)	-2.5
	20/40	5-methyl-pyrimidine (-0.39)	-1.8
	<b>50/100</b>	DMSO (-0.35)	-1.6
	100/200	5-methyl-pyrimidine (-0.39)	-1.5

## C.2 Solvent design

Lists of designed solvents for lignin dissolution and AAF is provided in the electronic supplements (`lignin_solvent_design.zip`).

### C.2.1 Hyperparameters of *PSEvolve*

*PSEvolve* was tested for several hyperparameter setups. The test runs were stopped after 100 generations. For the parameter test, the standard parameters were kept constant and each parameter listed in Tab. C.3 were independently varied. Standard parameters were mutation rate = 0.1, number of parents = 50, number of children = 100, population size = 1000. In this manner, the influence of the parameters on the mean lignin solubility in the population and the fittest designed molecules was systematically studied. The standard parameters which were used throughout the study (bold numbers) provided the best performance in terms of identifying the fittest molecules and mean lignin solubility of the population.

All parameters applied in *PSEvolve* are summarised in Tab. C.4. In the mutation step, random functional groups were added to the molecule. The pool of functional groups used for the design stage is provided in the electronic supplements (`lignin_solvent_design.zip`).

**Tab. C.4** Hyperparameters and constraints applied in the solvent design for high lignin solubilities and AAF using *PSEvolve*.

Parameter	Description	Value	Unit
Population size	Constant value describing the size of the population	1000	Number of molecules
Molecule types in the start population	n-Hexane	-	-
Fitness	$x_L$	Predicted by GNN	-
Number of parent molecules	Number of parent molecules selected for cross-over	50	Number of molecules
Number of offspring molecules	Number of offspring molecules generated by cross-over	100	Number of molecules
Mutation rate	Constant value describing the probability of occurring mutations	0.1	-
Maximum $M_{\text{solv}}$	Upper bound for molar weight	200	$\text{g mol}^{-1}$
SAS	Synthetic accessibility score <sup>441</sup> ranging from 1 (easily synthesisable) to 10 (difficult to synthesise)	3.5	-
Atom type constraints	Only C-, H-, O-, N-, S-atoms used	-	-
Functional group constraints	For AAF only: acid- and aldehyde- instable groups were eliminated (primary and secondary amines, aldehydes, aromatic N-heterocycles, isocyanates, amides, esters, hydrazides)	-	-
End criterium	Stop after a certain number of generations	1000	Number of generations

### C.2.1.1 GNN architecture

The solvent graph  $G = (V, E)$  consists of a set of nodes  $V$  connected by a set of edges  $E$ , representing the corresponding atoms and bonds, respectively. The atoms and bonds

were attributed with several features (Tab. C.5). The features constitute the vectorial representation of the corresponding nodes and edges within the graph. For each molecule, a matrix of atom features  $\mathbf{A}^{N_{\text{atoms}} \times N_{\text{atom features}}} \in [0, 1]$  and a matrix of bond features  $\mathbf{B}^{N_{\text{bonds}} \times N_{\text{bond features}}} \in [0, 1]$  was defined. The connectivity matrix  $\mathbf{C}^{2 \times 2 N_{\text{bonds}}} \in [0, 1]$  captures the indices of the source and receiver nodes and describes the connectivity between atoms and bonds. The features summarised in Tab. C.5 were selected to distinguish fundamental differences between atoms and bonds within a given molecule. One-hot-encoding was used to encode the atomic and bond information into fixed-size vectors for all molecules according to the dimensions shown in Tab. C.5. The connectivity matrix and all features were calculated using rdkit.

*PyTorch geometric* (version 2.3.1) and *PyTorch* (version 1.10.2) were used for the GNN setup. The model consists of 3 message passing layers operating with a hidden-dimension of 50. The *NNConv* architecture as proposed by Gilmer et al.<sup>501</sup> was used. A single hidden-layer neural network with dimension 64 and the ReLU activation function was used as the edge-transforming function. The batch normalisation proposed by Ioffe et al.<sup>502</sup> was applied after each message passing layer to enhance the training of the model. To update the node embeddings, the Leaky ReLU activation was used after the first and second message passing layers. The molecular fingerprint was obtained by using the max global pooling function on the final updated graph. A multi-layer perceptron was subsequently applied to regress the final solubility prediction from the molecular fingerprint. The multi-layer perceptron contains 2-hidden layers with dimensions 50 and 25. A dropout ratio of 0.1 was used in the message-passing layers and the final multi-layer perceptron to prevent overfitting. The model was trained „end-to-end“ from the molecular graph to the COSMO-RS lignin solubility using the AdamW optimizer with a learning rate of 0.001 and batches of 32 graphs. The mean squared error was used as the loss function. The training was performed for 100 epochs. A learning rate scheduler was used to decrease the learning rate by a factor of 0.8 using a patience of three epochs. The training was performed independently on 5 different train/validation splits resulting in 5 independent models. The final predictions were made by the ensemble of these 5 models by averaging their individual predictions. All hyperparameters were determined based on ablation studies assessed on the validation set.

### C.2.1.2 GNN hyperparameters

The GNN hyperparameters (number of message-passing layers, dropout probability, hidden-size of the message-passing layers, learning rate) were determined by comparing the model accuracy for different parameter sets. Tab. C.6 shows the values for several combinations of hyperparameters. The selected parameters (bold numbers) provide the lowest MAE.

**Tab. C.5** Atom and bond features considered in the GNN.

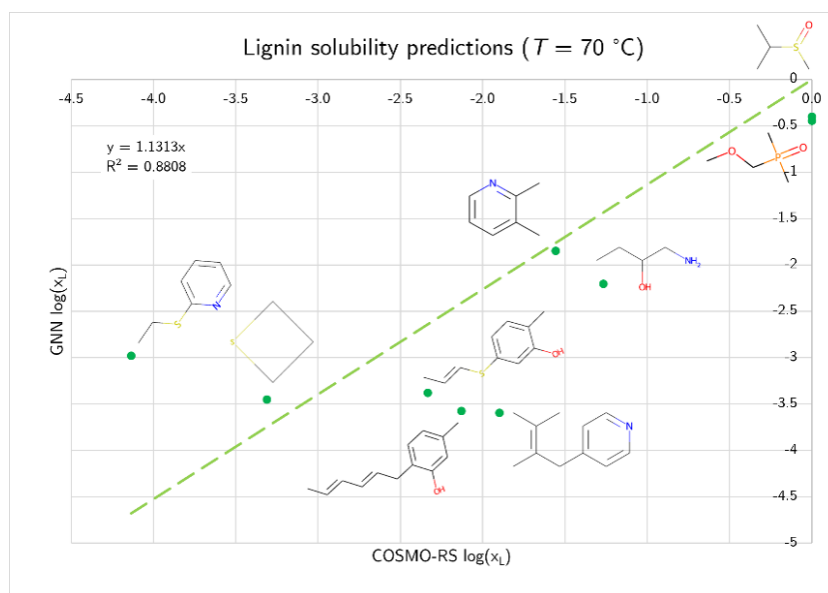
Feature type	Feature	Description	Dimension
Atom features	Atom type	(C, O, N, Cl, F, S, Si, Br, P, Se, I, B, As, Ge, Al)	15
	Ring	Is it in ring?	1
	Aromatic	Is it aromatic?	1
	Hybridisation	(sp, sp <sup>2</sup> , sp <sup>3</sup> , sp <sup>3</sup> d)	4
	Bonds	Number of bonds attached (0,1,2,3,4)	5
	Charge	Formal charge (0,1,-1,3)	4
	H's attached	Number of bonded H's (0,1,2,3)	4
Bond features	Bond type	(Single, double, triple, aromatic)	4
	Conjugated	Is it conjugated?	1
	Ring	Is it in ring?	1

**Tab. C.6** Parameter test for GNN solubility predictions. The selected parameters are marked with bold numbers.

Layers	Dropout	Hidden size	Learning rate	Validation MAE
2	0.1	50	0.001	0.294
3	0.1	25	0.001	0.307
3	0.1	50	0.0001	0.304
3	0.1	50	0.01	0.284
3	0.1	100	0.001	0.294
3	0.5	50	0.001	0.416
3	0.25	50	0.001	0.298
4	0.1	50	0.001	0.311
<b>3</b>	<b>0.1</b>	<b>50</b>	<b>0.001</b>	<b>0.281</b>

### C.2.1.3 Applicability range of the GNN

The trained GNN is capable to estimate lignin solubilities for solvents with structural similarity to the chemical classes contained in the training set at a temperature of  $T = 70$  celsius. The solubility predictions should not be treated as accurate absolute solubility values, but should rather be compared relative to each other. In this way, the GNN allows to qualitatively compare different solvents in their ability to dissolve lignin. For solubility predictions at different temperatures or using different solutes, a new model could be trained in a similar manner as presented within this thesis. To validate the applicability of the GNN for lignin solubility predictions for solvents outside the training and test set for the present study, several solvents that were designed by PSEvolve under the objective to maximise lignin solubility were selected. These solvents were chosen in a way that they reflect a broad solubility range (e.g. solvents with long aliphatic chains in the lower solubility range vs. the sulfoxide/phosphonate/pyridine-related structures in the higher solubility range) and a broad structural diversity. For these solvents, additional QM and COSMO-RS solubility predictions using the GGG-trimer as a model molecule for lignin were performed  $T = 70$  celsius according to the computational methodology described in the Sections 5.2.3.1 and 5.2.5. The GNN and COSMO-RS predicted solubilities for the newly designed molecules is very well in agreement ( $R^2 = 0.88$ ) which is comparable to that of the original test set ( $R^2 = 0.89$ , see Fig. 6.5). Therefore, the GNN is capable to predict lignin solubilities of solvent candidates designed in this study with an accuracy comparable to that of the original test set.



**Fig. C.2** Comparison of COSMO-RS and GNN lignin solubility predictions for solvent structures not included in the training set ( $T = 70$  °C). Reprinted from König-Mattern et al.<sup>395</sup> with permission from Elsevier.

**Tab. C.7** COSMO-RS lignin solubility predictions for selected solvents at  $T = 25\text{ }^{\circ}\text{C}$  and  $70\text{ }^{\circ}\text{C}$ 

Solvent	$\log_{10}(x_L)$	
	$T = 25\text{ }^{\circ}\text{C}$	$T = 70\text{ }^{\circ}\text{C}$
DMSO	0.0	0
Pyridine	-1.09	-1.02
Thiazole	-1.27	-1.12
1,4-Dioxane	-1.71	-1.35
Water	-6.5	-6.16
Di-n-butylether	-7.79	-6.32
n-Heptane	-9.12	-6.71

Additionally, the effect of the temperature on COSMO-RS lignin solubility predictions was studied. Since the COSMO-RS predictions are the foundation for the GNN solubility predictions, studying the temperature effects allows to assess possibilities of the solvent design framework for extrapolation to lower temperatures. COSMO-RS solubility predictions for the solvents summarised in Tab. C.7 at a temperature of  $T = 25\text{ }^{\circ}\text{C}$  were compared to COSMO-RS solubility predictions at  $T = 70\text{ }^{\circ}\text{C}$ . As expected, the solubilities decreased with decreasing temperatures. The ranking of solvents according to the highest lignin solubility remains the same, indicating that the GNN trained on lignin solubilities at  $T = 70\text{ }^{\circ}\text{C}$  can be applied to design solvents for lignin dissolution at lower temperatures as well.

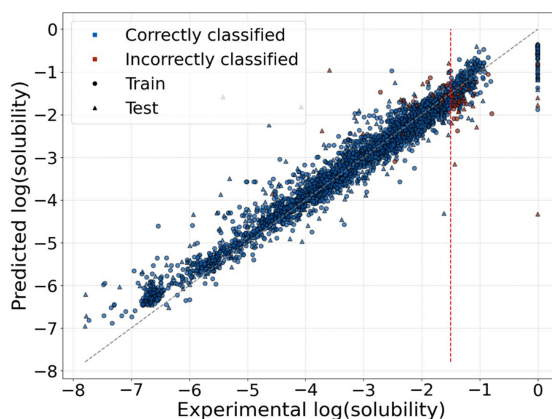
#### C.2.1.4 GNN explainability

To identify structural patterns in the solvent graph with the highest contribution to the predicted lignin solubilities, the IGs attribution method<sup>380,444</sup> was applied. The IG method is only applicable to a binary classification GNN. Therefore, a second GNN (referred to as „classification“) was trained to perform the classification of the molecules into „promising solvent“ and „non-promising solvent“. The binary classification threshold was set to  $\log_{10}(x_L) = -1.5$  according to the COSMO-RS predicted lignin solubilities. The same train and test splits as for the regression task were used. For developing the classification GNN, the chosen message-passing scheme corresponds to the *PyTorch geometric* implementation<sup>503</sup>. Two message-passing layers were used with a hidden dimension of 32. Then, a global sum pooling layer was used to obtain the molecular fingerprint. Finally, a multi-layer

perceptron of 2-hidden layers with a hidden size of 32 and the ReLU activation function was used to map the fingerprints to the binary solubility classes. The final two neurons of the multi-layer perceptron used the log-softmax activation function. Drop-out with a probability of 0.5 was used after the first hidden layer of the multi-layer perceptron to prevent overfitting. The class with the predicted probability of belonging was selected as the predicted solvent class. The classification GNN was trained using the negative log-likelihood as the loss function and the Adam optimizer. The training ran for 100 epochs with a learning rate of 0.001 and a batch size of 128 graphs. The overall accuracy was 96.9% and 82.17% of the solvents classified as “promising” (high lignin solubility) were correctly classified as such (true positive values), see Fig. C.3. These results highlight the relative agreement between the predictions of the regression GNN and the classification GNN. Discrepancies are mostly accumulated around the threshold of  $\log_{10}(x_L) = -1.5$ .

The classification GNN was coupled with the IG method to highlight the structural features of each input graph that were the most relevant for classifying the solvent as “promising” or “not-promising”. For this, the IG implementation from Captum52 (version 0.6.0) was used. The corresponding solvent graph with all node features equal to zero was used as a baseline for IG. The attribution scores were normalised for each graph to values between 0 and 1. These scores reflect the least and most important substructures of the graph for predicting the corresponding class, respectively. The default Gauss-Legendre quadrature rule as implemented in Captum was used for computing the integral of the gradients. It is important to highlight that the intention of gathering explainability scores by using IG and the classification GNN is to support or guide the scientist in the overall explainability and interpretation tasks. The attribution techniques should not be used as the solely ground truth for scientific discovery. Therefore, the explainability scores should be taken as an extra tool to support experimental discovery rather than as the scientific discovery *per se*.





**Fig. C.3** Parity plot of the classification GNN. Classes were „promising“ and „non-promising“ solvent, based on a threshold lignin solubility of  $\log(x_L) = -1.5$ .

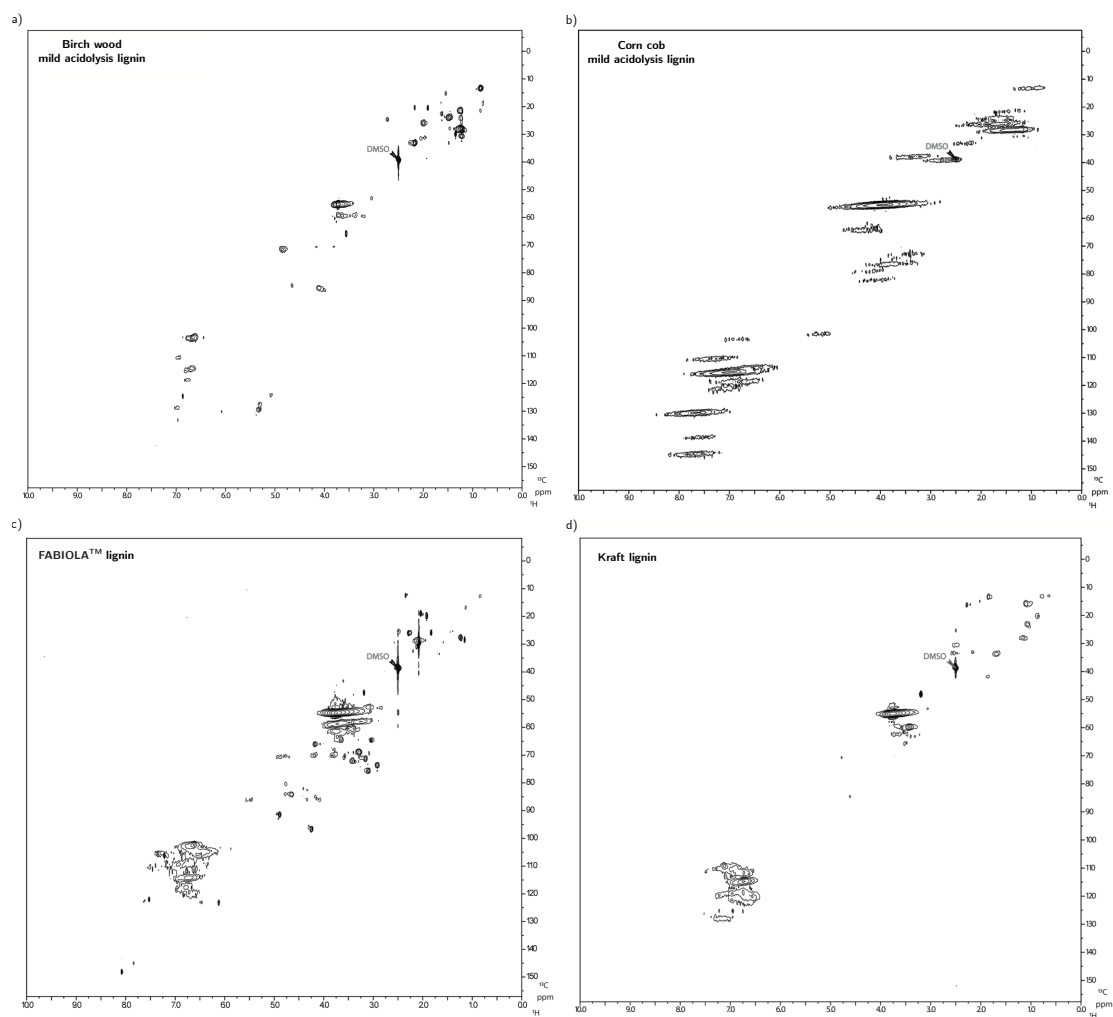
## C.3 Experimental

### C.3.1 2D HSQC NMR

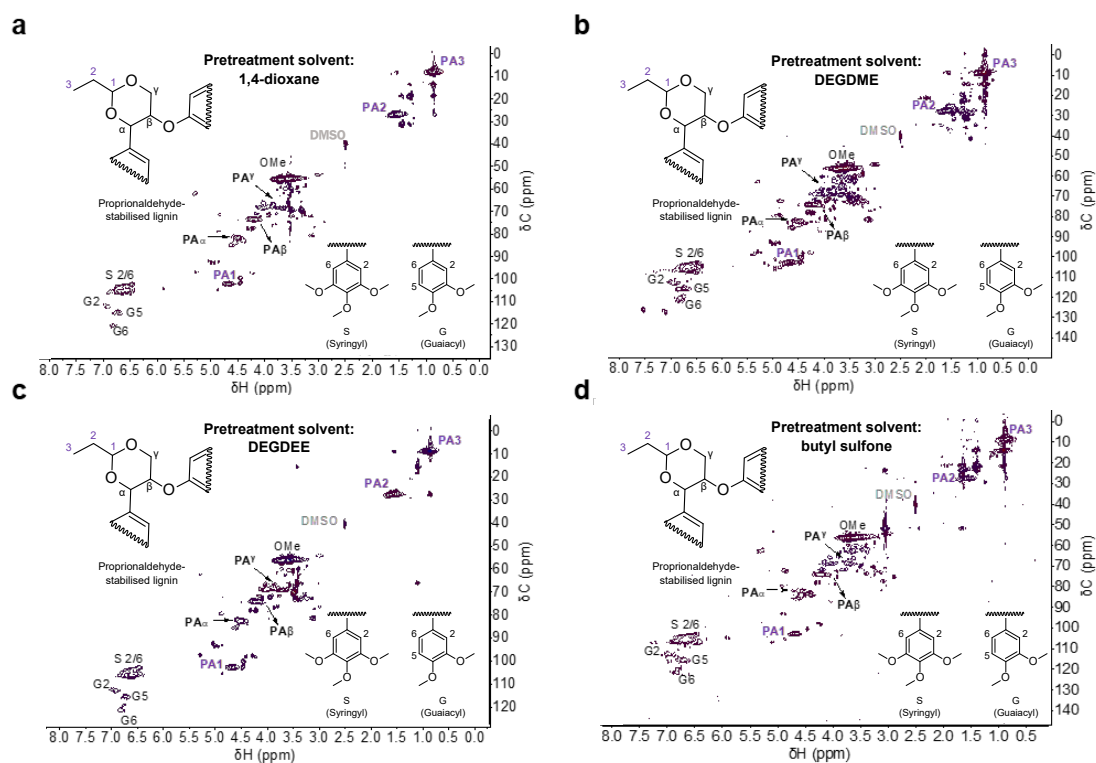
50 mg of the lignin samples was added to NMR tubes and 0.7 ml of DMSO-d was added to dissolve the lignin. Lignin dissolution was promoted by a sonication and vortex mixing. The NMR spectra were acquired on a Bruker Avance III 500 MHz spectrometer. The following parameters were applied in the HSQC-NMR analyses:

- NS (number of scans): 32
- D1 (delays): 1.5 s
- O1P (transmitter frequency offset): 4.700 p.p.m.
- SW (spectral width): 13.1536 p.p.m.
- DS (dummy scans): 32

The resulting scans of the lignin samples used for the solubility measurements in Section 6.3.2 are presented in Fig C.4. The aldehyde-stabilised lignin obtained from AAF using various solvents (Section 6.3.3) are presented in Fig. C.5.



**Fig. C.4** 2D-HSQC-NMR spectra of different lignin types used in this work. a) Birch MAL, b) corn cob MAL, c) FABIOLA™lignin from beechwood, and d) Kraft lignin from softwood species obtained from Berner Fachhochschule.



**Fig. C.5** 2D-HSQC-NMR spectra of aldehyde-stabilised lignin obtained from AAF using the benchmark solvent a) 1,4-dioxane, and the designed solvents b) diglyme (DEGDME), c) DEGDEE, and d) butyl sulfone.

### C.3.2 Lignin solubility measurements

Numerical values for the lignin solubility experiments and the GC FID response factors for each tested solvent are presented in Tab. C.8.

### C.3.3 Biomass composition of Birch wood

The composition of the birch wood applied in AAF experiments (Section 6.3.3) was determined according to Talebi Amiri et al.<sup>155</sup> and is provided in Tab. C.9.

### C.3.4 DPX and lignin monomer yields after AAF treatment

Numerical values for lignin monomer and DPX yields are provided in Tab. C.10.

**Tab. C.8** Experimental solubility data for Kraft, FABIOLA, as well as birch and corn cob MAL obtained by the GC method. Values in brackets correspond to additional datapoints obtained by the evaporation method. The asterisk indicates that the solubility limit in the corresponding solvent was reached. N/M corresponds to „not measured“. Solvents for which saturated lignin solutions were obtained are marked by an asterisk (\*).

Category	Solvent name	Lignin solubility [wt%]				GC response factor
		Kraft lignin	FABIOLA lignin	Corn cob MAL	Birch wood MAL	
Sulfoxides	DMSO	60	48	29	34	0.831
	DESO	19	21	33	23	0.689
Phosphonates	DMMP	23	20	20	27	1.372
	DEMP	20	20	33	25	0.708
	DEEP	20	20	21	26	0.691
Azines	4-Methyl-pyrimidine	36	34	28	22	0.565
	4-Pyrrolidinopyridine	22	19	19	26	0.404
	Pyrazine	21	18	19	24	0.584
	4-Methoxy-pyridine	24	19	20	26	0.547
	4-Piperidinopyridine	21	20	19	26	0.39
	Pyridine-n-oxide	30	25	29	N/M	0.466
	2-Picoline-n-oxide	49	32	29	N/M	0.529
2,5-Dimethyl-pyrazine	48	49	29	N/M	0.441	
Azoles	2-Methyl-2-oxazoline	43	34	29	21	0.735
	1-Methyl-imidazole	22	19	18	26	0.608
	5-Bromo-1-methyl-1H-imidazole	22	19	21	20	1.253
	Pyrazole	20	21	20	25	0.743
	Thiazole	49	49	28	N/M	0.773
	1,3-Aminopropyl-imidazole	31	34	19	N/M	0.623
Isoxazole	49	50	28	N/M	0.59	
Morpholine	4-(2-Hydroxyethyl)-morpholine	33	33	18	N/M	0.707
Sulfones	Butyl sulfone	31	30	30	N/M	0.451
	DMM-sulfonamide	21	22	17	25	1.036
Ethers	18-Crown-6-ether	22	22	32	27	0.703
	1,4-Dioxane	20	21	19	19	0.871
	DEGDME	51	50	25	29	0.759
	DEGDEE*	9	6	8	7	0.834
	2-MeTHF*	16 (20)	14 (17)	9 (9)	11 (14)	0.473
Di-n-butyl ether*	2 (5)	2 (1)	4 (3)	6 (7)	0.405	
Hydrocarbons	n-Heptane*	0 (1)	0 (0)	1 (2)	2 (6)	1.738

**Tab. C.9** Biomass composition of the birch wood used for AAF experiments.

Feedstock	Ash	Hydration	Extractives	Klason lignin	Acid-soluble lignin	Glucan	Xylan	Galactan	Arabinan	Mannan	Acetyl	Total
Birch wood	0.15%	6.09%	3.27%	17.94%	2.81%	32.16%	17.81	2.01%	0.56%	0.26%	5.89%	88.95%

**Tab. C.10** Experimental data of DPX and lignin monomer yields obtained from AAF experiments. N/M corresponds to „not measured“.

Solvent	Yield [wt. % <sub>raw biomass</sub> ]		
	Pulp	Lignin monomers	DPX
1,4-Dioxane	38.9	7.8	21.3
DEGDME	53.6	3.9	14.9
DEGDDE	37.4	4.7	22.3
18-Crown-6 ether	90.3	1.8	6.3
Butyl sulfone	42.2	4.9	21.5
DMM-sulfonamide	39.4	N/M	13.6
DEEP	69.4	N/M	< 0.5
DEMP	59.3	N/M	< 0.5
DMMP	65.3	N/M	< 0.5
DMSO	81.4	N/M	0.8



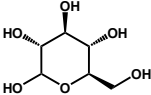
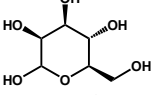
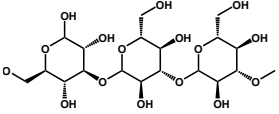
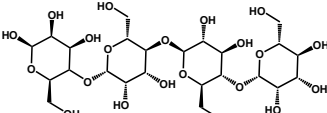
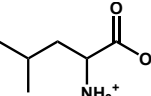
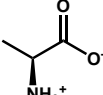
## D | Microalgae

### D.1 Solvent screening

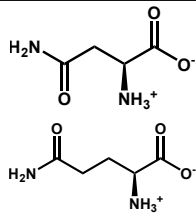
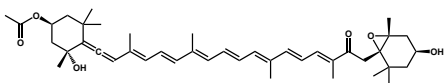
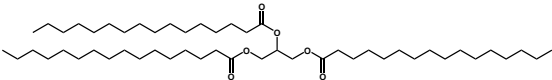
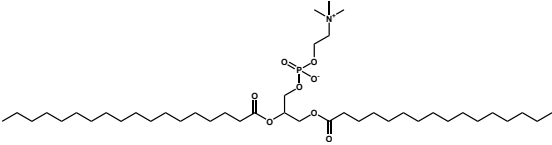
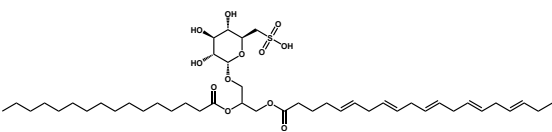
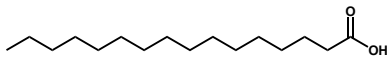
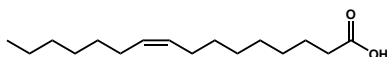
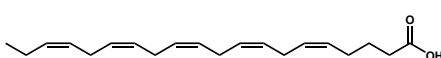
#### D.1.1 Selection of representative molecules for *P. tricornutum*

Representative molecules for *P. tricornutum* used for COSMO-RS solubility and partition coefficient predictions are provided in Tab. D.1.

**Tab. D.1** Representative molecules for *P. tricornutum* and their molecular structures.

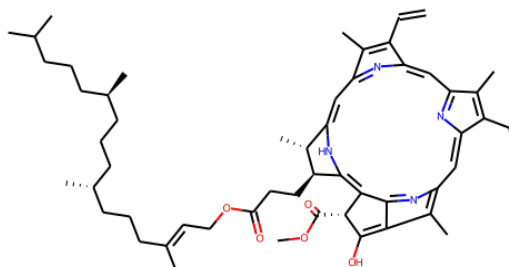
Fraction	Representative molecules	Structure
Carbohydrates	D-glucose	
	D-mannose	
	Laminarin	
	Glucomannan	
Proteins	L-leucine	
	L-alanine	
Continued on next page		

Tab. D.1 – Continued from previous page

	L-asparagine  L-glutamine	
Carotenoids	Fucoxanthin (Fx)	
Neutral lipids	Glycerol tripalmitate	
Polar lipids	Phosphatidyl choline (PC)	
	Sulfoquinovosyl diacyl glycerol (SQDG)	
	Palmitic acid (C16:0)	
	Palmitoleic acid (C16:1)	
	Eicosapentaenoic acid (EPA)	

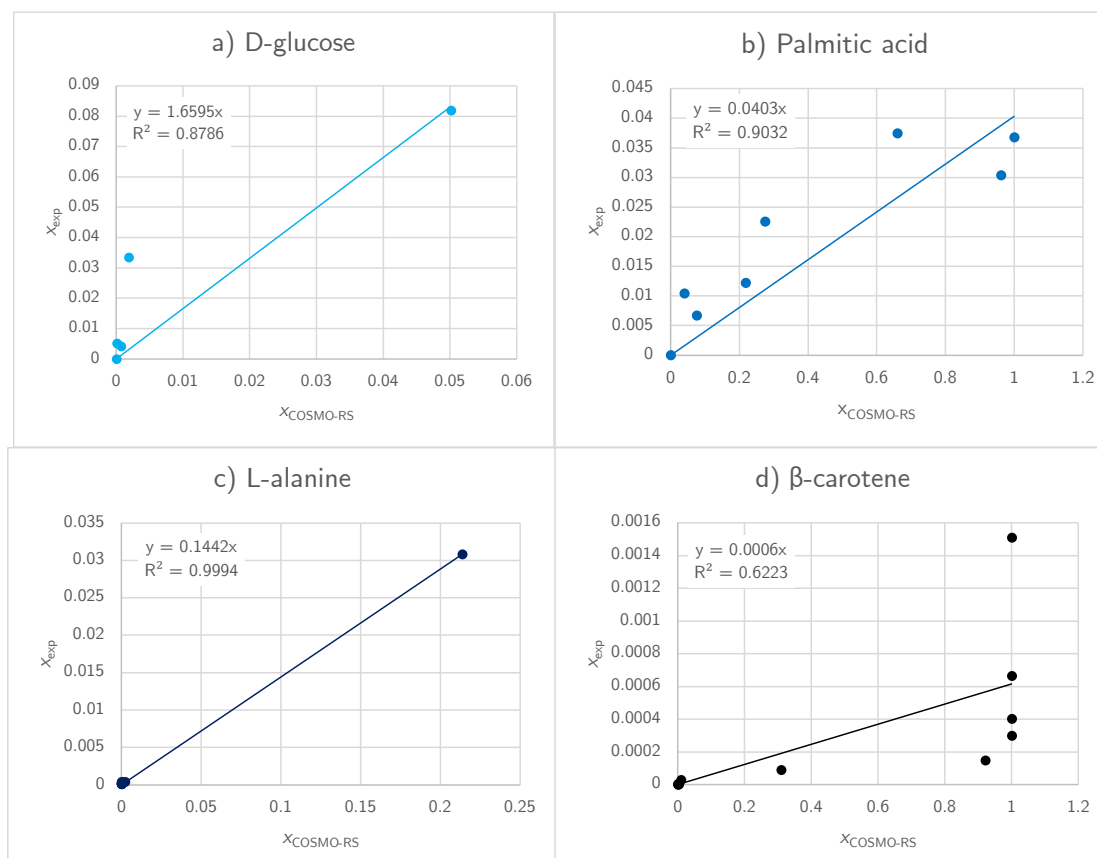
For 2-BuoH solubility predictions (Section 7.4), Fx was used as carotenoid representative, tripalmitin as a NL-representing molecule, lecithin was used to represent the PLs, and pheophytin a for chlorophylls. The chemical structure of pheophytin a is shown in Fig. D.1.





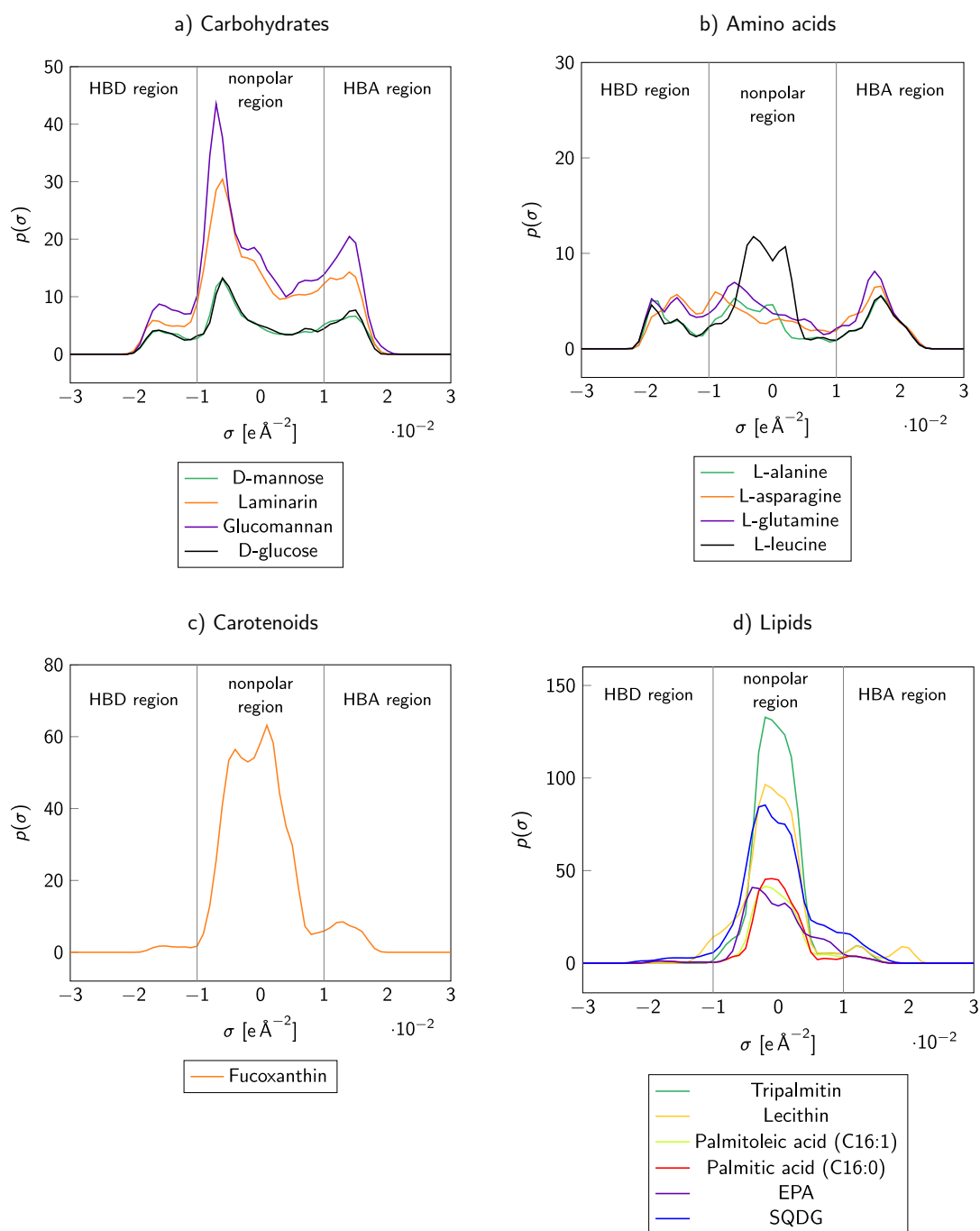
**Fig. D.1** Molecular structure of pheophytin a, the chlorophyll-representing molecule used in the COSMO-RS solubility predictions for 2-BuOH/water.

From each biomass fraction, one biomolecule was selected for a comparison between COSMO-RS solubility predictions and experimentally determined solubilities taken from literature<sup>504–506</sup> (Fig. D.2). Thus, five datapoints for D-glucose, eight datapoints for palmitic acid, 11 datapoints for L-alanine, and 11 datapoints for  $\beta$ -carotene were obtained. In general, COSMO-RS predictions and experimental data is qualitatively in agreement. For  $\beta$ -carotene, deviations between experiment and COSMO-RS predictions occur for higher solubility ranges. However, a broader range of solvents for a more detailed comparison. Especially for D-glucose and L-alanine, additional datapoints would be useful to analyse mid-range solubilities.



**Fig. D.2** Linear regression of experimental solubility measurements and COSMO-RS solubilities for a) D-glucose, b) palmitic acid, c) L-alanine, and d)  $\beta$ -carotene. Experimental values were taken from literature<sup>504–506</sup>.

The  $\sigma$ -profiles of all representative molecules are shown in Fig. D.3. The  $\sigma$ -profiles of the carbohydrates show peaks in the HBD, and HBA region, indicating the possibility for inter- and intramolecular hydrogen bonding. The peak in the nonpolar region is caused by the C-atoms of the sugar rings and increases with increasing length of the polymer chain. Therefore, a higher water-solubility is expected for the monomeric sugars D-glucose and D-mannose than for the polymers. The  $\sigma$ -profiles of the amino acids reflect their relatively high water-solubility as shown by the peaks in the HBD and HBA region. L-Leucine has an aliphatic side chain causing the slight peak in the nonpolar region. Both, the lipids and the carotenoid Fx show predominantly peaks in the nonpolar region. The NL tripalmitin has the strongest nonpolar character which is reflected by its extremely low water solubility. Fx and the PLs have slight peaks in the HBA region, and lower peaks in the nonpolar region, underlining their more polar characteristics compared to the NL tripalmitin.



**Fig. D.3**  $\sigma$ -profiles of the representative microalgal molecules: a) carbohydrates, b) amino acids representing the protein fraction, c) fucoxanthin representing the carotenoids, and d) lipids.

### D.1.2 Manually added solvents after EHS screening step

Solvents that were falsely excluded in the EHS screening step due to model uncertainties were manually re-added to the list of eligible solvents (Tab. D.2).

**Tab. D.2** Solvent candidates that were falsely excluded in the EHS property screening step and were manually re-added to the solvent list.

Name	CAS-Number
Methyl-trans-2-butenolate	000623-43-8
2-Methyl-pentanal	000123-15-9
Methyltrimethylacetate	000598-98-1
Isobutanol	000078-83-1
3,3-Dimethyl-2-butanone	000075-97-8
Butanal	000123-72-8
Butanethiol	000109-79-5
1,1-Diethoxyethane	000105-57-7
Diisopropylether	108-20-3
Methylmethacrylate	000080-62-6
2-Methyltetrahydrofuran	000096-47-9
Nitromethane	75-52-5
Tetrahydropyran	000142-68-7
Butylformate	000592-84-7
Pyridine	110-86-1
Acetylchloride	000075-36-5
5-Methylfurfural	000620-02-0
4-Methyl-2-pentanone	108-10-1
Ethylpropionate	105-37-3
Methyl-t-butylether	001634-04-4
Acetonitrile	75-05-8

### D.1.3 Solvent candidates identified for each biomass fraction

A list of solvent candidates identified after the solubility screening step for each biomass fraction is provided in the electronic supplements (microalgae\_solvent\_screening.zip).

#### D.1.3.1 $\sigma$ -profiles of the identified solvents

The  $\sigma$ -profiles of the solvents identified in the solvent screening is shown in Fig. D.4. All identified solvents exhibit peaks in the nonpolar region. N-hexane is the only solvent without any peaks in the HBA or HBD region, highlighting its strongly hydrophobic characteristics. Solvents exhibiting a moderate peak in the nonpolar region that are, in addition, able to donate or accept hydrogen bonds, extracted more lipids and pigments from wet *P. tricornutum* biomass compared to completely nonpolar solvents, such as n-hexane and vinyl ethers.

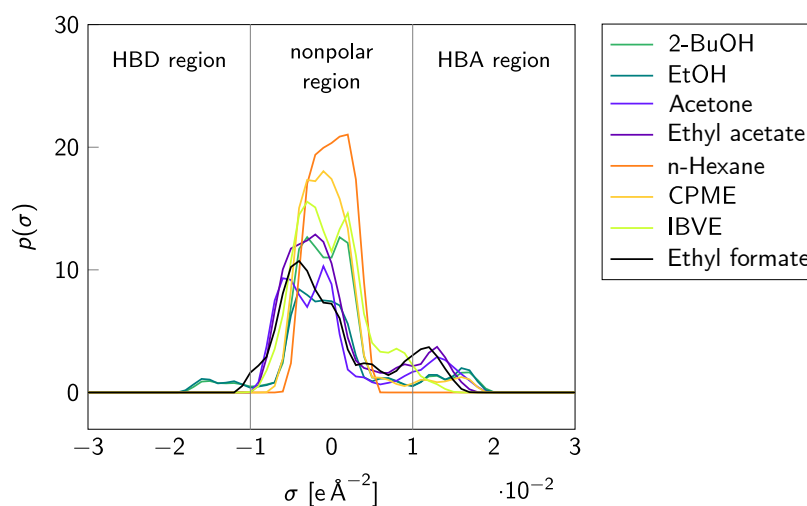
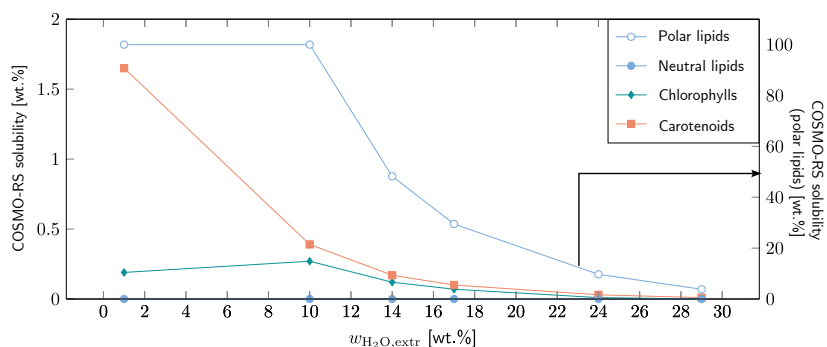


Fig. D.4  $\sigma$ -profiles of the solvents identified by the screening approach.

## D.2 COSMO-RS solubility prediction of representative molecules in 2-butanol/water-mixtures

The solubilities of selected representative molecules (Fx, tripalmitin, lecithin, and pheophytin a as a model molecule for the chlorophyll fraction) were predicted in 2-BuOH/water mixtures at  $T = 25$  °C (Fig. D.5). In line with the "like dissolves like"-principle, the solubilities of these lipophilic compounds decreases with increasing water content.

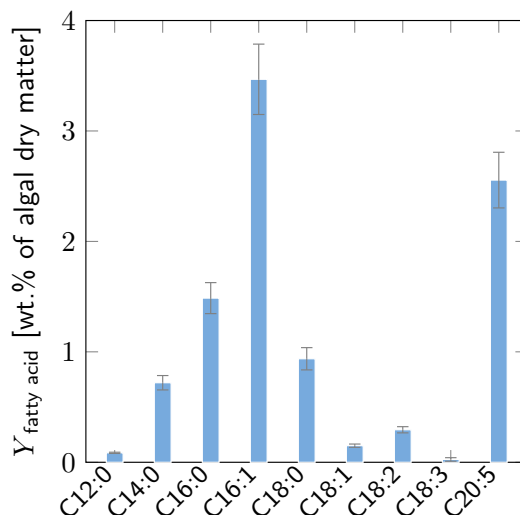


**Fig. D.5** COSMO-RS predicted solubilities for PLs, NLs, chlorophylls and carotenoids for 2-BuOH/water mixtures. Reprinted from König-Mattern et al.<sup>455</sup>

## D.3 Experimental details

### D.3.1 Biomass composition of *P.tricornutum*

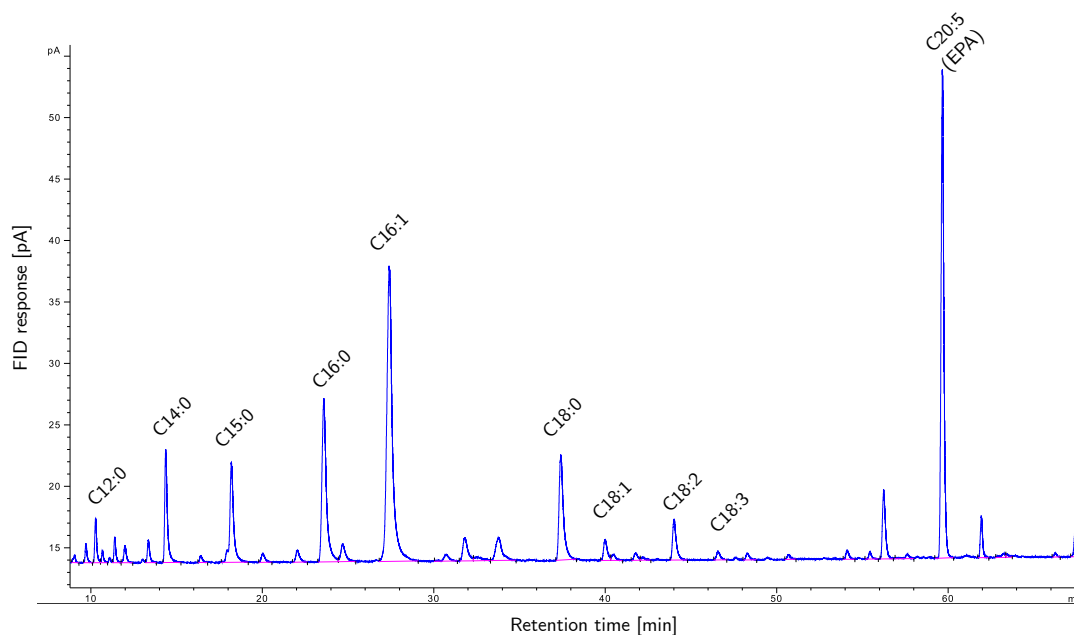
The biomass composition of the *P.tricornutum* biomass used for the experiments is provided in Tab. D.3. The fatty acid composition of the lipids as measured by GC-FID after conversion to FAMES and the corresponding chromatogram are shown in Fig. D.6 and D.7, respectively. The pigment composition as determined by HPLC and the corresponding chromatogram are provided in Fig. D.8 and D.9, respectively.



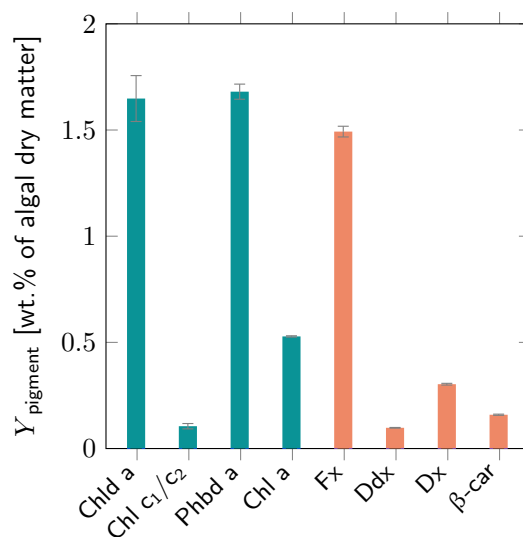
**Fig. D.6** Fatty acid composition of the *P. tricornutum* biomass used in this work (dried biomass, extracted with chloroform/methanol, 1/1, v/v). Reprinted from König-Mattern et al.<sup>455</sup>

**Tab. D.3** Biomass composition of the *P. tricornutum* applied in the extraction experiments. All values are given on a dry matter basis.

Fraction	Subfraction	Components	Mass [wt.% <sub>dry</sub> ]		
			Component	Subfraction	Fraction
Proteins	Water-soluble			18.13	39.85
	Water-insoluble			21.72	
Carbohydrates	Water-soluble	Laminarin	2.78	14.55	19.90
		Other	11.78		
	Water-insoluble			5.34	
Lipids	EPA			2.56	9.72
	PUFAs (non-EPA)			0.32	
	MUFAs			3.62	
	SFAs			3.23	
Pigments	Carotenoids	Fucoxanthin	1.49	2.05	7.61
		Other	0.56		
	Chlorophylls			5.56	
Ash					23.26

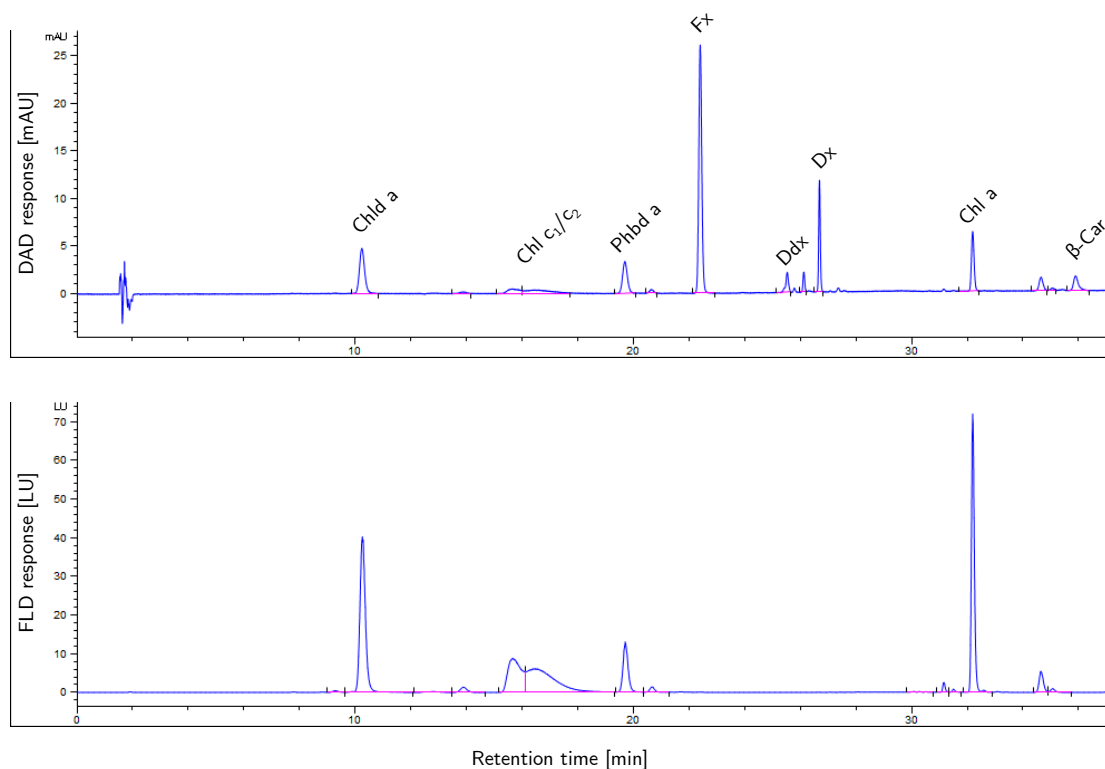


**Fig. D.7** GC-FID chromatogram of a lipid extract obtained from dried *P. tricornutum* using chloroform/methanol (1/1, v/v). The lipids were converted to FAMES prior to GC analysis.



**Fig. D.8** Chlorophyll and carotenoid composition of the *P. tricornutum* biomass used in this work (dried biomass, extracted with EtOH<sub>96</sub> vol.%). Reprinted from König-Mattern et al.<sup>455</sup>

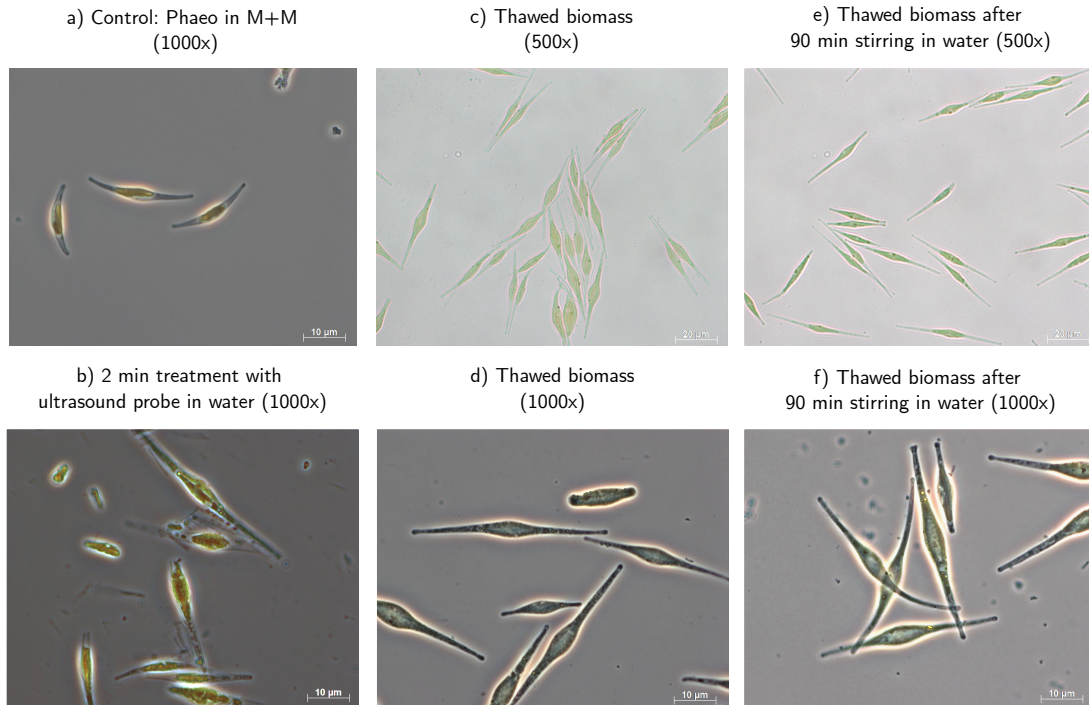




**Fig. D.9** HPLC chromatogram of a pigment extract obtained from dried *P. tricornutum* biomass using EtOH<sub>96</sub> vol.%.·

## D.4 Microscopic imaging of *P. tricornutum* cells

The biomass was inspected using a light microscope (Axio Imager A1, Carl Zeiss, Germany) during cultivation (Fig. D.10 a), after sonication treatment as described in the method section of the main manuscript (Fig. D.10 b), after thawing as used in the extraction experiments (Fig. D.10 c and d) and after incubation with H<sub>2</sub>O under conditions similar to the extraction experiments (Fig. D.10 e and f). Cell disruption using the ultrasound probe leads to cell breakage as clearly visible in (Fig. D.10 b). Cells after thawing appear swollen, presumably due to expansion caused by prior freezing at -20 °C (Fig. D.10 c and d). Incubation under magnetic stirring for 90 min as shown in Fig. D.10 e and f did not cause visual significant changes compared to Fig. D.10 c and d.



**Fig. D.10** Microscopy images of *P. tricornutum*. a) *P. tricornutum* in M+M medium in the final days of cultivation. b) *P. tricornutum* treated with ultrasound probe (2 min) in water. c) Thawed *P. tricornutum* biomass 500x magnified, and d) 1000x magnified. e) Thawed biomass after extraction with water after 90 min incubation similar to extraction experiments 500 x magnified and f) 1000x magnified. Reprinted from König-Mattern et al.<sup>455</sup>

## D.5 Economic value of *P. tricornutum* biomass

The economic value of the *P. tricornutum* biomass compounds is provided in Tab. D.4. In the scenario „complete biorefinery“, each biomass fraction was allocated to the product with the maximum economic value.

## D.6 Mass balances of the 2-butanol-based biorefinery process

The mass balance for the developed biorefinery process is given as the product of the incidence matrix  $\mathbf{A} \in \mathbb{R}^{p \times f}$  and the vector of mass flows  $\mathbf{F} \in \mathbb{R}^{f \times 1}$  :

$$\mathbf{0} = \mathbf{A} \cdot \mathbf{F}. \quad (\text{D.1})$$

The size of the incidence matrix  $\mathbf{A}$  is determined by the number of process units  $p$  and the number of mass flows connecting the process units  $f$ . The connection between the process

**Tab. D.4** Economic value of potential biorefinery products of *P. tricornutum*. The values were taken from Ruiz et al<sup>165</sup> unless otherwise stated.

Selling category	Product	Selling price [€ t <sup>-1</sup> ]	Biomass fractions
Biofuels	Biodiesel	710	Lipids
	Bioethanol	370	Carbohydrates Proteins
Chemicals	Biopolymers	2000	Lipids
		1400	Proteins
		2000	Carbohydrates (polysaccharides)
	Biolubricants	1500	Saturated FAMES
		3500	Unsaturated FAMES
Biopolymer additives	3700	Carbohydrates (polysaccharides)	
Food, feed	Proteins	1100	Proteins
		950	lipids
		750	Carbohydrates
		20000 <sup>507</sup>	Laminarin
Food additives	PUFAs (non-EPA)	5000	PUFAs (non-EPA)
	EPA	200000 <sup>508</sup>	EPA
	Functional protein	3300	Water-soluble protein
	Pigments	900000	Pigments
	Sterols	45000	Sterols
Cosmetics/health care	Lipid antioxidants	30000	PUFAs
	Carotenoid antioxidants	900000	Carotenoids
	Glyco- and phospholipids	6000	Polar lipids
	Bioactive sulfated polysaccharides	2500	Polysaccharides
	Chlorophyll colourant (E141)	12000 <sup>509</sup>	Chlorophylls
Complete biorefinery (max. profit)	EPA-food additive	200000 <sup>508</sup>	EPA
	Biolubricant (SFA)	1500	SFAs
	Biolubricant (MUFA)	3500	MUFAs
	Carbohydrate biopolymer	2000	Water-insoluble carbs
	Chlorophyll colorant (E141)	12000 <sup>509</sup>	Chlorophylls
	Cosmetics/health care protein	3500	Water-soluble protein
	Protein biopolymers	1400	Water-insoluble protein
	Laminarin powder	20000	Laminarin
	Lipid antioxidants	30000	PUFAs (non-EPA)
	Carotenoid antioxidants	900000	Carotenoids
	Food/feed: carbohydrates	750	Water-soluble carbohydrates (not laminarin)

units are given by  $\mathbf{A}$ . For flows leaving a given process unit, the corresponding value in  $\mathbf{A}$  equals -1, whereas 1 denotes flows entering the process unit. Each mass flow  $F_p$  in  $\mathbf{F}$  equals the sum of the component-specific mass flows  $F_{p,c}$ :

$$F_p = \sum_{c=1}^C F_{p,c}, \quad (\text{D.2})$$

where  $c$  denotes the total number of biomass components. The component-specific mass flows were determined for proteins, carbohydrates, lipids, carotenoids and chlorophylls were determined as

$$F_{p+1,c} = F_{p,c} \cdot \Phi_{p,c}, \quad (\text{D.3})$$

where the factor  $\Phi_{p,c} \in [0, 1]$  indicates the relative mass of component  $c$  that is transported from  $F_{p,c}$  to  $F_{p+1,c}$ . For the extraction unit,  $\Phi_{p,c}$  was described by the component yield. For phase separation,  $\Phi_{p,c}$  was given as the amount of component  $c$  that was transported to the organic or aqueous phase, respectively. The experimentally determined values for  $\Phi_{p,c}$  are summarised in Tab. D.5.

**Tab. D.5** Experimentally measured  $\Phi_{p,c}$ -factors to calculate the mass flows of the developed 2-BuOH-based biorefinery process.

Unit $p$	Component $c$	$\Phi_{p,c}$ [wt.%]	Note
2-BuOH <sub>79</sub> vol.% extraction ( $t = 90$ min)	Proteins	22.95	1, 3
	Carbohydrates	40.64	1, 3
	Lipids	89.57	1
	Carotenoids	19.94	1
	Chlorophylls	63.8	1
2-BuOH <sub>75</sub> vol.% extraction ( $t = 90$ min)	Proteins	25.37	2, 3
	Carbohydrates	58.35	2, 3
	Lipids	99.36	1
	Carotenoids	82.55	1
	Chlorophylls	100.00	1
Phase separation (organic phase)	Proteins	43.29	4
	Carbohydrates	17.57	4
	Lipids	88.45	4
	Carotenoids	97.92	4
	Chlorophylls	82.57	4
Phase separation (aqueous phase)	Proteins	56.71	4
	Carbohydrates	82.43	4
	Lipids	11.55	4
	Carotenoids	2.08	4
	Chlorophylls	17.43	4

1 - Based on extraction with 500 mg wet *P. tricornutum* biomass and 20 ml solvent/water-mix.

2 - Based on extraction with 1000 mg wet *P. tricornutum* biomass and 40 ml solvent/water-mix.

3 - The yield is based on the amount of water-soluble carbohydrates/proteins.

4 - Measured according to Section 7.5.11



## Bibliography

- [1] J. Rockström, W. Steffen, K. Noone, Å. Persson, F. S. Chapin, E. F. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, B. Nykvist, C. A. De Wit, T. Hughes, S. Van Der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. A. Foley, "A safe operating space for humanity," *Nature*, vol. 461, no. 7263, pp. 472–475, 2009.
- [2] K. Richardson, W. Steffen, W. Lucht, J. Bendtsen, S. E. Cornell, J. F. Donges, M. Drüke, I. Fetzer, G. Bala, W. Von Bloh, G. Feulner, S. Fiedler, D. Gerten, T. Gleeson, M. Hofmann, W. Huiskamp, M. Kummu, C. Mohan, D. Nogués-Bravo, S. Petri, M. Porkka, S. Rahmstorf, S. Schaphoff, K. Thonicke, A. Tobian, V. Virkki, L. Wang-Erlandsson, L. Weber, and J. Rockström, "Earth beyond six of nine planetary boundaries," *Science Advances*, vol. 9, no. 37, eadh2458, 2023.
- [3] G. S. Callendar, "The artificial production of carbon dioxide and its influence on temperature," *Quarterly Journal of the Royal Meteorological Society*, vol. 64, no. 275, pp. 223–240, 1938.
- [4] G. N. Plass, "The carbon dioxide theory of climatic change," *Tellus*, vol. 8, no. 2, pp. 140–154, 1956.
- [5] J. Tyndall, "XXIII. On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction.—the bakerian lecture," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 22, no. 146, pp. 169–194, 1861.
- [6] V. Ramanathan and Y. Feng, "Air pollution, greenhouse gases and climate change: Global and regional perspectives," *Atmospheric Environment*, vol. 43, no. 1, pp. 37–50, 2009.
- [7] Intergovernmental Panel on Climate Change (IPCC), "Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the Intergovernmental Panel on Climate Change," Geneva, Switzerland, Tech. Rep., 2023.

- [8] United Nations Environment Programme, *Paris Agreement*, <https://wedocs.unep.org/20.500.11822/20830>, last accessed: Jan 17, 2024, 2015.
- [9] European Chemical Industry Council, *The journey of petrochemicals explained: From raw materials to 95% of all manufactured goods*, <https://cefic.org/media-corner/newsroom/the-journey-of-petrochemicals-explained-from-raw-materials-to-95-of-all-manufactured-goods/>, last accessed: Jan 17, 2024.
- [10] European Statistical Office, *Final energy consumption by sector*, <https://ec.europa.eu/eurostat/web/products-datasets/-/ten00124>, DOI: 10.2908/TEN00124, last accessed: Jan 17, 2024, 2024.
- [11] D. Perez Sanchez, *International Energy Agency - Chemicals*, <https://www.iea.org/energy-system/industry/chemicals>, last accessed: Jan 17, 2024, 2024.
- [12] European Chemical Industry Council, *The chemical industry - a vital part of Europe's future*, [https://cefic.org/app/uploads/2023/12/2023\\_Facts\\_and\\_Figures\\_The\\_Leaflet.pdf](https://cefic.org/app/uploads/2023/12/2023_Facts_and_Figures_The_Leaflet.pdf), last accessed: Jan 17, 2024, 2023.
- [13] A. M. Bazzanella, F. Ausfelder, and DECHEMA Gesellschaft für Chemische Technik und Biotechnologie e.V., "Technology study: Low carbon energy and feedstock for the European chemical industry," Frankfurt am Main, Germany, Tech. Rep., 2017.
- [14] F. Meng, A. Wagner, A. B. Kremer, D. Kanazawa, J. J. Leung, P. Goult, M. Guan, S. Herrmann, E. Speelman, P. Sauter, S. Lingeswaran, M. M. Stuchtey, K. Hansen, E. Masanet, A. C. Serrenho, N. Ishii, Y. Kikuchi, and J. M. Cullen, "Planet-compatible pathways for transitioning the chemical industry," *Proceedings of the National Academy of Sciences*, vol. 120, no. 8, e2218294120, 2023.
- [15] G. Lopez, D. Keiner, M. Fasihi, T. Koiranen, and C. Breyer, "From fossil to green chemicals: Sustainable pathways and new carbon feedstocks for the global chemical industry," *Energy & Environmental Science*, vol. 16, no. 7, pp. 2879–2909, 2023.
- [16] "Roadmap Chemie 2050 auf dem Weg zu einer treibhausgasneutralen chemischen Industrie in Deutschland: eine Studie von DECHEMA und FutureCamp für den VCI," DECHEMA Gesellschaft für Chemische Technik und Biotechnologie e.V, Frankfurt am Main, Tech. Rep., 2019.
- [17] P. Münnich, J. Somers, J. Metz, U. Tillmann, C. Oliviera, R. Hermanns, A. Kalousdian, R. Meys, A. Bardow, and B. Winter, "Chemicals in transition - the three pillars for transforming chemical value chains," Berlin, Germany, Tech. Rep., 2023.



- [18] Circle Economy, *The circularity gap report*, <https://www.circularity-gap.world/2023>, 2023, last accessed: Jan 17, 2023.
- [19] European Commission. Directorate General for Internal Market, Industry, Entrepreneurship and SMEs., *Study on the Critical Raw Materials for the EU 2023: Final Report*. LU: Publications Office, 2023.
- [20] J. Hopewell, R. Dvorak, and E. Kosior, "Plastics recycling: Challenges and opportunities," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1526, pp. 2115–2126, 2009.
- [21] Circle Economy, *The circularity gap report*, <https://www.circularity-gap.world/2021>, 2021, last accessed: Jan 17, 2023.
- [22] P. Stegmann, M. Londo, and M. Junginger, "The circular bioeconomy: Its elements and role in European bioeconomy clusters," *Resources, Conservation & Recycling: X*, vol. 6, p. 100 029, 2020.
- [23] D. R. Dodds and R. A. Gross, "Chemicals from biomass," *Science*, vol. 318, no. 5854, pp. 1250–1251, 2007.
- [24] S. K. Maity, "Opportunities, recent trends and challenges of integrated biorefinery: Part I," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 1427–1445, 2015.
- [25] European Commission, "Bioeconomy: The European way to use our natural resources," Brussels, Belgium, Tech. Rep., 2018.
- [26] Federal Ministry of Education and Research and Federal Ministry of Food and Agriculture, "National bioeconomy strategy," Berlin, Germany, Tech. Rep., 2020.
- [27] Verband der Chemischen Industrie e.V. (VCI), "Biomass as a resource for the chemical industry," Frankfurt am Main, Germany, Tech. Rep., 2007.
- [28] Material Economics, "EU biomass use in a net-zero economy - a course correction for EU biomass," Tech. Rep., 2021.
- [29] M. Carus and L. Dammer, "The circular bioeconomy - concepts, opportunities, and limitations," *Industrial Biotechnology*, vol. 14, no. 2, pp. 83–91, 2018.
- [30] A. Muscat, E. M. De Olde, R. Ripoll-Bosch, H. H. E. Van Zanten, T. A. P. Metze, C. J. A. M. Termeeer, M. K. Van Ittersum, and I. J. M. De Boer, "Principles, drivers and opportunities of a circular bioeconomy," *Nature Food*, vol. 2, no. 8, pp. 561–566, 2021.
- [31] E. de Jong and IEA Bioenergy, "Biorefineries: Adding value to the sustainable utilisation of biomass," Tech. Rep., 2009.
- [32] B. Annevelink, L. Garcia Chavez, R. van Ree, and I. Vural Gursel, "Global biorefinery status report 2022," International Energy Agency, Tech. Rep., 2022.

- [33] Federal Ministry of Food, Agriculture and Consumer Protection (BMELV), Federal Ministry of Education and Research (BMBF), Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU), and Federal Ministry of Economics and Technology (BMW), "Biorefineries roadmap as part of the German federal government action plans for the material and energetic utilisation of renewable raw materials," p. 108, 2012.
- [34] F. Cherubini, G. Jungmeier, M. Wellisch, T. Willke, I. Skiadas, R. Van Ree, and E. De Jong, "Toward a common classification approach for biorefinery systems," *Biofuels, Bioproducts and Biorefining*, vol. 3, no. 5, pp. 534–546, 2009.
- [35] L. M. Schmidt, L. F. Andersen, C. Dieckmann, A. Lamp, and M. Kaltschmitt, "The biorefinery approach," in *Encyclopedia of Sustainability Science and Technology*, R. A. Meyers, Ed., New York, NY: Springer New York, 2018, pp. 1–30.
- [36] W. Schutyser, T. Renders, S. Van den Bosch, S.-F. Koelewijn, G. T. Beckham, and B. F. Sels, "Chemicals from lignin: An interplay of lignocellulose fractionation, depolymerisation, and upgrading," *Chemical Society Reviews*, vol. 47, no. 3, pp. 852–908, 2018.
- [37] J. C. Velasco Calderón, J. S. Arora, and S. H. Mushrif, "Mechanistic investigation into the formation of humins in acid-catalyzed biomass reactions," *ACS Omega*, vol. 7, no. 49, pp. 44 786–44 795, 2022.
- [38] L. Shuai, M. T. Amiri, Y. M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph, and J. S. Luterbacher, "Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization," *Science*, vol. 354, no. 6310, pp. 329–333, 2016.
- [39] L. Machado, G. Carvalho, and R. N. Pereira, "Effects of innovative processing methods on microalgae cell wall: Prospects towards digestibility of protein-rich biomass," *Biomass*, vol. 2, no. 2, pp. 80–102, 2022.
- [40] A. A. Kiss, J.-P. Lange, B. Schuur, D. Brillman, A. Van Der Ham, and S. R. Kersten, "Separation technology—making a difference in biorefineries," *Biomass and Bioenergy*, vol. 95, pp. 296–309, 2016.
- [41] F. G. Calvo-Flores and F. J. Martin-Martinez, "Biorefineries: Achievements and challenges for a bio-based economy," *Frontiers in Chemistry*, vol. 10, p. 973 417, 2022.
- [42] T. Zhou, K. McBride, S. Linke, Z. Song, and K. Sundmacher, "Computer-aided solvent selection and design for efficient chemical processes," *Current Opinion in Chemical Engineering*, vol. 27, pp. 35–44, 2020.
- [43] C. Reichardt, "Solvents and solvent effects: An introduction," *Organic Process Research & Development*, vol. 11, no. 1, pp. 105–113, 2007.

- [44] F. P. Byrne, S. Jin, G. Paggiola, T. H. M. Petchey, J. H. Clark, T. J. Farmer, A. J. Hunt, C. Robert McElroy, and J. Sherwood, "Tools and techniques for solvent selection: Green solvent selection guides," *Sustainable Chemical Processes*, vol. 4, no. 1, p. 7, 2016.
- [45] American Chemical Society, *12 principles of green chemistry*, <https://www.acs.org/greenchemistry/principles/12-principles-of-green-chemistry.html>, last accessed: Jan 22, 2024.
- [46] C. Jimenez-Gonzalez, "Life cycle considerations of solvents," *Current Opinion in Green and Sustainable Chemistry*, vol. 18, pp. 66–71, 2019.
- [47] M. J. Kamlet, P. W. Carr, R. W. Taft, and M. H. Abraham, "Linear solvation energy relationships. 13. Relationship between the hildebrand solubility parameter,  $\Delta H$ , and the solvatochromic parameter,  $\pi^*$ ," *Journal of the American Chemical Society*, vol. 103, no. 20, pp. 6062–6066, 1981.
- [48] C. M. Hansen, "The three dimensional solubility parameter and solvent diffusion coefficient - their importance in surface coating formulation," Ph.D. dissertation, Technical University of Denmark, Copenhagen, Denmark, 1967.
- [49] Z. Song, C. Zhang, Z. Qi, T. Zhou, and K. Sundmacher, "Computer-aided design of ionic liquids as solvents for extractive desulfurization," *AIChE Journal*, vol. 64, no. 3, pp. 1013–1025, 2018.
- [50] S. Linke, K. McBride, and K. Sundmacher, "Systematic green solvent selection for the hydroformylation of long-chain alkenes," *ACS Sustainable Chemistry & Engineering*, accsuschemeng.0c02611, 2020.
- [51] J. Scheffczyk, C. Redepinning, 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*, vol. 115, pp. 433–442, 2016.
- [52] Y. Chu and X. He, "MoDooop: An automated computational approach for COSMO-RS prediction of biopolymer solubilities in ionic liquids," *ACS Omega*, vol. 4, no. 1, pp. 2337–2343, 2019.
- [53] S. Rezaei Motlagh, R. Harun, D. R. Awang Biak, S. A. Hussain, R. Omar, and A. A. Elgharbawy, "COSMO-RS based prediction for alpha-linolenic acid (ALA) extraction from microalgae biomass using room temperature ionic liquids (RTILs)," *Marine Drugs*, vol. 18, no. 2, p. 108, 2020.

- [54] S. Rezaei Motlagh, R. Harun, D. Radiah Awang Biak, S. A. Hussain, A. A. Elgharbawy, R. Khezri, and C. D. Wilfred, "Prediction of potential ionic liquids (ILs) for the solid–liquid extraction of docosahexaenoic acid (DHA) from microalgae using COSMO-RS screening model," *Biomolecules*, vol. 10, no. 8, p. 1149, 2020.
- [55] S. Rezaei Motlagh, R. Harun, D. Awang Biak, S. Hussain, W. Wan Ab Karim Ghani, R. Khezri, C. Wilfred, and A. Elgharbawy, "Screening of suitable ionic liquids as green solvents for extraction of eicosapentaenoic acid (EPA) from microalgae biomass using COSMO-RS model," *Molecules*, vol. 24, no. 4, p. 713, 2019.
- [56] R. C. Kuhad and A. Singh, "Lignocellulose biotechnology: Current and future prospects," *Critical Reviews in Biotechnology*, vol. 13, no. 2, pp. 151–172, 1993.
- [57] S. Bertella and J. S. Luterbacher, "Lignin functionalization for the production of novel materials," *Trends in Chemistry*, vol. 2, no. 5, pp. 440–453, 2020.
- [58] F. H. Isikgor and C. R. Becer, "Lignocellulosic biomass: A sustainable platform for the production of bio-based chemicals and polymers," *Polymer Chemistry*, vol. 6, no. 25, pp. 4497–4559, 2015.
- [59] H. Klose and G. Paës, "Editorial: Understanding plant cell wall recalcitrance for efficient lignocellulose processing," *Frontiers in Plant Science*, vol. 14, p. 1284658, 2023.
- [60] J. Ralph, C. Lapierre, and W. Boerjan, "Lignin structure and its engineering," *Current Opinion in Biotechnology*, vol. 56, pp. 240–249, 2019.
- [61] D. Klemm, B. Heublein, H.-P. Fink, and A. Bohn, "Cellulose: Fascinating biopolymer and sustainable raw material," *Angewandte Chemie International Edition*, vol. 44, no. 22, pp. 3358–3393, 2005.
- [62] M. Wohlert, T. Bensselfelt, L. Wågberg, I. Furó, L. A. Berglund, and J. Wohlert, "Cellulose and the role of hydrogen bonds: Not in charge of everything," *Cellulose*, vol. 29, no. 1, pp. 1–23, 2022.
- [63] J. D. Kubicki, H. Yang, D. Sawada, H. O'Neill, D. Oehme, and D. Cosgrove, "The shape of native plant cellulose microfibrils," *Scientific Reports*, vol. 8, no. 1, p. 13983, 2018.
- [64] K. Kulasinski, S. Keten, S. V. Churakov, D. Derome, and J. Carmeliet, "A comparative molecular dynamics study of crystalline, paracrystalline and amorphous states of cellulose," *Cellulose*, vol. 21, no. 3, pp. 1103–1116, 2014.
- [65] M. C. Jarvis, "Hydrogen bonding and other non-covalent interactions at the surfaces of cellulose microfibrils," *Cellulose*, vol. 30, no. 2, pp. 667–687, 2023.

- [66] L. Alves, B. Medronho, F. E. Antunes, D. Topgaard, and B. Lindman, "Dissolution state of cellulose in aqueous systems. 1. Alkaline solvents," *Cellulose*, vol. 23, no. 1, pp. 247–258, 2016.
- [67] S. Fischer, K. Thümmler, B. Volkert, K. Hettrich, I. Schmidt, and K. Fischer, "Properties and applications of cellulose acetate," *Macromolecular Symposia*, vol. 262, no. 1, pp. 89–96, 2008.
- [68] L. Szabó, R. Milotskyi, G. Sharma, and K. Takahashi, "Cellulose processing in ionic liquids from a materials science perspective: Turning a versatile biopolymer into the cornerstone of our sustainable future," *Green Chemistry*, vol. 25, no. 14, pp. 5338–5389, 2023.
- [69] M. Ghasemi, M. Tsianou, and P. Alexandridis, "Assessment of solvents for cellulose dissolution," *Bioresource Technology*, vol. 228, pp. 330–338, 2017.
- [70] M. Dusselier, M. Mascal, and B. F. Sels, "Top chemical opportunities from carbohydrate biomass: A chemist's view of the biorefinery," in *Selective Catalysis for Renewable Feedstocks and Chemicals*, K. M. Nicholas, Ed., vol. 353, Cham: Springer International Publishing, 2014, pp. 1–40.
- [71] M. Sajid, X. Zhao, and D. Liu, "Production of 2,5-furandicarboxylic acid (FDCA) from 5-hydroxymethylfurfural (HMF): Recent progress focusing on the chemical-catalytic routes," *Green Chemistry*, vol. 20, no. 24, pp. 5427–5453, 2018.
- [72] C. Chatterjee, F. Pong, and A. Sen, "Chemical conversion pathways for carbohydrates," *Green Chemistry*, vol. 17, no. 1, pp. 40–71, 2015.
- [73] L. K. Rihko-Struckmann, O. Oluyinka, A. Sahni, K. McBride, M. Fchet, K. Ludwig, and K. Sundmacher, "Transformation of remnant algal biomass to 5-HMF and levulinic acid: Influence of a biphasic solvent system," *RSC Advances*, vol. 10, no. 42, pp. 24 753–24 763, 2020.
- [74] I. Van Zandvoort, E. J. Koers, M. Weingarth, P. C. A. Bruijninx, M. Baldus, and B. M. Weckhuysen, "Structural characterization of  $^{13}\text{C}$ -enriched humins and alkali-treated  $^{13}\text{C}$  humins by 2D solid-state NMR," *Green Chemistry*, vol. 17, no. 8, pp. 4383–4392, 2015.
- [75] H. V. Scheller and P. Ulvskov, "Hemicelluloses," *Annual Review of Plant Biology*, vol. 61, no. 1, pp. 263–289, 2010.
- [76] A. Kirui, W. Zhao, F. Deligey, H. Yang, X. Kang, F. Mentink-Vigier, and T. Wang, "Carbohydrate-aromatic interface and molecular architecture of lignocellulose," *Nature Communications*, vol. 13, no. 1, p. 538, 2022.

- [77] P. Sannigrahi, A. J. Ragauskas, and G. A. Tuskan, "Poplar as a feedstock for biofuels: A review of compositional characteristics," *Biofuels, Bioproducts and Biorefining*, vol. 4, no. 2, pp. 209–226, 2010.
- [78] V. B. Agbor, N. Cicek, R. Sparling, A. Berlin, and D. B. Levin, "Biomass pretreatment: Fundamentals toward application," *Biotechnology Advances*, vol. 29, no. 6, pp. 675–685, 2011.
- [79] A. O. Komarova, G. R. Dick, and J. S. Luterbacher, "Diformylxylose as a new polar aprotic solvent produced from renewable biomass," *Green Chemistry*, vol. 23, no. 13, pp. 4790–4799, 2021.
- [80] L. P. Manker, G. R. Dick, A. Demongeot, M. A. Hedou, C. Rayroud, T. Rambert, M. J. Jones, I. Sulaeva, M. Vieli, Y. Leterrier, A. Potthast, F. Maréchal, V. Michaud, H.-A. Klok, and J. S. Luterbacher, "Sustainable polyesters via direct functionalization of lignocellulosic sugars," *Nature Chemistry*, vol. 14, no. 9, pp. 976–984, 2022.
- [81] X. Kang, A. Kirui, M. C. Dickwella Widanage, F. Mentink-Vigier, D. J. Cosgrove, and T. Wang, "Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR," *Nature Communications*, vol. 10, no. 1, p. 347, 2019.
- [82] M. Chen, Y. Li, H. Liu, D. Zhang, Q.-S. Shi, X.-Q. Zhong, Y. Guo, and X.-B. Xie, "High value valorization of lignin as environmental benign antimicrobial," *Materials Today Bio*, vol. 18, p. 100520, 2023.
- [83] A. Tolbert, H. Akinosho, R. Khunsupat, A. K. Naskar, and A. J. Ragauskas, "Characterization and analysis of the molecular weight of lignin for biorefining studies," *Biofuels, Bioproducts and Biorefining*, vol. 8, no. 6, pp. 836–856, 2014.
- [84] J. Zakzeski, P. C. A. Bruijninx, A. L. Jongerius, and B. M. Weckhuysen, "The catalytic valorization of lignin for the production of renewable chemicals," *Chemical Reviews*, vol. 110, no. 6, pp. 3552–3599, 2010.
- [85] S. Baumberger, A. Abaecherli, M. Fasching, G. Gellerstedt, R. Gosselink, B. Hortling, J. Li, B. Saake, and E. De Jong, "Molar mass determination of lignins by size-exclusion chromatography: Towards standardisation of the method," *Holzforschung*, vol. 61, no. 4, pp. 459–468, 2007.
- [86] R. Vanholme, K. Morreel, C. Darrah, P. Oyarce, J. H. Grabber, J. Ralph, and W. Boerjan, "Metabolic engineering of novel lignin in biomass crops," *New Phytologist*, vol. 196, no. 4, pp. 978–1000, 2012.
- [87] M. V. Galkin and J. S. M. Samec, "Lignin valorization through catalytic lignocellulose fractionation: A fundamental platform for the future biorefinery," *ChemSusChem*, vol. 9, no. 13, pp. 1544–1558, 2016.

- [88] R. Parthasarathi, R. A. Romero, A. Redondo, and S. Gnanakaran, "Theoretical study of the remarkably diverse linkages in lignin," *The Journal of Physical Chemistry Letters*, vol. 2, no. 20, pp. 2660–2666, 2011.
- [89] X. Liu, F. P. Bouxin, J. Fan, V. L. Budarin, C. Hu, and J. H. Clark, "Recent advances in the catalytic depolymerization of lignin towards phenolic chemicals: A review," *ChemSusChem*, vol. 13, no. 17, pp. 4296–4317, 2020.
- [90] M. M. Abu-Omar, K. Barta, G. T. Beckham, J. S. Luterbacher, J. Ralph, R. Rinaldi, Y. Román-Leshkov, J. S. M. Samec, B. F. Sels, and F. Wang, "Guidelines for performing lignin-first biorefining," *Energy & Environmental Science*, vol. 14, no. 1, pp. 262–292, 2021.
- [91] R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijninx, and B. M. Weckhuysen, "Paving the way for lignin valorisation: Recent advances in bioengineering, biorefining and catalysis," *Angewandte Chemie International Edition*, vol. 55, no. 29, pp. 8164–8215, 2016.
- [92] A. J. Ragauskas, G. T. Beckham, M. J. Bidddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan, and C. E. Wyman, "Lignin valorization: Improving lignin processing in the biorefinery," *Science*, vol. 344, no. 6185, p. 1246843, 2014.
- [93] J.-B. Huang, S.-B. Wu, H. Cheng, M. Lei, J.-J. Liang, and H. Tong, "Theoretical study of bond dissociation energies for lignin model compounds," *Journal of Fuel Chemistry and Technology*, vol. 43, no. 4, pp. 429–436, 2015.
- [94] S. Constant, H. L. J. Wienk, A. E. Frissen, P. D. Peinder, R. Boelens, D. S. Van Es, R. J. H. Grisel, B. M. Weckhuysen, W. J. J. Huijgen, R. J. A. Gosselink, and P. C. A. Bruijninx, "New insights into the structure and composition of technical lignins: A comparative characterisation study," *Green Chemistry*, vol. 18, no. 9, pp. 2651–2665, 2016.
- [95] L. Gurrula, M. M. Kumar, and R. Vinu, "Catalytic hydrogenolysis of lignin to phenols: Effect of operating conditions on product distribution," in *Biomass, Biofuels, Biochemicals*, Elsevier, 2021, pp. 83–107.
- [96] E. Subbotina, T. Rukkijakan, M. D. Marquez-Medina, X. Yu, M. Johnsson, and J. S. M. Samec, "Oxidative cleavage of C–C bonds in lignin," *Nature Chemistry*, vol. 13, no. 11, pp. 1118–1125, 2021.

- [97] C. T. Palumbo, N. X. Gu, A. C. Bleem, K. P. Sullivan, R. Katahira, L. M. Stanley, J. K. Kenny, M. A. Ingraham, K. J. Ramirez, S. J. Haugen, C. R. Amendola, S. S. Stahl, and G. T. Beckham, "Catalytic carbon-carbon bond cleavage in lignin via manganese-zirconium-mediated autoxidation," *Nature Communications*, vol. 15, no. 1, p. 862, 2024.
- [98] Z. Luo, C. Liu, A. Radu, D. F. De Waard, Y. Wang, J. T. Behaghel De Bueren, P. D. Kouris, M. D. Boot, J. Xiao, H. Zhang, R. Xiao, J. S. Luterbacher, and E. J. M. Hensen, "Carbon-carbon bond cleavage for a lignin refinery," *Nature Chemical Engineering*, vol. 1, no. 1, pp. 61-72, 2024.
- [99] J. Ruwoldt, F. H. Blindheim, and G. Chinga-Carrasco, "Functional surfaces, films, and coatings with lignin – a critical review," *RSC Advances*, vol. 13, no. 18, pp. 12 529-12 553, 2023.
- [100] A. Dastpak, T. V. Lourencon, M. Balakshin, S. Farhan Hashmi, M. Lundström, and B. P. Wilson, "Solubility study of lignin in industrial organic solvents and investigation of electrochemical properties of spray-coated solutions," *Industrial Crops and Products*, vol. 148, p. 112 310, 2020.
- [101] J. Sameni, S. Krigstin, and M. Sain, "Solubility of lignin and acetylated lignin in organic solvents," *BioResources*, vol. 12, no. 1, pp. 1548-1565, 2017.
- [102] A. Manisekaran, P. Grysan, B. Duez, D. F. Schmidt, D. Lenoble, and J.-S. Thomann, "Solvents drive self-assembly mechanisms and inherent properties of kraft lignin nanoparticles (<50 nm)," *Journal of Colloid and Interface Science*, vol. 626, pp. 178-192, 2022.
- [103] P. Figueiredo, M. H. Lahtinen, M. B. Agustin, D. M. De Carvalho, S.-P. Hirvonen, P. A. Penttilä, and K. S. Mikkonen, "Green fabrication approaches of lignin nanoparticles from different technical lignins: A comparison study," *ChemSusChem*, vol. 14, no. 21, pp. 4718-4730, 2021.
- [104] H. Sadeghifar and A. Ragauskas, "Perspective on technical lignin fractionation," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 22, pp. 8086-8101, 2020.
- [105] C. Gioia, M. Colonna, A. Tagami, L. Medina, O. Sevastyanova, L. A. Berglund, and M. Lawoko, "Lignin-based epoxy resins: Unravelling the relationship between structure and material properties," *Biomacromolecules*, vol. 21, no. 5, pp. 1920-1928, 2020.
- [106] A. Bridgwater, "Review of fast pyrolysis of biomass and product upgrading," *Biomass and Bioenergy*, vol. 38, pp. 68-94, 2012.
- [107] V. Dhyani and T. Bhaskar, "A comprehensive review on the pyrolysis of lignocellulosic biomass," *Renewable Energy*, vol. 129, pp. 695-716, 2018.



- [108] H. Wang, J. Male, and Y. Wang, "Recent advances in hydrotreating of pyrolysis bio-oil and its oxygen-containing model compounds," *ACS Catalysis*, vol. 3, no. 5, pp. 1047–1070, 2013.
- [109] C. Liu, H. Wang, A. M. Karim, J. Sun, and Y. Wang, "Catalytic fast pyrolysis of lignocellulosic biomass," *Chem. Soc. Rev.*, vol. 43, no. 22, pp. 7594–7623, 2014.
- [110] D. A. Ruddy, J. A. Schaidle, J. R. Ferrell Iii, J. Wang, L. Moens, and J. E. Hensley, "Recent advances in heterogeneous catalysts for bio-oil upgrading via "ex situ catalytic fast pyrolysis": Catalyst development through the study of model compounds," *Green Chem.*, vol. 16, no. 2, pp. 454–490, 2014.
- [111] S. Xiu and A. Shahbazi, "Bio-oil production and upgrading research: A review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 7, pp. 4406–4414, 2012.
- [112] D. C. Makepa, C. H. Chihobo, W. R. Ruziwa, and D. Musademba, "A systematic review of the techno-economic assessment and biomass supply chain uncertainties of biofuels production from fast pyrolysis of lignocellulosic biomass," *Fuel Communications*, vol. 14, p. 100 086, 2023.
- [113] M. B. Figueirêdo, I. Hita, P. J. Deuss, R. H. Venderbosch, and H. J. Heeres, "Pyrolytic lignin: A promising biorefinery feedstock for the production of fuels and valuable chemicals," *Green Chemistry*, vol. 24, no. 12, pp. 4680–4702, 2022.
- [114] S. Ashoor, T. U. Khang, Y. H. Lee, J. S. Hyung, S. Y. Choi, S. E. Lim, J. Lee, S. J. Park, and J.-G. Na, "Bioupgrading of the aqueous phase of pyrolysis oil from lignocellulosic biomass: A platform for renewable chemicals and fuels from the whole fraction of biomass," *Bioresources and Bioprocessing*, vol. 10, no. 1, p. 34, 2023.
- [115] K. N. Yogalakshmi, T. Poornima Devi, P. Sivashanmugam, S. Kavitha, R. Yukesh Kannah, S. Varjani, S. AdishKumar, and G. Kumar, "Lignocellulosic biomass-based pyrolysis: A comprehensive review," *Chemosphere*, vol. 286, p. 131 824, 2022.
- [116] M. Ragnar, G. Henriksson, M. E. Lindström, M. Wimby, J. Blechschmidt, and S. Heinemann, "Pulp," in *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2000, pp. 3–88.
- [117] F. S. Chakar and A. J. Ragauskas, "Review of current and future softwood Kraft lignin process chemistry," *Industrial Crops and Products*, vol. 20, no. 2, pp. 131–141, 2004.
- [118] J. Gierer, "Chemical aspects of Kraft pulping," *Wood Science and Technology*, vol. 14, no. 4, pp. 241–266, 1980.
- [119] R. Patt, O. Kordsachia, R. Süttinger, Y. Ohtani, J. F. Hoesch, P. Ehrler, R. Eichinger, H. Holik, U. Hamm, and M. E. Rohmann, "Pulp," in *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2000, pp. 477–533.

- [120] M. Holtzapple, "Lignin," in *Encyclopedia of Food Sciences and Nutrition*, Elsevier, 2003, pp. 3535–3542.
- [121] P. Azadi, O. R. Inderwildi, R. Farnood, and D. A. King, "Liquid fuels, hydrogen and chemicals from lignin: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 506–523, 2013.
- [122] B. Saake and R. Lehnen, "Lignin," in *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2000, pp. 21–35.
- [123] F. G. Calvo-Flores and J. A. Dobado, "Lignin as renewable raw material," *ChemSusChem*, vol. 3, no. 11, pp. 1227–1235, 2010.
- [124] M. Lawoko and J. S. Samec, "Kraft lignin valorization: Biofuels and thermoset materials in focus," *Current Opinion in Green and Sustainable Chemistry*, vol. 40, p. 100738, 2023.
- [125] J. Jönsson, K. Pettersson, T. Berntsson, and S. Harvey, "Comparison of options for utilization of a potential steam surplus at kraft pulp mills—Economic performance and CO<sub>2</sub> emissions: Comparison of options for utilizing excess steam at kraft pulp mills," *International Journal of Energy Research*, vol. 37, no. 9, pp. 1017–1035, 2013.
- [126] Y. Wang, X. Yang, M. Sun, L. Ma, X. Li, and L. Shi, "Estimating carbon emissions from the pulp and paper industry: A case study," *Applied Energy*, vol. 184, pp. 779–789, 2016.
- [127] N. Brosse, M. H. Hussin, and A. A. Rahim, "Organosolv processes," in *Biorefineries*, K. Wagemann and N. Tippkötter, Eds., vol. 166, Cham: Springer International Publishing, 2017, pp. 153–176.
- [128] M.-F. Li, S. Yang, and R.-C. Sun, "Recent advances in alcohol and organic acid fractionation of lignocellulosic biomass," *Bioresource Technology*, vol. 200, pp. 971–980, 2016.
- [129] X. Zhao, K. Cheng, and D. Liu, "Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis," *Applied Microbiology and Biotechnology*, vol. 82, no. 5, pp. 815–827, 2009.
- [130] G. Tofani, E. Jasiukaitytė-Grojzdek, M. Grilc, and B. Likozar, "Organosolv biorefinery: Resource-based process optimisation, pilot technology scale-up and economics," *Green Chemistry*, vol. 26, no. 1, pp. 186–201, 2024.
- [131] W. J. J. Huijgen, A. T. Smit, J. H. Reith, and H. D. Uil, "Catalytic organosolv fractionation of willow wood and wheat straw as pretreatment for enzymatic cellulose hydrolysis," *Journal of Chemical Technology & Biotechnology*, vol. 86, no. 11, pp. 1428–1438, 2011.

- [132] F. Monteil-Rivera, M. Phuong, M. Ye, A. Halasz, and J. Hawari, "Isolation and characterization of herbaceous lignins for applications in biomaterials," *Industrial Crops and Products*, vol. 41, pp. 356–364, 2013.
- [133] C. S. Lancefield, I. Panovic, P. J. Deuss, K. Barta, and N. J. Westwood, "Pre-treatment of lignocellulosic feedstocks using biorenewable alcohols: Towards complete biomass valorisation," *Green Chemistry*, vol. 19, no. 1, pp. 202–214, 2017.
- [134] P. Sannigrahi, A. J. Ragauskas, and S. J. Miller, "Lignin structural modifications resulting from ethanol organosolv treatment of loblolly pine," *Energy & Fuels*, vol. 24, no. 1, pp. 683–689, 2010.
- [135] S. Laure, M. Leschinsky, M. Fröhling, F. Schultmann, and G. Unkelbach, "Assessment of an organosolv lignocellulose biorefinery concept based on a material flow analysis of a pilot plant," *Cellulose Chemistry and Technology*, no. 48, pp. 793–798, 2014.
- [136] L.R. Limited, *The Glycell™ process*, <http://www.leafresources.com.au/Technology-Overview/The-Glycell-Process/>, last accessed: Jan 23, 2024.
- [137] M. G. Alriols, A. Tejado, M. Blanco, I. Mondragon, and J. Labidi, "Agricultural palm oil tree residues as raw material for cellulose, lignin and hemicelluloses production by ethylene glycol pulping process," *Chemical Engineering Journal*, vol. 148, no. 1, pp. 106–114, 2009.
- [138] L. P. Novo, L. V. A. Gurgel, K. Marabezi, and A. A. D. S. Curvelo, "Delignification of sugarcane bagasse using glycerol–water mixtures to produce pulps for saccharification," *Bioresource Technology*, vol. 102, no. 21, pp. 10 040–10 046, 2011.
- [139] F. Sun and H. Chen, "Evaluation of enzymatic hydrolysis of wheat straw pretreated by atmospheric glycerol autocatalysis," *Journal of Chemical Technology & Biotechnology*, vol. 82, no. 11, pp. 1039–1044, 2007.
- [140] P. M. Grande, J. Viell, N. Theysen, W. Marquardt, P. Domínguez de María, and W. Leitner, "Fractionation of lignocellulosic biomass using the OrganoCat process," *Green Chemistry*, vol. 17, no. 6, pp. 3533–3539, 2015.
- [141] J. Snelders, E. Dornez, B. Benjelloun-Mlayah, W. J. Huijgen, P. J. de Wild, R. J. Goselink, J. Gerritsma, and C. M. Courtin, "Biorefining of wheat straw using an acetic and formic acid based organosolv fractionation process," *Bioresource Technology*, vol. 156, pp. 275–282, 2014.
- [142] M. Delmas, "Vegetal refining and agrochemistry," *Chemical Engineering & Technology*, vol. 31, no. 5, pp. 792–797, 2008.

- [143] F. Abdelkafi, H. Ammar, B. Rousseau, M. Tessier, R. El Gharbi, and A. Fradet, "Structural analysis of Alfa grass (*Stipa tenacissima* L.) lignin obtained by acetic acid/formic acid delignification," *Biomacromolecules*, vol. 12, no. 11, pp. 3895–3902, 2011.
- [144] L. Jiménez, M. De La Torre, J. Bonilla, and J. Ferrer, "Organosolv pulping of wheat straw by use of acetone-water mixtures," *Process Biochemistry*, vol. 33, no. 4, pp. 401–408, 1998.
- [145] W. J. J. Huijgen, J. H. Reith, and H. Den Uil, "Pretreatment and fractionation of wheat straw by an acetone-based organosolv process," *Industrial & Engineering Chemistry Research*, vol. 49, no. 20, pp. 10 132–10 140, 2010.
- [146] R. Katahira, A. Mittal, K. McKinney, P. N. Ciesielski, B. S. Donohoe, S. K. Black, D. K. Johnson, M. J. Bidy, and G. T. Beckham, "Evaluation of clean fractionation pretreatment for the production of renewable fuels and chemicals from corn stover," *ACS Sustainable Chemistry & Engineering*, vol. 2, no. 6, pp. 1364–1376, 2014.
- [147] J. J. Bozell, S. K. Black, M. Myers, D. Cahill, W. P. Miller, and S. Park, "Solvent fractionation of renewable woody feedstocks: Organosolv generation of biorefinery process streams for the production of biobased chemicals," *Biomass and Bioenergy*, vol. 35, no. 10, pp. 4197–4208, 2011.
- [148] M. Chen, F. Malaret, A. E. J. Firth, P. Verdía, A. R. Aboulela, Y. Chen, and J. P. Hallett, "Design of a combined ionosolv-organosolv biomass fractionation process for biofuel production and high value-added lignin valorisation," *Green Chemistry*, vol. 22, no. 15, pp. 5161–5178, 2020.
- [149] A. A. Vaidya, K. D. Murton, D. A. Smith, and G. Dedual, "A review on organosolv pretreatment of softwood with a focus on enzymatic hydrolysis of cellulose," *Biomass Conversion and Biorefinery*, vol. 12, no. 11, pp. 5427–5442, 2022.
- [150] J. Xu, C. Li, L. Dai, C. Xu, Y. Zhong, F. Yu, and C. Si, "Biomass fractionation and lignin fractionation towards lignin valorization," *ChemSusChem*, vol. 13, no. 17, pp. 4284–4295, 2020.
- [151] S. Van den Bosch, W. Schutyser, R. Vanholme, T. Driessen, S.-F. Koelewijn, T. Renders, B. De Meester, W. J. J. Huijgen, W. Dehaen, C. M. Courtin, B. Lagrain, W. Boerjan, and B. F. Sels, "Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps," *Energy & Environmental Science*, vol. 8, no. 6, pp. 1748–1763, 2015.
- [152] Y. Liu, N. Deak, Z. Wang, H. Yu, L. Hameleers, E. Jurak, P. J. Deuss, and K. Barta, "Tunable and functional deep eutectic solvents for lignocellulose valorization," *Nature Communications*, vol. 12, no. 1, p. 5424, 2021.

- [153] W. Lan, M. T. Amiri, C. M. Hunston, and J. S. Luterbacher, "Protection group effects during  $\alpha,\Gamma$ -diol lignin stabilization promote high-selectivity monomer production," *Angewandte Chemie International Edition*, vol. 57, no. 5, pp. 1356–1360, 2018.
- [154] A. De Santi, M. V. Galkin, C. W. Lahive, P. J. Deuss, and K. Barta, "Lignin-first fractionation of softwood lignocellulose using a mild dimethyl carbonate and ethylene glycol organosolv process," *ChemSusChem*, vol. 13, no. 17, pp. 4468–4477, 2020.
- [155] M. Talebi Amiri, G. R. Dick, Y. M. Questell-Santiago, and J. S. Luterbacher, "Fractionation of lignocellulosic biomass to produce uncondensed aldehyde-stabilized lignin," *Nature Protocols*, vol. 14, no. 3, pp. 921–954, 2019.
- [156] *CompTox Chemicals Dashboard*, <https://comptox.epa.gov/dashboard/>, last accessed: Sep 27, 2023.
- [157] B. Elisabeth, F. Rayen, and T. Behnam, "Microalgae culture quality indicators: A review," *Critical Reviews in Biotechnology*, vol. 41, no. 4, pp. 457–473, 2021.
- [158] I. A. Levine, "Algae," in *Microalgae in Health and Disease Prevention*, Elsevier, 2018, pp. 1–10.
- [159] A. C. Guedes, H. M. Amaro, I. Sousa-Pinto, and F. X. Malcata, "Applications of spent biomass," in *Biofuels from Algae*, 1st ed., Burlington, USA: Elsevier, 2014, pp. 205–233.
- [160] X.-W. Wang, L. Huang, P.-Y. Ji, C.-P. Chen, X.-S. Li, Y.-H. Gao, and J.-R. Liang, "Using a mixture of wastewater and seawater as the growth medium for wastewater treatment and lipid production by the marine diatom *Phaeodactylum tricornutum*," *Bioresource Technology*, vol. 289, p. 121 681, 2019.
- [161] M. M. A. Nur, W. Muizelaar, P. Boelen, and A. G. J. Buma, "Environmental and nutrient conditions influence fucoxanthin productivity of the marine diatom *Phaeodactylum tricornutum* grown on palm oil mill effluent," *Journal of Applied Phycology*, vol. 31, no. 1, pp. 111–122, 2019.
- [162] A. P. Batista, L. Gouveia, N. M. Bandarra, J. M. Franco, and A. Raymundo, "Comparison of microalgal biomass profiles as novel functional ingredient for food products," *Algal Research*, vol. 2, no. 2, pp. 164–173, 2013.
- [163] R. Chandra, H. M. Iqbal, G. Vishal, H.-S. Lee, and S. Nagra, "Algal biorefinery: A sustainable approach to valorize algal-based biomass towards multiple product recovery," *Bioresource Technology*, vol. 278, pp. 346–359, 2019.
- [164] C. G. Khoo, Y. K. Dasan, M. K. Lam, and K. T. Lee, "Algae biorefinery: Review on a broad spectrum of downstream processes and products," *Bioresour Technol*, vol. 292, p. 121 964, 2019. PMID: 31451339.

- [165] J. Ruiz, G. Olivieri, J. de Vree, R. Bosma, P. Willems, J. H. Reith, M. H. M. Eppink, D. M. M. Kleinegris, R. H. Wijffels, and M. J. Barbosa, "Towards industrial products from microalgae," *Energy & Environmental Science*, vol. 9, no. 10, pp. 3036–3043, 2016.
- [166] R. Araújo, F. Vázquez Calderón, J. Sánchez López, I. C. Azevedo, A. Bruhn, S. Fluch, M. Garcia Tasende, F. Ghaderiardakani, T. Ilmjärv, M. Laurans, M. Mac Monagail, S. Mangini, C. Peteiro, C. Rebours, T. Stefansson, and J. Ullmann, "Current status of the algae production industry in europe: An emerging sector of the blue bioeconomy," *Frontiers in Marine Science*, vol. 7, p. 626 389, 2021.
- [167] P. Spolaore, C. Joannis-Cassan, E. Duran, and A. Isambert, "Commercial applications of microalgae," *J Biosci Bioeng*, vol. 101, no. 2, pp. 87–96, 2006.
- [168] I. Barkia, N. Saari, and S. R. Manning, "Microalgae for high-value products towards human health and nutrition," *Marine Drugs*, vol. 17, no. 5, p. 304, 2019.
- [169] K. Yamaguchi, "Recent advances in microalgal bioscience in Japan, with special reference to utilization of biomass and metabolites: A review," *Journal of Applied Phycology*, vol. 8, no. 6, pp. 487–502, 1996.
- [170] K. E. Apt and P. W. Behrens, "Commercial developments in microalgal biotechnology," *Journal of Phycology*, vol. 35, no. 2, pp. 215–226, 1999.
- [171] A. Muller-Feuga, "The role of microalgae in aquaculture: Situation and trends," *Journal of Applied Phycology*, vol. 12, no. 3/5, pp. 527–534, 2000.
- [172] M. A. Borowitzka, "Microalgae for aquaculture: Opportunities and constraints," *Journal of Applied Phycology*, vol. 9, no. 5, pp. 393–401, 1997.
- [173] M. T. Russo, A. Rogato, M. Jaubert, B. J. Karas, and A. Falciatore, "*Phaeodactylum tricornutum* : An established model species for diatom molecular research and an emerging chassis for algal synthetic biology," *Journal of Phycology*, vol. 59, no. 6, pp. 1114–1122, 2023.
- [174] M. R. Brown, "The amino-acid and sugar composition of 16 species of microalgae used in mariculture," *Journal of Experimental Marine Biology and Ecology*, vol. 145, no. 1, pp. 79–99, 1991.
- [175] X. Bai, H. Song, M. Lavoie, K. Zhu, Y. Su, H. Ye, S. Chen, Z. Fu, and H. Qian, "Proteomic analyses bring new insights into the effect of a dark stress on lipid biosynthesis in *Phaeodactylum tricornutum*," *Scientific Reports*, vol. 6, no. 1, p. 25 494, 2016.

- [176] I. Megía-Hervás, A. Sánchez-Bayo, L. F. Bautista, V. Morales, F. G. Witt-Sousa, M. Segura-Fornieles, and G. Vicente, "Scale-up cultivation of *phaeodactylum tricornutum* to produce biocrude by hydrothermal liquefaction," *Processes*, vol. 8, no. 9, p. 1072, 2020.
- [177] T. Butler, R. V. Kapoore, and S. Vaidyanathan, "*Phaeodactylum tricornutum*: a diatom cell factory," *Trends in Biotechnology*, vol. 38, no. 6, pp. 606–622, 2020.
- [178] W. Yongmanitchai and O. P. Ward, "Growth of and omega-3 fatty acid production by *Phaeodactylum tricornutum* under different culture conditions," *Applied and Environmental Microbiology*, vol. 57, no. 2, pp. 419–425, 1991.
- [179] D. D. McClure, A. Luiz, B. Gerber, G. W. Barton, and J. M. Kavanagh, "An investigation into the effect of culture conditions on fucoxanthin production using the marine microalgae *Phaeodactylum tricornutum*," *Algal Research*, vol. 29, pp. 41–48, 2018.
- [180] F. Derwenskus, F. Metz, A. Gille, U. Schmid-Staiger, K. Briviba, U. Schließmann, and T. Hirth, "Pressurized extraction of unsaturated fatty acids and carotenoids from wet *Chlorella vulgaris* and *Phaeodactylum tricornutum* biomass using subcritical liquids," *GCB Bioenergy*, vol. 11, no. 1, pp. 335–344, 2019.
- [181] E. Ryckebosch, C. Bruneel, R. Termote-Verhalle, K. Muylaert, and I. Foubert, "Influence of extraction solvent system on extractability of lipid components from different microalgae species," *Algal Research*, vol. 3, pp. 36–43, 2014.
- [182] R. Vazhappilly and F. Chen, "Eicosapentaenoic acid and docosahexaenoic acid production potential of microalgae and their heterotrophic growth," *Journal of the American Oil Chemists' Society*, vol. 75, no. 3, pp. 393–397, 1998.
- [183] M. Arena, D. Auteri, S. Barmaz, G. Bellisai, A. Brancato, D. Brocca, L. Bura, H. Byers, A. Chiusolo, D. Court Marques, F. Crivellente, C. De Lentdecker, M. De Maglie, M. Egsmose, Z. Erdos, G. Fait, L. Ferreira, M. Goumenou, L. Greco, A. Ippolito, F. Istace, S. Jarrah, D. Kardassi, R. Leuschner, C. Lythgo, J. O. Magrans, P. Medina, I. Miron, T. Molnar, A. Nougadere, L. Padovani, J. M. Parra Morte, R. Pedersen, H. Reich, A. Sacchi, M. Santos, R. Serafimova, R. Sharp, A. Stanek, F. Streissl, J. Sturma, C. Szentes, J. Tarazona, A. Terron, A. Theobald, B. Vagenende, A. Verani, and L. Villamar-Bouza, "Peer review of the pesticide risk assessment of the active substance laminarin," *EFSA Journal*, vol. 15, no. 6, European Food Safety Authority (EFSA), Ed., p. 16, 2017.
- [184] M. I. Kusaikin, S. P. Ermakova, N. M. Shevchenko, V. V. Isakov, A. G. Gorshkov, A. L. Vereshchagin, M. A. Grachev, and T. N. Zvyagintseva, "Structural characteristics and

- antitumor activity of a new chrysolaminaran from the diatom alga *Synedra acus*," *Chemistry of Natural Compounds*, vol. 46, no. 1, pp. 1–4, 2010.
- [185] M. Sakai, "Current research status of fish immunostimulants," *Aquaculture*, vol. 172, no. 1-2, pp. 63–92, 1999.
- [186] R. A. Dalmo, B. Martinsen, T. E. Horsberg, A. Ramstad, C. Syvertsen, R. Seljelid, and K. Ingebrigtsen, "Prophylactic effect of  $\beta(1,3)$ -D-glucan (laminaran) against experimental *Aeromonas salmonicida* and *Vibrio salmonicida* infections: Prophylactic effect of laminaran," *Journal of Fish Diseases*, vol. 21, no. 6, pp. 459–462, 1998.
- [187] B. Morales-Lange, J. Bethke, P. Schmitt, and L. Mercado, "Phenotypical parameters as a tool to evaluate the immunostimulatory effects of laminarin in *Oncorhynchus mykiss*," *Aquaculture Research*, vol. 46, no. 11, pp. 2707–2715, 2015.
- [188] M. Huete-Ortega, K. Okurowska, R. V. Kapoore, M. P. Johnson, D. J. Gilmour, and S. Vaidyanathan, "Effect of ammonium and high light intensity on the accumulation of lipids in *Nannochloropsis oceanica* (CCAP 849/10) and *Phaeodactylum tricorutum* (CCAP 1055/1)," *Biotechnology for Biofuels*, vol. 11, no. 1, p. 60, 2018.
- [189] L. Alipanah, P. Winge, J. Rohloff, J. Najafi, T. Brembu, and A. M. Bones, "Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom *Phaeodactylum tricorutum*," *PLOS ONE*, vol. 13, no. 2, S. Lin, Ed., e0193335, 2018.
- [190] H. Abida, L.-J. Dolch, C. Mei, V. Villanova, M. Conte, M. A. Block, G. Finazzi, O. Bastien, L. Tirichine, C. Bowler, F. Rébeillé, D. Petroutsos, J. Jouhet, and E. Maréchal, "Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricorutum*," *Plant Physiology*, vol. 167, no. 1, pp. 118–136, 2015.
- [191] L. Alipanah, J. Rohloff, P. Winge, A. M. Bones, and T. Brembu, "Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricorutum*," *Journal of Experimental Botany*, vol. 66, no. 20, pp. 6281–6296, 2015.
- [192] L. Stiefvatter, U. Neumann, A. Rings, K. Frick, U. Schmid-Staiger, and S. C. Bischoff, "The microalgae *Phaeodactylum tricorutum* is well suited as a food with positive effects on the intestinal microbiota and the generation of SCFA: Results from a pre-clinical study," *Nutrients*, vol. 14, no. 12, p. 2504, 2022.
- [193] R. Nagao, K. Kato, M. Kumazawa, K. Ifuku, M. Yokono, T. Suzuki, N. Dohmae, F. Akita, S. Akimoto, N. Miyazaki, and J.-R. Shen, "Structural basis for different types of hetero-tetrameric light-harvesting complexes in a diatom PSII-FCPII super-complex," *Nature Communications*, vol. 13, no. 1, p. 1764, 2022.



- [194] W. Wang, L.-J. Yu, C. Xu, T. Tomizaki, S. Zhao, Y. Umena, X. Chen, X. Qin, Y. Xin, M. Suga, G. Han, T. Kuang, and J.-R. Shen, "Structural basis for blue-green light harvesting and energy dissipation in diatoms," *Science*, vol. 363, no. 6427, eaav0365, 2019.
- [195] R. Croce and H. Van Amerongen, "Natural strategies for photosynthetic light harvesting," *Nature Chemical Biology*, vol. 10, no. 7, pp. 492–501, 2014.
- [196] N. Adir, H. Zer, S. Shochat, and I. Ohad, "Photoinhibition – a historical perspective," *Photosynthesis Research*, vol. 76, no. 1/3, pp. 343–370, 2003.
- [197] J. I. Carreto and J. A. Catoggio, "Variations in pigment contents of the diatom *Phaeodactylum tricornutum* during growth," *Marine Biology*, vol. 36, no. 2, pp. 105–112, 1976.
- [198] J. Lupette, A. Jaussaud, K. Seddiki, C. Morabito, S. Brugière, H. Schaller, M. Kuntz, J.-L. Putaux, P.-H. Jouneau, F. Rébeillé, D. Falconet, Y. Couté, J. Jouhet, M. Tardif, J. Salvaing, and E. Maréchal, "The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*," *Algal Research*, vol. 38, p. 101415, 2019.
- [199] A. Kawee-ai, A. Kuntiya, and S. M. Kim, "Anticholinesterase and antioxidant activities of fucoxanthin purified from the microalga *Phaeodactylum tricornutum*," *Natural Product Communications*, vol. 8, no. 10, 2013.
- [200] D. B. Rodriguez-Amaya and R. Carle, "Alterations of natural pigments," in *Chemical Changes During Processing and Storage of Foods*, Elsevier, 2021, pp. 265–327.
- [201] W. Terpstra, "Chlorophyllase in *Phaeodactylum tricornutum* photosynthetic membranes. Extractability, small-scale purification and molecular weight determination by SDS-gel-electrophoresis," *Physiologia Plantarum*, vol. 44, no. 4, pp. 329–334, 1978.
- [202] Y.-F. Yang, D.-W. Li, S. Balamurugan, X. Wang, W.-D. Yang, and H.-Y. Li, "Chrysolaminarin biosynthesis in the diatom is enhanced by overexpression of 1,6- $\beta$ -transglycosylase," *Algal Research*, vol. 66, p. 102817, 2022.
- [203] W. Huang, I. Haferkamp, B. Lepetit, M. Molchanova, S. Hou, W. Jeblick, C. Río Bártulos, and P. G. Kroth, "Reduced vacuolar  $\beta$ -1,3-glucan synthesis affects carbohydrate metabolism as well as plastid homeostasis and structure in *Phaeodactylum tricornutum*," *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, pp. 4791–4796, 2018.
- [204] T. Le Costaouëc, C. Unamunzaga, L. Mantecon, and W. Helbert, "New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*," *Algal Research*, vol. 26, pp. 172–179, 2017.

- [205] C. W. Ford and E. Percival, "1298. Carbohydrates of *Phaeodactylum tricornutum*. Part I. preliminary examination of the organism, and characterisation of low molecular weight material and of a glucan," *Journal of the Chemical Society*, pp. 7035–7041, 1965.
- [206] C. W. Ford and E. Percival, "1299. Carbohydrates of *Phaeodactylum tricornutum*. Part II. a sulphated glucuronomannan," *Journal of the Chemical Society*, pp. 7042–7046, 1965.
- [207] C.-Y. Chuang, P. H. Santschi, Y. Jiang, Y.-F. Ho, A. Quigg, L. Guo, M. Ayrarov, and D. Schumann, "Important role of biomolecules from diatoms in the scavenging of particle-reactive radionuclides of thorium, protactinium, lead, polonium, and beryllium in the ocean: A case study with *Phaeodactylum tricornutum*," *Limnology and Oceanography*, vol. 59, no. 4, pp. 1256–1266, 2014.
- [208] A. Mulgund, "Increasing lipid accumulation in microalgae through environmental manipulation, metabolic and genetic engineering: A review in the energy NEXUS framework," *Energy Nexus*, vol. 5, p. 100 054, 2022.
- [209] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, and A. Darzins, "Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances," *The Plant Journal*, vol. 54, no. 4, pp. 621–639, 2008.
- [210] L. Galas, C. Burel, D. Schapman, M. Ropitiaux, S. Bernard, M. Bénard, and M. Bardor, "Comparative structural and functional analyses of the fusiform, oval, and triradiate morphotypes of *Phaeodactylum tricornutum* Pt3 strain," *Frontiers in Plant Science*, vol. 12, p. 638 181, 2021.
- [211] Y.-H. Yang, L. Du, M. Hosokawa, K. Miyashita, Y. Kokubun, H. Arai, and H. Taro-da, "Fatty acid and lipid class composition of the microalga *Phaeodactylum tricornutum*," *Journal of Oleo Science*, vol. 66, no. 4, pp. 363–368, 2017.
- [212] E. Ryckebosch, C. Bruneel, R. Termote-Verhalle, K. Goiris, K. Muylaert, and I. Foubert, "Nutritional evaluation of microalgae oils rich in omega-3 long chain polyunsaturated fatty acids as an alternative for fish oil," *Food Chemistry*, vol. 160, pp. 393–400, 2014.
- [213] M. L. Hamilton, J. Warwick, A. Terry, M. J. Allen, J. A. Napier, and O. Sayanova, "Towards the industrial production of omega-3 long chain polyunsaturated fatty acids from a genetically modified diatom *Phaeodactylum tricornutum*," *PLOS ONE*, vol. 10, no. 12, A. Ianora, Ed., e0144054, 2015.

- [214] I. M. Remmers, D. E. Martens, R. H. Wijffels, and P. P. Lamers, "Dynamics of triacylglycerol and EPA production in *Phaeodactylum tricornutum* under nitrogen starvation at different light intensities," *PLOS ONE*, vol. 12, no. 4, A. Iancora, Ed., e0175630, 2017.
- [215] L. Grossmann, J. Hinrichs, and J. Weiss, "Solubility of extracted proteins from *Chlorella sorokiniana*, *Phaeodactylum tricornutum*, and *Nannochloropsis oceanica*: Impact of pH-value," *LWT*, vol. 105, pp. 408–416, 2019.
- [216] V. Aishvarya, J. Jena, N. Pradhan, P. K. Panda, and L. B. Sukla, "Microalgae: Cultivation and application," in *Environmental Microbial Biotechnology*, L. B. Sukla, N. Pradhan, S. Panda, and B. K. Mishra, Eds., vol. 45, Cham: Springer International Publishing, 2015, pp. 289–311.
- [217] J. Jerney and K. Spilling, "Large scale cultivation of microalgae: Open and closed systems," in *Biofuels from Algae*, K. Spilling, Ed., vol. 1980, New York, NY: Springer New York, 2018, pp. 1–8.
- [218] D. Chaumont, "Biotechnology of algal biomass production: A review of systems for outdoor mass culture," *Journal of Applied Phycology*, vol. 5, no. 6, pp. 593–604, 1993.
- [219] B. Aslanbay Guler, I. Deniz, Z. Demirel, S. S. Oncel, and E. Imamoglu, "Comparison of different photobioreactor configurations and empirical computational fluid dynamics simulation for fucoxanthin production," *Algal Research*, vol. 37, pp. 195–204, 2019.
- [220] R. R. L. Guillard and J. H. Ryther, "Studies of marine planktonic diatoms: I. *Cyclotella nana hustedt* and *Detonula confervacea* (Cleve) gran," *Canadian Journal of Microbiology*, vol. 8, no. 2, pp. 229–239, 1962.
- [221] J. E. Mann and J. Myers, "On pigments, growth, and photosynthesis of *Phaeodactylum tricornutum*," *Journal of Phycology*, vol. 4, no. 4, pp. 349–355, 1968.
- [222] G. Singh and S. Patidar, "Microalgae harvesting techniques: A review," *Journal of Environmental Management*, vol. 217, pp. 499–508, 2018.
- [223] L. Brennan and P. Owende, "Biofuels from microalgae - A review of technologies for production, processing, and extractions of biofuels and co-products," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 2, pp. 557–577, 2010.
- [224] G. M. Kim and Y.-K. Kim, "Drying techniques of microalgal biomass: A review," *Applied Chemistry for Engineering*, vol. 33, no. 2, pp. 145–150, 2022.
- [225] C.-L. Chen, J.-S. Chang, and D.-J. Lee, "Dewatering and drying methods for microalgae," *Drying Technology*, vol. 33, no. 4, pp. 443–454, 2015.

- [226] L. Lardon, A. Helias, B. Sialve, J. P. Steyer, and O. Bernard, "Life-cycle assessment of biodiesel production from microalgae," *Environ Sci Technol*, vol. 43, no. 17, pp. 6475–81, 2009. PMID: 19764204.
- [227] E. Günerken, E. D'Hondt, M. Eppink, L. Garcia-Gonzalez, K. Elst, and R. Wijffels, "Cell disruption for microalgae biorefineries," *Biotechnology Advances*, vol. 33, no. 2, pp. 243–260, 2015.
- [228] S. Ebert, L. Grossmann, J. Hinrichs, and J. Weiss, "Emulsifying properties of water-soluble proteins extracted from the microalgae *Chlorella sorokiniana* and *Phaeodactylum tricornutum*," *Food & Function*, vol. 10, no. 2, pp. 754–764, 2019.
- [229] M. Sørensen, K. Kousoulaki, R. Hammerø, M. Kokkali, D. Kleinegris, F. J. Marti-Quijal, F. J. Barba, A. M. Palihawadana, E. S. Egeland, C. A. Johnsen, O. H. Romarheim, S. Bisa, and V. Kiron, "Mechanical processing of *Phaeodactylum tricornutum* and *Tetraselmis chui* biomass affects phenolic and antioxidant compound availability, nutrient digestibility and deposition of carotenoids in Atlantic salmon," *Aquaculture*, vol. 569, p. 739 395, 2023.
- [230] T. O. Butler, G. Padmaperuma, A. M. Lizzul, J. McDonald, and S. Vaidyanathan, "Towards a *Phaeodactylum tricornutum* biorefinery in an outdoor UK environment," *Bioresource Technology*, vol. 344, p. 126 320, 2022.
- [231] G. P. 'T Lam, J. A. Van Der Kolk, A. Chordia, M. H. Vermuë, G. Olivieri, M. H. M. Eppink, and R. H. Wijffels, "Mild and selective protein release of cell wall deficient microalgae with pulsed electric field," *ACS Sustainable Chemistry & Engineering*, vol. 5, no. 7, pp. 6046–6053, 2017.
- [232] Y. Chisti, "Biodiesel from microalgae," *Biotechnology Advances*, vol. 25, no. 3, pp. 294–306, 2007.
- [233] T. M. Mata, A. A. Martins, and N. S. Caetano, "Microalgae for biodiesel production and other applications: A review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 217–232, 2010.
- [234] K. Sander and G. S. Murthy, "Life cycle analysis of algae biodiesel," *The International Journal of Life Cycle Assessment*, vol. 15, no. 7, pp. 704–714, 2010.
- [235] M. K. Lam and K. T. Lee, "Microalgae biofuels: A critical review of issues, problems and the way forward," *Biotechnology Advances*, vol. 30, no. 3, pp. 673–690, 2012.
- [236] S. Grierson, V. Strezov, and J. Bengtsson, "Life cycle assessment of a microalgae biomass cultivation, bio-oil extraction and pyrolysis processing regime," *Algal Research*, vol. 2, no. 3, pp. 299–311, 2013.

- [237] A. Westervelt, *Big oil firms touted algae as climate solution. Now all have pulled funding.* <https://www.theguardian.com/environment/2023/mar/17/big-oil-algae-biofuel-funding-cut-exxonmobil>, The Guardian, Mar 17, 2023, last accessed: Mar 11, 2024.
- [238] D. Prat, O. Pardigon, H.-W. Flemming, S. Letestu, V. Ducandas, P. Isnard, E. Guntrum, T. Senac, S. Ruisseau, P. Cruciani, and P. Hosek, "Sanofi's solvent selection guide: A step toward more sustainable processes," *Organic Process Research & Development*, vol. 17, no. 12, pp. 1517–1525, 2013.
- [239] D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehada, and P. J. Dunn, "CHEM21 selection guide of classical- and less classical-solvents," *Green Chemistry*, vol. 18, no. 1, pp. 288–296, 2016.
- [240] R. K. Henderson, C. Jiménez-González, D. J. C. Constable, S. R. Alston, G. G. A. Inglis, G. Fisher, J. Sherwood, S. P. Binks, and A. D. Curzons, "Expanding GSK's solvent selection guide – embedding sustainability into solvent selection starting at medicinal chemistry," *Green Chemistry*, vol. 13, no. 4, p. 854, 2011.
- [241] P. H. Chen and J. C. Quinn, "Microalgae to biofuels through hydrothermal liquefaction: Open-source techno-economic analysis and life cycle assessment," *Applied Energy*, vol. 289, p. 116613, 2021.
- [242] C. Tian, B. Li, Z. Liu, Y. Zhang, and H. Lu, "Hydrothermal liquefaction for algal biorefinery: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 933–950, 2014.
- [243] K. Pirwitz, L. Rihko-Struckmann, and K. Sundmacher, "Valorization of the aqueous phase obtained from hydrothermally treated *Dunaliella salina* remnant biomass," *Bioresource Technology*, vol. 219, pp. 64–71, 2016.
- [244] J.-M. Roux, H. Lamotte, and J.-L. Achard, "An overview of microalgae lipid extraction in a biorefinery framework," *Energy Procedia*, vol. 112, pp. 680–688, 2017.
- [245] J.-Y. Park, M. S. Park, Y.-C. Lee, and J.-W. Yang, "Advances in direct transesterification of algal oils from wet biomass," *Bioresource Technology*, vol. 184, pp. 267–275, 2015.
- [246] M. Vanthoor-Koopmans, R. H. Wijffels, M. J. Barbosa, and M. H. Eppink, "Biorefinery of microalgae for food and fuel," *Bioresource Technology*, vol. 135, pp. 142–149, 2013.
- [247] E. Angles, P. Jaouen, J. Pruvost, and L. Marchal, "Wet lipid extraction from the microalga *Nannochloropsis sp.*: disruption, physiological effects and solvent screening," *Algal Research*, vol. 21, pp. 27–34, 2017.

- [248] S. S. de Jesus, G. F. Ferreira, L. S. Moreira, M. R. Wolf Maciel, and R. Maciel Filho, "Comparison of several methods for effective lipid extraction from wet microalgae using green solvents," *Renewable Energy*, vol. 143, pp. 130–141, 2019.
- [249] P. Subramanian and C. Anandharamakrishnan, "Extraction of bioactive compounds," in *Industrial Application of Functional Foods, Ingredients and Nutraceuticals*, Elsevier, 2023, pp. 45–87.
- [250] E. G. Bligh and W. J. Dyer, "A rapid method of total lipid extraction and purification," *Canadian Journal of Biochemistry and Physiology*, vol. 37, no. 8, pp. 911–917, 1959.
- [251] J. Folch, M. Lees, and G. H. Sloane Stanley, "A simple method for the isolation and purification of total lipides from animal tissues," *The Journal of Biological Chemistry*, vol. 226, no. 1, pp. 497–509, 1957.
- [252] M. Zapata, F. Rodríguez, and J. Garrido, "Separation of chlorophylls and carotenoids from marine phytoplankton: A new HPLC method using a reversed phase C<sub>8</sub> column and pyridine-containing mobile phases," *Marine Ecology Progress Series*, vol. 195, pp. 29–45, 2000.
- [253] A. Jain, B. Behera, and B. Paramasivan, "Evaluation of physicochemical procedures for pigment extraction from mixed microalgal consortium," *Bioresource Technology Reports*, vol. 15, p. 100775, 2021.
- [254] K. H. Wiltshire, M. Boersma, A. Möller, and H. Buhtz, "Extraction of pigments and fatty acids from the green alga *Scenedesmus obliquus* (Chlorophyceae)," *Aquatic Ecology*, vol. 34, no. 2, pp. 119–126, 2000.
- [255] J.-C. Kim, "Solvent extraction of fucoxanthin from *Phaeodactylum tricornutum*," *Separation Science and Technology*, vol. 49, no. 3, pp. 410–415, 2014.
- [256] F. Ghasemi Naghdi, L. M. González González, W. Chan, and P. M. Schenk, "Progress on lipid extraction from wet algal biomass for biodiesel production," *Microbial Biotechnology*, vol. 9, no. 6, pp. 718–726, 2016.
- [257] S. Patel and D. C. Kannan, "A method of wet algal lipid recovery for biofuel production," *Algal Research*, vol. 55, p. 102237, 2021.
- [258] M. H. Eppink, S. P. Ventura, J. A. Coutinho, and R. H. Wijffels, "Multiproduct microalgae biorefineries mediated by ionic liquids," *Trends in Biotechnology*, vol. 39, no. 11, pp. 1131–1143, 2021.
- [259] Y. Zhu, X. Li, Y. Wang, L. Ren, and Q. Zhao, "Lutein extraction by imidazolium-based ionic liquid-water mixture from dried and fresh *Chlorella* sp.," *Algal Research*, vol. 60, p. 102528, 2021.

- [260] R. K. Desai, M. Streefland, R. H. Wijffels, and M. H. M. Eppink, "Novel astaxanthin extraction from *Haematococcus pluvialis* using cell permeabilising ionic liquids," *Green Chemistry*, vol. 18, no. 5, pp. 1261–1267, 2016.
- [261] V. C. A. Orr, N. V. Plechkova, K. R. Seddon, and L. Rehmann, "Disruption and wet extraction of the microalgae *Chlorella vulgaris* using room-temperature ionic liquids," *ACS Sustainable Chemistry & Engineering*, vol. 4, no. 2, pp. 591–600, 2016.
- [262] I. Fatima, M. Munir, R. Qureshi, U. Hanif, N. Gulzar, and A. A. Sheikh, "Advanced methods of algal pigments extraction: A review," *Critical Reviews in Food Science and Nutrition*, pp. 1–18, 2023.
- [263] C. Fan, Y. Liu, Y. Shan, and X. Cao, "A priori design of new natural deep eutectic solvent for lutein recovery from microalgae," *Food Chemistry*, vol. 376, p. 131930, 2022.
- [264] C. Samorì, C. Torri, G. Samorì, D. Fabbri, P. Galletti, F. Guerrini, R. Pistocchi, and E. Tagliavini, "Extraction of hydrocarbons from microalga *Botryococcus braunii* with switchable solvents," *Bioresource Technology*, vol. 101, no. 9, pp. 3274–3279, 2010.
- [265] D. Sanyal, G. Venkata Subhash, N. Saxena, W. Kargupta, A. Sapre, and S. Dasgupta, "Switchable green solvents for lipids extraction from microalgae," in *Green Sustainable Process for Chemical and Environmental Engineering and Science*, Elsevier, 2022, pp. 157–176.
- [266] L. D'Elia, P. Imbimbo, D. Liberti, F. Bolinesi, A. Pollio, O. Mangoni, W. Brillman, G. Olivieri, and D. M. Monti, "Switchable solvent selective extraction of hydrophobic antioxidants from *Synechococcus bigranulatus*," *ACS Sustainable Chemistry & Engineering*, vol. 9, no. 41, pp. 13798–13806, 2021.
- [267] Y. Du, B. Schuur, C. Samorì, E. Tagliavini, and D. W. F. Brillman, "Secondary amines as switchable solvents for lipid extraction from non-broken microalgae," *Bioresource Technology*, vol. 149, pp. 253–260, 2013.
- [268] R. L. Mendes, A. D. Reis, and A. F. Palavra, "Supercritical CO<sub>2</sub> extraction of  $\gamma$ -linolenic acid and other lipids from *Arthrospira (Spirulina) maxima*: Comparison with organic solvent extraction," *Food Chemistry*, vol. 99, no. 1, pp. 57–63, 2006.
- [269] B.-C. Liao, C.-T. Shen, F.-P. Liang, S.-E. Hong, S.-L. Hsu, T.-T. Jong, and C.-M. J. Chang, "Supercritical fluids extraction and anti-solvent purification of carotenoids from microalgae and associated bioactivity," *The Journal of Supercritical Fluids*, vol. 55, no. 1, pp. 169–175, 2010.
- [270] S. Tzima, I. Georgiopoulou, V. Louli, and K. Magoulas, "Recent advances in supercritical CO<sub>2</sub> extraction of pigments, lipids and bioactive compounds from microalgae," *Molecules*, vol. 28, no. 3, p. 1410, 2023.

- [271] C. Picot-Allain, M. F. Mahomoodally, G. Ak, and G. Zengin, "Conventional versus green extraction techniques — a comparative perspective," *Current Opinion in Food Science*, vol. 40, pp. 144–156, 2021.
- [272] M. Macías-Sánchez, C. Mantell, M. Rodríguez, E. Martínez De La Ossa, L. Lubián, and O. Montero, "Supercritical fluid extraction of carotenoids and chlorophyll a from *Nannochloropsis gaditana*," *Journal of Food Engineering*, vol. 66, no. 2, pp. 245–251, 2005.
- [273] K. Fujii, "Process integration of supercritical carbon dioxide extraction and acid treatment for astaxanthin extraction from a vegetative microalga," *Food and Bioproducts Processing*, vol. 90, no. 4, pp. 762–766, 2012.
- [274] C. Basheer, K. Alhooshani, A. Nuhu, S. Kanimozhi, and H. Lee, "Sample Preparation of Complex Biological Samples in the Analysis of Trace-Level Contaminants," in *Comprehensive Sampling and Sample Preparation*, Elsevier, 2012, pp. 681–700.
- [275] Q. Wang, K. Oshita, and M. Takaoka, "Effective lipid extraction from undewatered microalgae liquid using subcritical dimethyl ether," *Biotechnology for Biofuels*, vol. 14, no. 1, p. 17, 2021.
- [276] F. Derwenskus, B. Schäfer, J. Müller, K. Frick, A. Gille, K. Briviba, U. Schmid-Staiger, and T. Hirth, "Coproducts of EPA and fucoxanthin with *Phaeodactylum tricornutum* – a promising approach for up- and downstream processing," *Chemie Ingenieur Technik*, vol. 92, no. 11, pp. 1780–1789, 2020.
- [277] K. Ludwig, L. Rihko-Struckmann, G. Brinitzer, G. Unkelbach, and K. Sundmacher, " $\beta$ -Carotene extraction from *Dunaliella salina* by supercritical CO<sub>2</sub>," *Journal of Applied Phycology*, vol. 33, no. 3, pp. 1435–1445, 2021.
- [278] B. Chalermthai, A. Giwa, N. Moheimani, and H. Taher, "Techno-economic strategies for improving economic viability of  $\beta$ -carotene extraction using natural oil and supercritical solvent: A comparative assessment," *Algal Research*, vol. 68, p. 102875, 2022.
- [279] M. Mubarak, A. Shaija, and T. Suchithra, "A review on the extraction of lipid from microalgae for biodiesel production," *Algal Research*, vol. 7, pp. 117–123, 2015.
- [280] A. F. Ferreira, L. A. Ribeiro, A. P. Batista, P. A. Marques, B. P. Nobre, A. M. Palavra, P. P. Da Silva, L. Gouveia, and C. Silva, "A biorefinery from *Nannochloropsis* sp. microalga – energy and CO<sub>2</sub> emission and economic analyses," *Bioresource Technology*, vol. 138, pp. 235–244, 2013.



- [281] F. A. Ansari, A. Shrivastav, S. K. Gupta, I. Rawat, and F. Bux, "Exploration of microalgae biorefinery by optimizing sequential extraction of major metabolites from *Scenedesmus obliquus*," *Industrial & Engineering Chemistry Research*, vol. 56, no. 12, pp. 3407–3412, 2017.
- [282] S. Sadukha, B. Mehta, S. Chatterjee, A. Ghosh, and R. Dineshkumar, "Sequential Downstream Process for Concurrent Extraction of Lutein, Phytol, and Biochemicals from Marine Microalgal Biomass as a Sustainable Biorefinery," *ACS Sustainable Chemistry & Engineering*, vol. 11, no. 2, pp. 547–558, 2023.
- [283] J. M. Juárez, J. Vladoic, S. B. Rodríguez, and S. Vidovic, "Sequential valorisation of microalgae biomass grown in pig manure treatment photobioreactors," *Algal Research*, vol. 50, p. 101972, 2020.
- [284] I. Papachristou, S. Akaberi, A. Silve, E. Navarro-López, R. Wüstner, K. Leber, N. Nazarova, G. Müller, and W. Frey, "Analysis of the lipid extraction performance in a cascade process for *Scenedesmus almeriensis* biorefinery," *Biotechnology for Biofuels*, vol. 14, no. 1, p. 20, 2021.
- [285] C. A. Suarez Ruiz, J. Kwaijtaal, O. C. Peinado, C. van den Berg, R. H. Wijffels, and M. H. Eppink, "Multistep fractionation of microalgal biomolecules using selective aqueous two-phase systems," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 6, pp. 2441–2452, 2020.
- [286] K. Mayolo-Deloya, J. Benavides, and M. Rito-Palomares, "General concepts and definitions of aqueous two-phase systems," in *Aqueous Two-Phase Systems for Bioprocess Development for the Recovery of Biological Products*, M. Rito-Palomares and J. Benavides, Eds., Cham: Springer International Publishing, 2017, pp. 1–18.
- [287] Y. K. Leong, J. C.-W. Lan, H.-S. Loh, T. C. Ling, C. W. Ooi, and P. L. Show, "Thermoseparating aqueous two-phase systems: Recent trends and mechanisms," *Journal of Separation Science*, vol. 39, no. 4, pp. 640–647, 2016.
- [288] C. A. Suarez Ruiz, S. Z. Baca, L. A. Van Den Broek, C. Van Den Berg, R. H. Wijffels, and M. H. Eppink, "Selective fractionation of free glucose and starch from microalgae using aqueous two-phase systems," *Algal Research*, vol. 46, p. 101801, 2020.
- [289] C. A. Suarez Ruiz, D. P. Emmery, R. H. Wijffels, M. H. Eppink, and C. van den Berg, "Selective and mild fractionation of microalgal proteins and pigments using aqueous two-phase systems: Selective and mild fractionation of microalgal," *Journal of Chemical Technology & Biotechnology*, vol. 93, no. 9, pp. 2774–2783, 2018.

- [290] E. Suarez Garcia, C. A. Suarez Ruiz, T. Tilaye, M. H. Eppink, R. H. Wijffels, and C. van den Berg, "Fractionation of proteins and carbohydrates from crude microalgae extracts using an ionic liquid based-aqueous two phase system," *Separation and Purification Technology*, vol. 204, pp. 56–65, 2018.
- [291] "New Horizons in chemical space," *Nature Reviews Drug Discovery*, vol. 3, no. 5, pp. 375–375, 2004.
- [292] J.-L. Reymond, "The chemical space project," *Accounts of Chemical Research*, vol. 48, no. 3, pp. 722–730, 2015.
- [293] T. Sterling and J. J. Irwin, "ZINC 15 – ligand discovery for everyone," *Journal of Chemical Information and Modeling*, vol. 55, no. 11, pp. 2324–2337, 2015.
- [294] N. D. Austin, N. V. Sahinidis, and D. W. Trahan, "Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques," *Chemical Engineering Research and Design*, vol. 116, pp. 2–26, 2016.
- [295] R. M. Dias, L. C. G. Petrin, F. H. B. Sosa, A. M. Da Costa Lopes, J. A. P. Coutinho, and M. C. Da Costa, "Investigation of kraft lignin solubility in protic ionic liquids and their aqueous solutions," *Industrial & Engineering Chemistry Research*, vol. 59, no. 40, pp. 18 193–18 202, 2020.
- [296] R. P. Swatloski, S. K. Spear, J. D. Holbrey, and R. D. Rogers, "Dissolution of cellulose with ionic liquids," *Journal of the American Chemical Society*, vol. 124, no. 18, pp. 4974–4975, 2002.
- [297] R. Gani, "Chemical product design: Challenges and opportunities," *Computers & Chemical Engineering*, vol. 28, no. 12, pp. 2441–2457, 2004.
- [298] C. Gertig, K. Leonhard, and A. Bardow, "Computer-aided molecular and processes design based on quantum chemistry: Current status and future prospects," *Current Opinion in Chemical Engineering*, vol. 27, pp. 89–97, 2020.
- [299] O. Odele and S. Macchietto, "Computer aided molecular design: A novel method for optimal solvent selection," *Fluid Phase Equilibria*, vol. 82, pp. 47–54, 1993.
- [300] Z. Song, T. Zhou, Z. Qi, and K. Sundmacher, "Systematic method for screening ionic liquids as extraction solvents exemplified by an extractive desulfurization process," *ACS Sustainable Chemistry & Engineering*, vol. 5, no. 4, pp. 3382–3389, 2017.
- [301] Z. Song, X. Hu, H. Wu, M. Mei, S. Linke, T. Zhou, Z. Qi, and K. Sundmacher, "Systematic screening of deep eutectic solvents as sustainable separation media exemplified by the CO<sub>2</sub> capture process," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 23, pp. 8741–8751, 2020.

- [302] A. D. Sendek, Q. Yang, E. D. Cubuk, K.-A. N. Duerloo, Y. Cui, and E. J. Reed, "Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials," *Energy & Environmental Science*, vol. 10, no. 1, pp. 306–320, 2017.
- [303] Z. Song, T. Zhou, J. Zhang, H. Cheng, L. Chen, and Z. Qi, "Screening of ionic liquids for solvent-sensitive extraction –with deep desulfurization as an example," *Chemical Engineering Science*, vol. 129, pp. 69–77, 2015.
- [304] C. Balaji, T. Banerjee, and V. V. Goud, "COSMO-RS Based Predictions for the Extraction of Lignin from Lignocellulosic Biomass Using Ionic Liquids: Effect of Cation and Anion Combination," *Journal of Solution Chemistry*, vol. 41, no. 9, pp. 1610–1630, 2012.
- [305] E. C. Achinivu, M. Mohan, H. Choudhary, L. Das, K. Huang, H. D. Magurudeniya, V. R. Pidatala, A. George, B. A. Simmons, and J. M. Gladden, "A predictive toolset for the identification of effective lignocellulosic pretreatment solvents: A case study of solvents tailored for lignin extraction," *Green Chemistry*, vol. 23, no. 18, pp. 7269–7289, 2021.
- [306] J. Kahlen, K. Masuch, and K. Leonhard, "Modelling cellulose solubilities in ionic liquids using COSMO-RS," *Green Chemistry*, vol. 12, no. 12, p. 2172, 2010.
- [307] Y. Chu, X. Zhang, M. Hillestad, and X. He, "Computational prediction of cellulose solubilities in ionic liquids based on COSMO-RS," *Fluid Phase Equilibria*, vol. 475, pp. 25–36, 2018.
- [308] M. Mohan, C. Balaji, V. V. Goud, and T. Banerjee, "Thermodynamic insights in the separation of cellulose/hemicellulose components from lignocellulosic biomass using ionic liquids," *Journal of Solution Chemistry*, vol. 44, no. 3-4, pp. 538–557, 2015.
- [309] M. Mohan, P. Viswanath, T. Banerjee, and V. V. Goud, "Multiscale modelling strategies and experimental insights for the solvation of cellulose and hemicellulose in ionic liquids," *Molecular Physics*, vol. 116, no. 15-16, pp. 2108–2128, 2018.
- [310] P. Yamin, "COSMO-RS-based methods for improved modelling of complex chemical systems," Ph.D. dissertation, RWTH Aachen, Aachen, 2019.
- [311] A. Casas, J. Palomar, M. V. Alonso, M. Oliet, S. Omar, and F. Rodriguez, "Comparison of lignin and cellulose solubilities in ionic liquids by COSMO-RS analysis and experimental validation," *Industrial Crops and Products*, vol. 37, no. 1, pp. 155–163, 2012.
- [312] A. Casas, S. Omar, J. Palomar, M. Oliet, M. V. Alonso, and F. Rodriguez, "Relation between differential solubility of cellulose and lignin in ionic liquids and activity coefficients," *RSC Advances*, vol. 3, no. 10, p. 3453, 2013.

- [313] N. G. Chemmangattuvalappil, "Development of solvent design methodologies using computer-aided molecular design tools," *Current Opinion in Chemical Engineering*, vol. 27, pp. 51–59, 2020.
- [314] R. Gani and E. Brignole, "Molecular design of solvents for liquid extraction based on UNIFAC," *Fluid Phase Equilibria*, vol. 13, pp. 331–340, 1983.
- [315] P. M. Harper, R. Gani, P. Kolar, and T. Ishikawa, "Computer-aided molecular design with combined molecular modeling and group contribution," *Fluid Phase Equilibria*, vol. 158–160, pp. 337–347, 1999.
- [316] C. S. Adjiman, C. A. Schweiger, and C. A. Floudas, "Mixed-integer nonlinear optimization in process synthesis," in *Handbook of Combinatorial Optimization*, D.-Z. Du and P. M. Pardalos, Eds., Boston, MA: Springer US, 1998, pp. 1–76.
- [317] M. A. Duran and I. E. Grossmann, "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," *Mathematical Programming*, vol. 36, no. 3, pp. 307–339, 1986.
- [318] N. V. Sahinidis, M. Tawarmalani, and M. Yu, "Design of alternative refrigerants via global optimization," *AIChE Journal*, vol. 49, no. 7, pp. 1761–1775, 2003.
- [319] M. Sinha, L. E. Achenie, and G. M. Ostrovsky, "Environmentally benign solvent design by global optimization," *Computers & Chemical Engineering*, vol. 23, no. 10, pp. 1381–1394, 1999.
- [320] M. Sinha, L. E. K. Achenie, and R. Gani, "Blanket Wash Solvent Blend Design Using Interval Analysis," *Industrial & Engineering Chemistry Research*, vol. 42, no. 3, pp. 516–527, 2003.
- [321] A. I. Papadopoulos, I. Tsivintzelis, P. Linke, and P. Seferlis, "Computer-aided molecular design: Fundamentals, methods, and applications," in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, 2018, B9780124095472143422.
- [322] J. O. Spiegel and J. D. Durrant, "AutoGrow4: An open-source genetic algorithm for de novo drug design and lead optimization," *Journal of Cheminformatics*, vol. 12, no. 1, p. 25, 2020.
- [323] R. Laplaza, S. Gallarati, and C. Corminboeuf, "Genetic optimization of homogeneous catalysts," *Chemistry–Methods*, vol. 2, no. 6, e202100107, 2022.
- [324] T. Zhou, J. Wang, K. McBride, and K. Sundmacher, "Optimal design of solvents for extractive reaction processes," *AIChE Journal*, vol. 62, no. 9, pp. 3238–3249, 2016.
- [325] J. H. Jensen, "A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space," *Chemical Science*, vol. 10, no. 12, pp. 3567–3572, 2019.

- [326] A. K. Nigam, P. Friederich, M. Krenn, and A. Aspuru-Guzik, *Augmenting genetic algorithms with deep neural networks for exploring the chemical space*, 2019. arXiv: 1909.11655v4.
- [327] V. Venkatasubramanian, K. Chan, and J. Caruthers, "Computer-aided molecular design using genetic algorithms," *Computers & Chemical Engineering*, vol. 18, no. 9, pp. 833–844, 1994.
- [328] Z. Michalewicz and M. Schoenauer, "Evolutionary algorithms for constrained parameter optimization problems," *Evolutionary Computation*, vol. 4, no. 1, pp. 1–32, 1996.
- [329] D. P. Kingma and M. Welling, *Auto-encoding variational Bayes*, 2013. arXiv: 1312.6114v11.
- [330] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik, "Automatic chemical design using a data-driven continuous representation of molecules," *ACS Central Science*, vol. 4, no. 2, pp. 268–276, 2018.
- [331] Y. J. Lee, H. Kahng, and S. B. Kim, "Generative adversarial networks for de novo molecular design," *Molecular Informatics*, vol. 40, no. 10, p. 2100045, 2021.
- [332] G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, and A. Aspuru-Guzik, *Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models*, 2018. arXiv: 1705.10843v3.
- [333] F. Grisoni, M. Moret, R. Lingwood, and G. Schneider, "Bidirectional molecule generation with recurrent neural networks," *Journal of Chemical Information and Modeling*, vol. 60, no. 3, pp. 1175–1183, 2020.
- [334] A. E. Blanchard, C. Stanley, and D. Bhowmik, "Using GANs with adaptive training data to search for new molecules," *Journal of Cheminformatics*, vol. 13, no. 1, p. 14, 2021.
- [335] X. Liu, W. Zhang, X. Tong, F. Zhong, Z. Li, Z. Xiong, J. Xiong, X. Wu, Z. Fu, X. Tan, Z. Liu, S. Zhang, H. Jiang, X. Li, and M. Zheng, "MolFilterGAN: A progressively augmented generative adversarial network for triaging AI-designed molecules," *Journal of Cheminformatics*, vol. 15, no. 1, p. 42, 2023.
- [336] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360–365, 2018.
- [337] M. J. Kusner, B. Paige, and J. M. Hernández-Lobato, *Grammar variational autoencoder*, 2017. arXiv: 1703.01925v1.

- [338] H. Dai, Y. Tian, B. Dai, S. Skiena, and L. Song, *Syntax-directed variational autoencoder for structured data*, 2018. arXiv: 1802.08786v1.
- [339] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, "Deep learning for molecular design—a review of the state of the art," *Molecular Systems Design & Engineering*, vol. 4, no. 4, pp. 828–849, 2019.
- [340] X. Liu, A. P. IJzerman, and G. J. P. Van Westen, "Computational approaches for de novo drug design: Past, present, and future," in *Artificial Neural Networks*, H. Cartwright, Ed., vol. 2190, New York, NY: Springer US, 2021, pp. 139–165.
- [341] H. D. Baehr and S. Kabelac, *Thermodynamik: Grundlagen und technische Anwendungen* (Springer-Lehrbuch). Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 249–259.
- [342] C. Lüdecke and D. Lüdecke, "Thermodynamik: Physikalisch-chemische Grundlagen für Naturwissenschaftler und Ingenieure der thermischen Verfahrenstechnik," in Berlin, Heidelberg: Springer, 2020, pp. 703–705.
- [343] G. M. Wilson and C. H. Deal, "Activity coefficients and molecular structure. Activity coefficients in changing environments-solutions of groups," *Industrial & Engineering Chemistry Fundamentals*, vol. 1, no. 1, pp. 20–23, 1962.
- [344] H. Renon and J. M. Prausnitz, "Local compositions in thermodynamic excess functions for liquid mixtures," *AIChE Journal*, vol. 14, no. 1, pp. 135–144, 1968.
- [345] A. Fredenslund, J. Gmehling, and P. Rasmussen, *Vapor-Liquid Equilibria Using UNI-FAC: A Group Contribution Method*. New York: Elsevier, 1977, p. 380.
- [346] A. Fredenslund, R. L. Jones, and J. M. Prausnitz, "Group-contribution estimation of activity coefficients in nonideal liquid mixtures," *AIChE Journal*, vol. 21, no. 6, pp. 1086–1099, 1975.
- [347] G. Soave, "Equilibrium constants from a modified Redlich-Kwong equation of state," *Chemical Engineering Science*, vol. 27, no. 6, pp. 1197–1203, 1972.
- [348] D.-Y. Peng and D. B. Robinson, "A New Two-Constant Equation of State," *Industrial & Engineering Chemistry Fundamentals*, vol. 15, no. 1, pp. 59–64, 1976.
- [349] J. Gross and G. Sadowski, "Perturbed-chain SAFT: an equation of state based on a perturbation theory for chain molecules," *Industrial & Engineering Chemistry Research*, vol. 40, no. 4, pp. 1244–1260, 2001.
- [350] J. Gross and G. Sadowski, "Application of the perturbed-chain SAFT equation of state to associating systems," *Industrial & Engineering Chemistry Research*, vol. 41, no. 22, pp. 5510–5515, 2002.
- [351] *Dortmund Data Bank*, [www.ddbst.com](http://www.ddbst.com), last accessed: Apr 17, 2024, 2024.

- [352] Z. S. Baird, P. Uusi-Kyyny, J.-P. Pokki, E. Pedegert, and V. Alopaeus, "Vapor pressures, densities, and PC-SAFT parameters for 11 bio-compounds," *International Journal of Thermophysics*, vol. 40, no. 11, p. 102, 2019.
- [353] Z. Song, T. Zhou, Z. Qi, and K. Sundmacher, "Extending the UNIFAC model for ionic liquid–solute systems by combining experimental and computational databases," *AIChE Journal*, vol. 66, no. 2, e16821, 2020.
- [354] C. M. Hansen, *Hansen Solubility Parameters: A User's Handbook*, 2nd ed. Boca Raton: CRC Press, 2007, pp. 261–291.
- [355] S. Venkatram, C. Kim, A. Chandrasekaran, and R. Ramprasad, "Critical assessment of the hildebrand and hansen solubility parameters for polymers," *Journal of Chemical Information and Modeling*, vol. 59, no. 10, pp. 4188–4194, 2019.
- [356] A. K. Dwamena, "Recent advances in hydrophobic deep eutectic solvents for extraction," *Separations*, vol. 6, no. 1, 2019.
- [357] J. Ruwoldt, M. Tanase-Opedal, and K. Syverud, "Ultraviolet spectrophotometry of lignin revisited: Exploring solvents with low harmfulness, lignin purity, hansen solubility parameter, and determination of phenolic hydroxyl groups," *ACS Omega*, vol. 7, no. 50, pp. 46 371–46 383, 2022.
- [358] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, no. 6, pp. 1049–1070, 1926.
- [359] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review*, vol. 136, no. 3B, B864–B871, 1964.
- [360] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, "Best-practice DFT protocols for basic molecular computational chemistry," *Angewandte Chemie*, vol. 134, no. 42, e202205735, 2022.
- [361] B. Mennucci, "Continuum solvation models: What else can we learn from them?" *The Journal of Physical Chemistry Letters*, vol. 1, no. 10, pp. 1666–1674, 2010.
- [362] 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*, vol. 159, pp. 84–92, 2017.
- [363] T. J. Sheldon, M. Folić, and C. S. Adjiman, "Solvent design using a quantum mechanical continuum solvation model," *Industrial & Engineering Chemistry Research*, vol. 45, no. 3, pp. 1128–1140, 2006.
- [364] A. Lehmann and C. D. Maranas, "Molecular design using quantum chemical calculations for property estimation," *Industrial & Engineering Chemistry Research*, vol. 43, no. 13, pp. 3419–3432, 2004.

- [365] H. Weiß, P. Deglmann, P. J. In 'T Veld, M. Cetinkaya, and E. Schreiner, "Multiscale materials modeling in an industrial environment," *Annual Review of Chemical and Biomolecular Engineering*, vol. 7, no. 1, pp. 65–86, 2016.
- [366] A. Klamt and G. Schüürmann, "COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *J. Chem. Soc., Perkin Trans. 2*, no. 5, pp. 799–805, 1993.
- [367] *TURBOMOLE V7.3 2018, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007.*
- [368] F. Eckert and A. Klamt, "Fast solvent screening via quantum chemistry: COSMO-RS approach," *AIChE Journal*, vol. 48, no. 2, pp. 369–385, 2002.
- [369] A. Klamt, "Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena," *The Journal of Physical Chemistry*, vol. 99, no. 7, pp. 2224–2235, 1995.
- [370] A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz, "Refinement and parametrization of COSMO-RS," *The Journal of Physical Chemistry A*, vol. 102, no. 26, pp. 5074–5085, 1998.
- [371] Z. Song, H. Shi, X. Zhang, and T. Zhou, "Prediction of CO<sub>2</sub> solubility in ionic liquids using machine learning methods," *Chemical Engineering Science*, vol. 223, p. 115752, 2020.
- [372] Z. Ye and D. Ouyang, "Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms," *Journal of Cheminformatics*, vol. 13, no. 1, p. 98, 2021.
- [373] E. I. Sanchez Medina, S. Linke, M. Stoll, and K. Sundmacher, "Graph neural networks for the prediction of infinite dilution activity coefficients," *Digital Discovery*, vol. 1, no. 3, pp. 216–225, 2022.
- [374] E. I. Sanchez Medina, S. Linke, M. Stoll, and K. Sundmacher, "Gibbs–Helmholtz graph neural network: Capturing the temperature dependency of activity coefficients at infinite dilution," *Digital Discovery*, vol. 2, no. 3, pp. 781–798, 2023.
- [375] F. Jirasek, R. A. S. Alves, J. Damay, R. A. Vandermeulen, R. Bamler, M. Bortz, S. Mandt, M. Kloft, and H. Hasse, "Machine learning in thermodynamics: Prediction of activity coefficients by matrix completion," *The Journal of Physical Chemistry Letters*, vol. 11, no. 3, pp. 981–985, 2020.
- [376] B. Winter, C. Winter, J. Schilling, and A. Bardow, "A smile is all you need: Predicting limiting activity coefficients from SMILES with natural language processing," *Digital Discovery*, vol. 1, no. 6, pp. 859–869, 2022.



- [377] S. Lopez-Zamora, J. Kong, S. Escobedo, and H. D. Lasa, "Thermodynamics and machine learning based approaches for vapor–liquid–liquid phase equilibria in n-octane/water, as a naphtha–water surrogate in water blends," *Processes*, vol. 9, no. 3, p. 413, 2021.
- [378] L. David, A. Thakkar, R. Mercado, and O. Engkvist, "Molecular representations in AI-driven drug discovery: A review and practical guide," *Journal of Cheminformatics*, vol. 12, no. 1, p. 56, 2020.
- [379] K. Beckh, S. Müller, M. Jakobs, V. Toborek, H. Tan, R. Fischer, P. Welke, S. Houben, and L. von Rueden, "Explainable machine learning with prior knowledge: An overview," 2021.
- [380] B. Sanchez-Lengeling, J. Wei, B. Lee, E. Reif, P. Wang, W. Qian, K. McCloskey, L. Colwell, and A. Wiltschko, "Evaluating attribution for graph neural networks," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., 2020, pp. 5898–5910.
- [381] I. K. Adu, H. Sugiyama, U. Fischer, and K. Hungerbühler, "Comparison of methods for assessing environmental, health and safety (EHS) hazards in early phases of chemical process design," *Process Safety and Environmental Protection*, vol. 86, no. 2, pp. 77–93, 2008.
- [382] European Chemical Agency, *Registration, evaluation, authorisation and restriction of chemicals: REACH*, 2006.
- [383] A. S. Alshehri, A. K. Tula, F. You, and R. Gani, "Next generation pure component property estimation models: With and without machine learning techniques," *AIChE Journal*, vol. 68, no. 6, e17469, 2022.
- [384] E. Sanchez Medina, S. Linke, and K. Sundmacher, "Prediction of bioconcentration factors (BCF) using graph neural networks," in *Computer Aided Chemical Engineering*, vol. 50, Elsevier, 2021, pp. 991–997.
- [385] E. Benefenati, A. Manganaro, and G. Gini, "VEGA-QSAR: AI inside a platform for predictive toxicology," *CEUR Workshop Proceedings*, no. 1107, pp. 21–28, 2013.
- [386] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, pp. 31–36, 1988.
- [387] D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for generation of unique SMILES notation," *Journal of Chemical Information and Computer Sciences*, vol. 29, no. 2, pp. 97–101, 1989.

- [388] SMARTS - A language for describing molecular patterns, <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, last accessed: Feb 2, 2024, 2024.
- [389] R. Schmidt, E. S. R. Ehmki, F. Ohm, H.-C. Ehrlich, A. Mashychev, and M. Rarey, "Comparing molecular patterns using the example of SMARTS: Theory and algorithms," *Journal of Chemical Information and Modeling*, vol. 59, no. 6, pp. 2560–2571, 2019.
- [390] M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, "Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation," *Machine Learning: Science and Technology*, vol. 1, no. 4, p. 045 024, 2020.
- [391] M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, R. F. Lameiro, D. Lemm, A. Lo, S. M. Moosavi, J. M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk Von Rudorff, A. Wang, A. D. White, A. Young, R. Yu, and A. Aspuru-Guzik, "SELFIES and the future of molecular string representations," *Patterns*, vol. 3, no. 10, p. 100 588, 2022.
- [392] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: Moving beyond fingerprints," *Journal of Computer-Aided Molecular Design*, vol. 30, no. 8, pp. 595–608, 2016.
- [393] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande, "MoleculeNet: A benchmark for molecular machine learning," *Chemical Science*, vol. 9, no. 2, pp. 513–530, 2018.
- [394] L. König-Mattern, S. Linke, L. Rihko-Struckmann, and K. Sundmacher, "Computer-aided solvent screening for the fractionation of wet microalgae biomass," *Green Chemistry*, 10.1039.D1GC03471E, 2021.
- [395] L. König-Mattern, A. O. Komarova, A. Ghosh, S. Linke, L. K. Rihko-Struckmann, J. Luterbacher, and K. Sundmacher, "High-throughput computational solvent screening for lignocellulosic biomass processing," *Chemical Engineering Journal*, vol. 452, p. 139 476, 2023.
- [396] L. König-Mattern, E. I. Sanchez Medina, A. O. Komarova, S. Linke, L. Rihko-Struckmann, J. Luterbacher, and K. Sundmacher, "Machine learning-supported solvent design for lignin-first biorefineries and lignin upgrading," *Chemical Engineering Journal*, vol. 495, p. 153 524, 2024.
- [397] L. Moity, M. Durand, A. Benazzouz, C. Pierlot, V. Molinier, and J.-M. Aubry, "Panorama of sustainable solvents using the COSMO-RS approach," *Green Chemistry*, vol. 14, no. 4, pp. 1132–1145, 2012.

- [398] J. E. Camp, "Bio-available solvent cyrene: Synthesis, derivatization, and applications," *ChemSusChem*, vol. 11, no. 18, pp. 3048–3055, 2018.
- [399] Q. Zhang, K. De Oliveira Vigier, S. Royer, and F. Jérôme, "Deep eutectic solvents: Syntheses, properties and applications," *Chemical Society Reviews*, vol. 41, no. 21, pp. 7108–7146, 2012.
- [400] R. Kohli, "Applications of ionic liquids in removal of surface contaminants," in *Developments in Surface Contamination and Cleaning: Applications of Cleaning Techniques*, Elsevier, 2019, pp. 619–680.
- [401] E. Slupek, P. Makoś, and J. Gebicki, "Theoretical and economic evaluation of low-cost deep eutectic solvents for effective biogas upgrading to bio-methane," *Energies*, vol. 13, no. 13, p. 3379, 2020.
- [402] A. Paiva, R. Craveiro, I. Aroso, M. Martins, R. L. Reis, and A. R. C. Duarte, "Natural deep eutectic solvents – solvents for the 21st century," *ACS Sustainable Chemistry & Engineering*, vol. 2, no. 5, pp. 1063–1071, 2014.
- [403] K. Jeong, M. Yang, Y. Jin, E. Kim, J. Ko, and J. Lee, "Identification of major flavone C-glycosides and their optimized extraction from cymbidium kanran using deep eutectic solvents," *Molecules*, vol. 22, no. 11, p. 2006, 2017.
- [404] K. Radošević, N. Ćurko, V. Gaurina Srček, M. Cvjetko Bubalo, M. Tomašević, K. Kovačević Ganić, and I. Radojčić Redovniković, "Natural deep eutectic solvents as beneficial extractants for enhancement of plant extracts bioactivity," *LWT-Food Science and Technology*, vol. 73, pp. 45–51, 2016.
- [405] T. Křížek, M. Bursová, R. Horsley, M. Kuchař, P. Tůma, R. Čabala, and T. Hložek, "Menthol-based hydrophobic deep eutectic solvents: Towards greener and efficient extraction of phytocannabinoids," *Journal of Cleaner Production*, vol. 193, pp. 391–396, 2018.
- [406] M. W. Nam, J. Zhao, M. S. Lee, J. H. Jeong, and J. Lee, "Enhanced extraction of bioactive natural products using tailor-made deep eutectic solvents: Application to flavonoid extraction from *Flos sophorae*," *Green Chemistry*, vol. 17, no. 3, pp. 1718–1727, 2015.
- [407] C. Florindo, L. C. Branco, and I. M. Marrucho, "Quest for green-solvent design: From hydrophilic to hydrophobic (deep) eutectic solvents," *ChemSusChem*, vol. 12, no. 8, pp. 1549–1559, 2019.
- [408] W. Tang, Y. Dai, and K. H. Row, "Evaluation of fatty acid/alcohol-based hydrophobic deep eutectic solvents as media for extracting antibiotics from environmental water," *Analytical and Bioanalytical Chemistry*, vol. 410, no. 28, pp. 7325–7336, 2018.

- [409] H. Vanda, Y. Dai, E. G. Wilson, R. Verpoorte, and Y. H. Choi, "Green solvents from ionic liquids and deep eutectic solvents to natural deep eutectic solvents," *Comptes Rendus Chimie*, vol. 21, no. 6, pp. 628–638, 2018.
- [410] Iolitec, *Iolitec Catalogue*, [https://iolitec.de/products/ionic\\_liquids/catalogue](https://iolitec.de/products/ionic_liquids/catalogue), last accessed: Apr 20, 2020.
- [411] SigmaAldrich, "ChemFiles Vol. 5 No. 6," 2005.
- [412] Dassault Systèmes company, "COSMOtherm Reference Manual," 2019.
- [413] M. Hayyan, M. A. Hashim, A. Hayyan, M. A. Al-Saadi, I. M. AlNashef, M. E. Mirghani, and O. K. Saheed, "Are deep eutectic solvents benign or toxic?" *Chemosphere*, vol. 90, no. 7, pp. 2193–2195, 2013.
- [414] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, "PubChem 2023 update," *Nucleic Acids Research*, vol. 51, no. D1, pp. D1373–D1380, 2023.
- [415] G. Landrum, *RDKit: Open-source cheminformatics*, <https://www.rdkit.org>, last accessed: June 13, 2022.
- [416] J.-P. Ebejer, G. M. Morris, and C. M. Deane, "Freely available conformer generation methods: How good are they?" *Journal of Chemical Information and Modeling*, vol. 52, no. 5, pp. 1146–1158, 2012.
- [417] *COSMOtherm, Release 19; COSMOlogic GmbH & Co. KG, a Dassault Systèmes company.*
- [418] A. Klamt and F. Eckert, "COSMO-RS: A novel and efficient method for the a priori prediction of thermophysical data of liquids," *Fluid Phase Equilibria*, vol. 172, no. 1, pp. 43–72, 2000.
- [419] F. Bezold, M. E. Weinberger, and M. Minceva, "Computational solvent system screening for the separation of tocopherols with centrifugal partition chromatography using deep eutectic solvent-based biphasic systems," *Journal of Chromatography A*, vol. 1491, pp. 153–158, 2017.
- [420] H. F. Hizaddin, A. Ramalingam, M. A. Hashim, and M. K. Hadj-Kali, "Evaluating the performance of deep eutectic solvents for use in extractive denitrification of liquid fuels by the conductor-like screening model for real solvents," *Journal of Chemical & Engineering Data*, vol. 59, no. 11, pp. 3470–3487, 2014.
- [421] M. Diedenhofen and A. Klamt, "COSMO-RS as a tool for property prediction of IL mixtures—A review," *Fluid Phase Equilibria*, vol. 294, no. 1-2, pp. 31–38, 2010.
- [422] A. Hagberg, P. J. Swart, and D. A. Schult, "Exploring network structure, dynamics, and function using NetworkX," United States, 2008.

- [423] J. M. Schmidt, "A simple test on 2-vertex- and 2-edge-connectivity," *Information Processing Letters*, vol. 113, no. 7, pp. 241–244, 2013.
- [424] A. Kanevsky, "Finding all minimum-size separating vertex sets in a graph," *Networks*, vol. 23, no. 6, pp. 533–541, 1993.
- [425] A. Brandt, J. Gräsvik, J. P. Hallett, and T. Welton, "Deconstruction of lignocellulosic biomass with ionic liquids," *Green Chemistry*, vol. 15, no. 3, p. 550, 2013.
- [426] C. Li, B. Knierim, C. Manisseri, R. Arora, H. V. Scheller, M. Auer, K. P. Vogel, B. A. Simmons, and S. Singh, "Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification," *Bioresource Technology*, vol. 101, no. 13, pp. 4900–4906, 2010.
- [427] T. C. R. Brennan, S. Datta, H. W. Blanch, B. A. Simmons, and B. M. Holmes, "Recovery of sugars from ionic liquid biomass liquor by solvent extraction," *BioEnergy Research*, vol. 3, no. 2, pp. 123–133, 2010.
- [428] B. Pereira and V. Arantes, "Nanocelluloses from sugarcane biomass," in *Advances in Sugarcane Biorefinery*, Amsterdam, Netherlands: Elsevier, 2018, p. 181.
- [429] L. A. Alves, J. B. Almeida E Silva, and M. Giulietti, "Solubility of D-glucose in water and ethanol/water mixtures," *Journal of Chemical & Engineering Data*, vol. 52, no. 6, pp. 2166–2170, 2007.
- [430] S. Arora, N. Gupta, and V. Singh, "Choline based basic ionic liquid (BIL)/acidic DES mediated cellulose rich fractionation of agricultural waste biomass and valorization to 5-HMF," *Waste and Biomass Valorization*, vol. 11, no. 7, pp. 3345–3354, 2020.
- [431] K. Glińska, J. Gitalt, E. Torrens, N. Plechkova, and C. Bengoa, "Extraction of cellulose from corn stover using designed ionic liquids with improved reusing capabilities," *Process Safety and Environmental Protection*, vol. 147, pp. 181–191, 2021.
- [432] F. P. Burns, P.-A. Themens, and K. Ghandi, "Assessment of phosphonium ionic liquid-dimethylformamide mixtures for dissolution of cellulose," *Composite Interfaces*, vol. 21, no. 1, pp. 59–73, 2014.
- [433] P. Mäki-Arvela, I. Anugwom, P. Virtanen, R. Sjöholm, and J. Mikkola, "Dissolution of lignocellulosic materials and its constituents using ionic liquids—a review," *Industrial Crops and Products*, vol. 32, no. 3, pp. 175–201, 2010.
- [434] M. Zavrel, D. Bross, M. Funke, J. Büchs, and A. C. Spiess, "High-throughput screening for ionic liquids dissolving (ligno-)cellulose," *Bioresource Technology*, vol. 100, no. 9, pp. 2580–2587, 2009.
- [435] J. Vitz, T. Erdmenger, C. Haensch, and U. S. Schubert, "Extended dissolution studies of cellulose in imidazolium based ionic liquids," *Green Chemistry*, vol. 11, no. 3, p. 417, 2009.

- [436] Y.-L. Chen, X. Zhang, T.-T. You, and F. Xu, "Deep eutectic solvents (DESs) for cellulose dissolution: A mini-review," *Cellulose*, vol. 26, no. 1, pp. 205–213, 2019.
- [437] CDH Fine Chemicals, *18-Crown-6-Ether Safety Data Sheet*, [https://www.cdhfinechemical.com/images/product/msds/34\\_1958546453\\_18-Crown-6-Ether-CASNO-17455-13-9-MSDS.pdf](https://www.cdhfinechemical.com/images/product/msds/34_1958546453_18-Crown-6-Ether-CASNO-17455-13-9-MSDS.pdf), last accessed: Mar 18, 2024.
- [438] S. Jadhav, A. Lidhure, S. Thakre, and V. Ganvir, "Modified Lyocell process to improve dissolution of cellulosic pulp and pulp blends in NMMO solvent," *Cellulose*, vol. 28, no. 2, pp. 973–990, 2021.
- [439] A. S. Filatov, E. Block, and M. A. Petrukhina, "Dimethyl selenoxide," *Acta Crystallographica Section C Crystal Structure Communications*, vol. 61, no. 10, o596–o598, 2005.
- [440] P. Linstrom, *NIST Chemistry WebBook, NIST Standard Reference Database 69*, 1997.
- [441] P. Ertl and A. Schuffenhauer, "Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions," *Journal of Cheminformatics*, vol. 1, no. 1, p. 8, 2009.
- [442] J. V. Vermaas, M. F. Crowley, and G. T. Beckham, "Molecular lignin solubility and structure in organic solvents," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 48, pp. 17 839–17 850, 2020.
- [443] A. T. Smit, M. Verges, P. Schulze, A. Van Zomeren, and H. Lorenz, "Laboratory-to pilot-scale fractionation of lignocellulosic biomass using an acetone organosolv process," *ACS Sustainable Chemistry & Engineering*, vol. 10, no. 32, pp. 10 503–10 513, 2022.
- [444] M. Sundararajan, A. Taly, and Q. Yan, *Axiomatic attribution for deep networks*, 2017. arXiv: 1703.01365v2.
- [445] Z. Sumer and R. C. Van Lehn, "Heuristic computational model for predicting lignin solubility in tailored organic solvents," *ACS Sustainable Chemistry & Engineering*, vol. 11, no. 1, pp. 187–198, 2023.
- [446] L. Petridis and J. C. Smith, "Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy," *Nature Reviews Chemistry*, vol. 2, no. 11, pp. 382–389, 2018.
- [447] A. Das, A. Rahimi, A. Ulbrich, M. Alherech, A. H. Motagamwala, A. Bhalla, L. da Costa Sousa, V. Balan, J. A. Dumesic, E. L. Hegg, B. E. Dale, J. Ralph, J. J. Coon, and S. S. Stahl, "Lignin conversion to low-molecular-weight aromatics via an aerobic oxidation-hydrolysis sequence: Comparison of different lignin sources," *ACS Sustainable Chemistry & Engineering*, vol. 6, no. 3, pp. 3367–3374, 2018.

- [448] G. R. Dick, A. O. Komarova, and J. S. Luterbacher, "Controlling lignin solubility and hydrogenolysis selectivity by acetal-mediated functionalization," *Green Chemistry*, vol. 24, no. 3, pp. 1285–1293, 2022.
- [449] J. Leguy, T. Cauchy, M. Glavatskikh, B. Duval, and B. Da Mota, "EvoMol: A flexible and interpretable evolutionary algorithm for unbiased de novo molecular generation," *Journal of Cheminformatics*, vol. 12, no. 1, p. 55, 2020.
- [450] R. H. Wijffels, O. Kruse, and K. J. Hellingwerf, "Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae," *Current Opinion in Biotechnology*, vol. 24, no. 3, pp. 405–413, 2013.
- [451] L. Zhu, "Biorefinery as a promising approach to promote microalgae industry: An innovative framework," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1376–1384, 2015.
- [452] K. Wagemann and N. Tippkötter, Eds., *Biorefineries (Advances in Biochemical Engineering/Biotechnology)*. Cham: Springer International Publishing, 2019, vol. 166, pp. 99-125.
- [453] A. Steriti, R. Rossi, A. Concas, and G. Cao, "A novel cell disruption technique to enhance lipid extraction from microalgae," *Bioresour Technol*, vol. 164, pp. 70–7, 2014.
- [454] A. K. Lee, D. M. Lewis, and P. J. Ashman, "Disruption of microalgal cells for the extraction of lipids for biofuels: Processes and specific energy requirements," *Biomass and Bioenergy*, vol. 46, pp. 89–101, 2012.
- [455] L. König-Mattern, L. Rihko-Struckmann, and K. Sundmacher, "Systematic solvent selection enables the fractionation of wet microalgal biomass," *Separation and Purification Technology*, 2024, accepted.
- [456] K. Solymosi and B. Mysliwa-Kurdziel, "Chlorophylls and their derivatives used in food industry and medicine," *Mini-Reviews in Medicinal Chemistry*, vol. 17, no. 13, 2017.
- [457] J. A. Read, P. J. Beale, D. H. Volker, N. Smith, A. Childs, and S. J. Clarke, "Nutrition intervention using an eicosapentaenoic acid (EPA)-containing supplement in patients with advanced colorectal cancer. Effects on nutritional and inflammatory status: A phase II trial," *Supportive Care in Cancer*, vol. 15, no. 3, pp. 301–307, 2007.
- [458] C. Carballo, E. G. Chronopoulou, S. Letsiou, C. Maya, N. E. Labrou, C. Infante, D. M. Power, and M. Manchado, "Antioxidant capacity and immunomodulatory effects of a chrysolaminarin-enriched extract in Senegalese sole," *Fish & Shellfish Immunology*, vol. 82, pp. 1–8, 2018.

- [459] Y. Wang, J. Cai, Y. Jiang, X. Jiang, and D. Zhang, "Preparation of biosilica structures from frustules of diatoms and their applications: Current state and perspectives," *Applied Microbiology and Biotechnology*, vol. 97, no. 2, pp. 453–460, 2013.
- [460] A. Sardo, I. Orefice, S. Balzano, L. Barra, and G. Romano, "Mini-review: Potential of diatom-derived silica for biomedical applications," *Applied Sciences*, vol. 11, no. 10, p. 4533, 2021.
- [461] J. Grafmüller, A. Böhm, Y. Zhuang, S. Spahr, P. Müller, T. N. Otto, T. D. Bucheli, J. Leifeld, R. Giger, M. Tobler, H.-P. Schmidt, N. Dahmen, and N. Hagemann, "Wood ash as an additive in biomass pyrolysis: Effects on biochar yield, properties, and agricultural performance," *ACS Sustainable Chemistry & Engineering*, vol. 10, no. 8, pp. 2720–2729, 2022.
- [462] R. Halim, M. K. Danquah, and P. A. Webley, "Extraction of oil from microalgae for biodiesel production: A review," *Biotechnol Adv*, vol. 30, no. 3, pp. 709–32, 2012. PMID: 22266377.
- [463] A. Gugliuzza, "Solvent swollen polymer," in *Encyclopedia of Membranes*, E. Drioli and L. Giorno, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 1801–1802.
- [464] X. Pi, S. Zhao, W. Wang, D. Liu, C. Xu, G. Han, T. Kuang, S.-F. Sui, and J.-R. Shen, "The pigment-protein network of a diatom photosystem II–light-harvesting antenna supercomplex," *Science*, vol. 365, no. 6452, eaax4406, 2019.
- [465] B. Gügi, T. Le Costaouec, C. Burel, P. Lerouge, W. Helbert, and M. Bardor, "Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms," *Marine Drugs*, vol. 13, no. 9, pp. 5993–6018, 2015.
- [466] S. Guzmán, A. Gato, M. Lamela, M. Freire-Garabal, and J. M. Calleja, "Anti-inflammatory and immunomodulatory activities of polysaccharide from *Chlorella stigmatophora* and *Phaeodactylum tricornutum*," *Phytotherapy Research*, vol. 17, no. 6, pp. 665–670, 2003.
- [467] M. P. Andersson, J. H. Jensen, and S. L. S. Stipp, "Predicting pKa for proteins using COSMO-RS," *PeerJ*, vol. 1, e198, 2013.
- [468] "Hierarchical structure of proteins," in *Molecular Cell Biology*, ser. Media Connected, H. Lodish, Ed., 4th ed., New York, NY: Freeman, 2002.
- [469] "Properties of amino acids," in *CRC Handbook of Chemistry and Physics*, J. R. Rumble, Ed., 102 (Internet Version), Boca Raton, FL: CRC Press/Taylor & Francis, 2021.



- [470] J. Zhou, H. Sui, Z. Jia, Z. Yang, L. He, and X. Li, "Recovery and purification of ionic liquids from solutions: A review," *RSC Advances*, vol. 8, no. 57, pp. 32 832–32 864, 2018.
- [471] Y. Dai, J. van Spronsen, G.-J. Witkamp, R. Verpoorte, and Y. H. Choi, "Ionic liquids and deep eutectic solvents in natural products research: Mixtures of solids as extraction solvents," *Journal of Natural Products*, vol. 76, no. 11, pp. 2162–2173, 2013.
- [472] E. Lladosa, J. B. Montón, J. de la Torre, and N. F. Martínez, "Liquid-liquid and vapor-liquid-liquid equilibrium of the 2-butanone + 2-butanol + water system," *Journal of Chemical & Engineering Data*, vol. 56, no. 5, pp. 1755–1761, 2011.
- [473] M. Letcher, J. D. Sewry, and D. Naran, "Phase equilibria for mixtures of ethylether + water + an alcohol at 25°C," *Fluid Phase Equilibria*, pp. 187–193, 1989.
- [474] Y. Zhang, X. Guo, J. Xu, Y. Wu, and M. Lu, "Liquid–Liquid Equilibrium for Ternary Systems, Water + 5-Hydroxymethylfurfural + (1-Butanol, Isobutanol, Methyl Isobutyl Ketone), at 313.15, 323.15, and 333.15 K," *Journal of Chemical & Engineering Data*, vol. 63, no. 8, pp. 2775–2782, 2018.
- [475] S. Çehreli, D. Özmen, and U. Dramur, "(Liquid + liquid) equilibria of (water + 1-propanol + solvent) at T=298.2K," *Fluid Phase Equilibria*, vol. 239, no. 2, pp. 156–160, 2006.
- [476] M. Trofimova, A. Sadaev, A. Samarov, M. Toikka, and A. Toikka, "Solubility, liquid-liquid equilibrium and critical states for the quaternary system formic acid – ethanol – ethyl formate – water at 298.15 K and 308.15 K," *Fluid Phase Equilibria*, vol. 485, pp. 111–119, 2019.
- [477] R. M. Stephenson, "Mutual solubilities: Water-ketones, water-ethers, and water-gasoline-alcohols," *Journal of Chemical & Engineering Data*, vol. 37, no. 1, pp. 80–95, 1992.
- [478] C.-T. Hsieh, W.-Y. Ji, H.-m. Lin, and M.-J. Lee, "Multiphase equilibria for mixtures containing water, acetic acid, propionic acid, methyl acetate and methyl propionate," *Fluid Phase Equilibria*, vol. 271, no. 1-2, pp. 69–75, 2008.
- [479] CAMEO Chemicals, *Butylvinylether*; CAS: 111-34-2, <https://cameochemicals.noaa.gov/chemical/2745>, last accessed: Oct 17, 2023.
- [480] H. Zhang, G. Liu, C. Li, and L. Zhang, "Liquid–liquid equilibria of water + acetic acid + cyclopentyl methyl ether (CPME) system at different temperatures," *Journal of Chemical & Engineering Data*, vol. 57, no. 11, pp. 2942–2946, 2012.

- [481] D. Özmen, "(Liquid+liquid) equilibria of (water+propionic acid+dipropyl ether or diisopropyl ether) at T=298.2K," *The Journal of Chemical Thermodynamics*, vol. 39, no. 1, pp. 123–127, 2007.
- [482] I.-C. Hwang, J.-I. Kim, S.-J. Park, and S.-J. In, "Liquid–liquid equilibrium for ternary systems of propyl vinyl ether + C<sub>3</sub> or C<sub>4</sub> alcohols + water at 298.15 K and excess molar enthalpies for ternary and constituent binary systems of propyl vinyl ether + ethanol + isooctane at 303.15 K," *Journal of Chemical & Engineering Data*, vol. 53, no. 2, pp. 475–480, 2008.
- [483] C.-Z. Liu, S. Zheng, L. Xu, F. Wang, and C. Guo, "Algal oil extraction from wet biomass of *Botryococcus braunii* by 1,2-dimethoxyethane," *Applied Energy*, vol. 102, pp. 971–974, 2013.
- [484] S. Sarkar, M. S. Manna, T. K. Bhowmick, and K. Gayen, "Extraction of chlorophylls and carotenoids from dry and wet biomass of isolated *Chlorella thermophila*: Optimization of process parameters and modelling by artificial neural network," *Process Biochemistry*, vol. 96, pp. 58–72, 2020.
- [485] W. Zhang, F. Wang, B. Gao, L. Huang, and C. Zhang, "An integrated biorefinery process: Stepwise extraction of fucoxanthin, eicosapentaenoic acid and chrysolaminarin from the same *Phaeodactylum tricornutum* biomass," *Algal Research*, vol. 32, pp. 193–200, 2018.
- [486] X. Ren, C. Wei, Q. Yan, X. Shan, M. Wu, X. Zhao, and Y. Song, "Optimization of a novel lipid extraction process from microalgae," *Scientific Reports*, vol. 11, no. 1, p. 20 221, 2021.
- [487] S. Sridhar, B. Smitha, and A. A. Reddy, "Separation of 2-butanol–water mixtures by pervaporation through PVA–NYL 66 blend membranes," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 280, no. 1-3, pp. 95–102, 2006.
- [488] A. Meiser, U. Schmid-Staiger, and W. Trösch, "Optimization of eicosapentaenoic acid production by *Phaeodactylum tricornutum* in the flat panel airlift (FPA) reactor," *Journal of Applied Phycology*, vol. 16, no. 3, pp. 215–225, 2004.
- [489] G. Breuer, P. P. Lamers, D. E. Martens, R. B. Draaisma, and R. H. Wijffels, "Effect of light intensity, pH, and temperature on triacylglycerol (TAG) accumulation induced by nitrogen starvation in *Scenedesmus obliquus*," *Bioresource Technology*, vol. 143, pp. 1–9, 2013.
- [490] G. Breuer, W. A. C. Evers, J. H. De Vree, D. M. M. Kleinegris, D. E. Martens, R. H. Wijffels, and P. P. Lamers, "Analysis of fatty acid content and composition in microalgae," *Journal of Visualized Experiments*, no. 80, p. 50 628, 2013.

- [491] M. Szymczak-Żyła, G. Kowalewska, and J. W. Louda, "The influence of microorganisms on chlorophyll a degradation in the marine environment," *Limnology and Oceanography*, vol. 53, no. 2, pp. 851–862, 2008.
- [492] M. M. Maroneze, L. Q. Zepka, E. J. Lopes, A. Pérez-Gálvez, and M. Roca, "Chlorophyll oxidative metabolism during the phototrophic and heterotrophic growth of *Scenedesmus obliquus*," *Antioxidants*, vol. 8, no. 12, p. 600, 2019.
- [493] O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, "Protein measurement with the Folin phenol reagent," *The Journal of Biological Chemistry*, vol. 193, no. 1, pp. 265–275, 1951.
- [494] S. S. Nielsen, "Phenol-sulfuric acid method for total carbohydrates," in *Food Analysis Laboratory Manual*, S. S. Nielsen, Ed., Boston, MA: Springer US, 2010, pp. 47–53.
- [495] M. F. Monteiro, M. H. Moura-Neto, A. L. N. Mota, C. S. Figueiredo, J. R. P. Ciambelli, J. F. Do Nascimento, L. S. Pereira, and O. Chiavone-Filho, "Salt solubility and saturated electrical conductivity data for water + monoethylene glycol + NaCl in a wide temperature range," *Journal of Chemical & Engineering Data*, vol. 66, no. 4, pp. 1676–1683, 2021.
- [496] J. Kopton, L. K. Rihko-Struckmann, L. König-Mattern, and K. Sundmacher, "Superstructure optimization of a microalgal biorefinery design with life cycle assessment-based and economic objectives," *Biofuels, Bioproducts and Biorefining*, vol. 17, no. 6, pp. 1515–1527, 2023.
- [497] A. Ben-Naim, "Standard thermodynamics of transfer. Uses and misuses," *The Journal of Physical Chemistry*, vol. 82, no. 7, pp. 792–803, 1978.
- [498] A. Klamt, F. Eckert, M. Hornig, M. E. Beck, and T. Bürger, "Prediction of aqueous solubility of drugs and pesticides with COSMO-RS," *Journal of Computational Chemistry*, vol. 23, no. 2, pp. 275–281, 2002.
- [499] J. Sameni, S. Krigstin, and M. Sain, "Characterization of lignins isolated from industrial residues and their beneficial uses," *BioResources*, vol. 11, no. 4, pp. 8435–8456, 2016.
- [500] H. Zhao, G. A. Baker, Z. Song, O. Olubajo, T. Crittle, and D. Peters, "Designing enzyme-compatible ionic liquids that can dissolve carbohydrates," *Green Chemistry*, vol. 10, no. 6, p. 696, 2008.
- [501] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, *Neural message passing for quantum chemistry*, 2017. arXiv: 1704.01212.
- [502] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015. arXiv: 1502.03167v3.

- [503] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, *Weisfeiler and Leman go neural: Higher-order graph neural networks*, 2018. arXiv: 1810.02244v5.
- [504] M. An, J. Qiu, D. Yi, H. Liu, S. Hu, J. Han, H. Huang, H. He, C. Liu, Z. Zhao, Y. Shi, and P. Wang, "Measurement and correlation for solubility of L-alanine in pure and binary solvents at temperatures from 283.15 to 323.15 K," *Journal of Chemical & Engineering Data*, vol. 65, no. 2, pp. 549–560, 2020.
- [505] N. E. Craft and J. H. Soares, "Relative solubility, stability, and absorptivity of lutein and  $\beta$ -carotene in organic solvents," *Journal of Agricultural and Food Chemistry*, vol. 40, no. 3, pp. 431–434, 1992.
- [506] *Chemister Database*, <http://chemister.ru/Database/search-en.php>, last accessed: Oct 27, 2021.
- [507] M. Koller, A. Muhr, and G. Braunegg, "Microalgae as versatile cellular factories for valued products," *Algal Research*, vol. 6, pp. 52–63, 2014.
- [508] F. Derwenskus, S. Weickert, I. Lewandowski, U. Schmid-Staiger, and T. Hirth, "Economic evaluation of up- and downstream scenarios for the co-production of fucoxanthin and eicosapentaenoic acid with *Phaeodactylum tricornutum* using flat-panel airlift photobioreactors with artificial light," *Algal Research*, vol. 51, p. 102 078, 2020.
- [509] iHerb, *iHerb Liquid Chlorophyll*, <https://de.iherb.com/pr/now-foods-liquid-chlorophyll-natural-mint-16-fl-oz-473-ml/>, last accessed: May 15, 2023.

## List of Figures

1.1	Overview of a) a linear economy, and b) a circular bioeconomy. Adapted from Stegmann et al. <sup>22</sup> . . . . .	3
2.1	Lignocellulosic biomass with the chemical structures of its three main components: cellulose, hemicellulose (here exemplified by xylan), and lignin. These three biopolymers are located in the cell wall where cellulose fibrils are surrounded by a matrix of hemicellulose and lignin. These three main components form the lignin-carbohydrate complex. The structure of lignin is based on G-, S-, and H-units linked by bonding motifs as proposed by Ralph et al. <sup>60</sup>	10
2.2	Process flow diagram of organosolv pulping. As an alternative to filtration, centrifugation can be applied. . . . .	15
2.3	Overview over the AAF approach (lab-scale) based on Talebi Amiri et al. <sup>155</sup> a) Process flow diagram. b) Reaction mechanism of a lignin fragment with and without aldehyde. For recovering the aldehyde-stabilised xylose, multiple process units are applicable as indicated by the gray rectangle. c) Reaction mechanism of the hemicellulose fraction (represented by xylan) during AAF.	17
3.1	<i>P. tricornutum</i> (fusiform shape) and its potential biorefinery products. a) Scheme of a <i>P. tricornutum</i> cell (fusiform state). b) Chemical structures of Fx, EPA, and the $\beta$ -1,3 backbone of chrysolaminarin. . . . .	21
3.2	Process flow diagram for microalgal processing. After harvest, different process routes can be taken, such as HTL (purple), lipid extraction for biodiesel production (blue), or biomass fractionation in a biorefinery (green). . . . .	23
4.1	Scheme of molecular design as an inverse problem. . . . .	30
4.2	Workflow for thermodynamic property predictions using the COSMO-RS. a) COSMO-RS predictions rely on QC-based COSMO calculations that can be stored in a database. b) $\sigma$ -profiles and c) $\sigma$ -potentials of water, DMSO, and n-hexane. Adapted from Eckert and Klamt. <sup>368</sup> . . . . .	39
5.1	Overview of the solvent screening procedure. . . . .	46

5.2	Overview of the PSEvolve algorithm. . . . .	55
5.3	Deleting edges and nodes from the molecular graph. a) In the hexane graph, all edges are bridges. Therefore, the fragmentation algorithm could delete each bond to achieve the desired fragmentation of the graph (blue highlights). In contrast, in the bond deletion operation, fragmentation is not desired. Therefore, this operation is not applicable to the hexane graph. b) In the benzene graph, there are no bridges. Therefore, a pair of non-neighbouring edges must be selected to achieve a fragmentation, as exemplified by the blue highlights. In contrast, in the bond deletion operation, each bond could be chosen, since no fragmentation would occur. c) For atom deletion, a fragmentation of the graph is not desired. Therefore, only the oxygen-atoms attached to the branches (blue highlights) can be deleted. . . . .	58
6.1	Representative molecules investigated for the cellulose, hemicellulose, and lignin fraction. The abundance of each fraction within the biomass is given as wt.% on a dry matter basis. Adapted from König-Mattern et al. <sup>395</sup> with permission from Elsevier. . . . .	64
6.2	Numbers of eligible solvent candidates after each screening step. Reprinted from König-Mattern et al. <sup>395</sup> with permission from Elsevier. . . . .	67
6.3	Solubilities of the biomass fraction for all of the 3525 solvent candidates that were identified using the screening procedure. The datapoints are colour coded corresponding to their EHS score. a) Logarithmic molar solubilities of cellulose ( $\log_{10}(x_C)$ ), lignin ( $\log_{10}(x_L)$ ), and hemicellulose ( $\log_{10}(x_H)$ ). b) Logarithmic molar solubilities of cellulose ( $\log_{10}(x_C)$ ), lignin ( $\log_{10}(x_L)$ ). The locations of distinct solvent candidates are highlighted. Adapted and reprinted from König-Mattern et al. <sup>395</sup> with permission from Elsevier. . . .	68
6.4	Connecting the molecular design algorithm <i>PSEvolve</i> with a GNN for lignin solubility predictions. The genetic algorithm <i>PSEvolve</i> optimises the structure of a molecule tailored towards high lignin solubilities as predicted by the GNN. Adapted from König-Mattern et al. <sup>396</sup> with permission from Elsevier. . . .	75
6.5	GNN-training with COSMO-RS solubility data for the GGG-trimer which serves as a lignin-representing structure. a) Parity plot for GNN vs. COSMO-RS predictions of the training and the test set. The coefficient of determination $R^2$ and the MAE are given for the test set. b) Chemical classes of the training set as computed by the Classyfire toolbox. Adapted from König-Mattern et al. <sup>396</sup> with permission from Elsevier. . . . .	77

- 6.6 Application of the solvent design framework for lignin upgrading. a) T-SNE plot of the designed molecules with highlighted lignin solubility. b) Exploration of chemical space during molecule optimisation. c) Evolution of the molar lignin solubility during molecular optimisation. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier. . . . . 79
- 6.7 Application of the solvent design framework for AAF. a) T-SNE plot of the designed molecules with highlighted lignin solubility. b) Exploration of chemical space during molecule optimisation. c) Evolution of the molar lignin solubility during molecular optimisation. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier. . . . . 81
- 6.8 Experimental validation of the GNN predictions and attribution of structural features to the predicted lignin solubilities. a) Experimental lignin solubilities in the designed solvents for Kraft lignin, FABIOLA<sup>TM</sup> lignin, and MAL isolated from corn cob. Arrows indicate that lignin saturation was not yet reached, however, the high viscosity of the solution hindered measurements with higher lignin loadings. b) Normalised attributions for each discovered solvent class. A higher attribution score of the highlighted structural feature indicates higher importance for the lignin solubility prediction. Abbreviations: Dimethylsulfoxide (DMSO), diethylsulfoxide (DMSO), n,n-dimethylmethanesulfonamide (DMM-sulfonamide), dimethyl methylphosphonate (DMMP), diethyl methylphosphonate (DEMP), diethyl ethylphosphonate (DEMP), 5-bromo-1-methyl-1H-imidazole (5-Br-1-Me-1H-imidazole), 4-(2-hydroxyethyl)morpholine (4-(2-HE)morpholine), diethylene glycol dimethyl ether (DEGDME), diethylene glycol diethyl ether (DEGDDEE), 2-methyltetrahydrofuran (2-MeTHF). Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier. . . . . 85
- 6.9 Results of propionaldehyde-assisted pretreatment of birch wood ( $T = 85\text{ }^{\circ}\text{C}$ ,  $t = 3\text{ h}$ ). a) Cellulose-rich pulp after pretreatment with DEGDEE (above) and DMSO (below), b) DPX yield on raw biomass basis, c) lignin monomer yield. The yields are provided on a raw biomass basis. Adapted from König-Mattern et al.<sup>396</sup> with permission from Elsevier. . . . . 89
- 7.1 Biomass composition of moisture-free *P. tricornutum* on a) weight basis and b) based on economic value. The high-value components, EPA and fucoxanthin, render the lipid and pigment fraction the most valuable biomass components. Reprinted from König-Mattern et al.<sup>455</sup> with permission from Elsevier. 100
- 7.2 Representative algae molecules of the diatom *P. tricornutum* with their  $\sigma$ -surfaces. Adapted from König-Mattern et al.<sup>394</sup> with permission from the Royal Society of Chemistry. . . . . 106

7.3	Screening progress. As a first step, solvents meeting the structural constraints, $T_m$ and $T_b$ limits, and EHS criteria were identified. Subsequently, the solubilities of the biomass fractions in the identified solvents were predicted. Solvents with high solubilities for the lipophilic fractions (NLs, PLs, and carotenoids) were selected for predictions of their LLE with water ( $T = 25\text{ }^\circ\text{C}$ ). In this way, the identified solvents were classified as being WM, PWM, or NWM. . . . .	108
7.4	COSMO-RS predicted solubilities of NLs, PLs, and carotenoids in the selected solvents. . . . .	109
7.5	Effect of the water content in the organic phases ( $x_{\text{H}_2\text{O}}^{\text{org}}$ ) of the identified solvent on COSMO-RS predicted a) solubilities and b) partition coefficients. The solubility in water saturated solvents decreases for all lipophilic fraction with increasing water content, especially for neutral lipids. Partition coefficients are approaching zero with increasing water content. Adapted and reprinted from König-Mattern et al. <sup>394</sup> with permission from the Royal Society of Chemistry. . . . .	110
7.6	Lipid extraction ( $t = 90\text{ min}$ ) from wet <i>P. tricornutum</i> paste, with and without sonication treatment. Even without sonication, 2-BuOH <sub>75 vol.%</sub> , EtOH <sub>96 vol.%</sub> and acetone extracted more than 86 wt.% of the lipids. n-Hexane and all other practically water-immiscible solvents in contrast, attained yields below 33 wt.% with sonication treatment. Adapted from König-Mattern et al. <sup>455</sup> with permission from Elsevier. . . . .	115
7.7	Chlorophyll and carotenoid extraction ( $t = 90\text{ min}$ ) from wet <i>P. tricornutum</i> paste ( $MC = 81 - 85\text{ wt.}\%$ ), a) with sonication ( $t = 2\text{ min}$ ) and b) without sonication. Adapted from König-Mattern et al. <sup>455</sup> with permission from Elsevier. . . . .	116
7.8	Dry extractions ( $t = 90\text{ min}$ ) of freeze-dried <i>P. tricornutum</i> biomass (moisture content = 8.5 wt.%) compared to wet <i>P. tricornutum</i> paste (moisture content = 81 - 85 wt.%). The ratio of algal dry matter to solvent was held constant. The yields of carotenoids, chlorophylls and lipids are shown in a), b), and c), respectively. Adapted from König-Mattern et al. <sup>455</sup> with permission from Elsevier. . . . .	117
7.9	Influence of the water content $w_{\text{water,extr}}$ and the incubation time $t$ on lipid, carotenoid, and chlorophyll yields during extraction of wet <i>P. tricornutum</i> biomass ( $MC = 81 - 85\text{ wt.}\%$ , no sonication). a) Varying $w_{\text{water,extr}}$ at constant incubation time ( $t = 90\text{ min}$ ), b) varying incubation time $t$ at constant water content ( $w_{\text{water,extr}} = 32\text{ wt.}\%$ ). Adapted from König-Mattern et al. <sup>455</sup> with permission from Elsevier. . . . .	120



7.10	Overview of the developed biorefinery process for wet <i>P. tricornutum</i> biomass. a) Process flow diagram. b) Corresponding mass flows modelled on the basis of experimental measurements. The final yields are based on the weight of the fractions in the initial biomass. Reprinted from König-Mattern et al. <sup>455</sup> with permission from Elsevier. . . . .	122
C.1	$\sigma$ -profiles for representative molecules of a) cellulose, b) lignin, and c) hemi-cellulose. Reprinted from <sup>395</sup> with permission from Elsevier. . . . .	147
C.2	Comparison of COSMO-RS and GNN lignin solubility predictions for solvent structures not included in the training set ( $T = 70\text{ }^{\circ}\text{C}$ ). Reprinted from König-Mattern et al. <sup>395</sup> with permission from Elsevier. . . . .	154
C.3	Parity plot of the classification GNN. Classes were „promising“and „non-promising“solvent, based on a threshold lignin solubility of $\log(x_L) = -1.5$ . . . . .	157
C.4	2D-HSQC-NMR spectra of different lignin types used in this work. a) Birch MAL, b) corn cob MAL, c) FABIOLA™lignin from beechwood, and d) Kraft lignin from softwood species obtained from Berner Fachhochschule. . . . .	158
C.5	2D-HSQC-NMR spectra of aldehyde-stabilised lignin obtained from AAF using the benchmark solvent a) 1,4-dioxane, and the designed solvents b) diglyme (DEGDME), c) DEGDEE, and d) butyl sulfone. . . . .	159
D.1	Molecular structure of pheophytin a, the chlorophyll-representing molecule used in the COSMO-RS solubility predictions for 2-BuOH/water. . . . .	165
D.2	Linear regression of experimental solubility measurements and COSMO-RS solubilities for a) D-glucose, b) palmitic acid, c) L-alanine, and d) $\beta$ -carotene. Experimental values were taken from literature <sup>504–506</sup> . . . . .	166
D.3	$\sigma$ -profiles of the representative microalgal molecules: a) carbohydrates, b) amino acids representing the protein fraction, c) fucoxanthin representing the carotenoids, and d) lipids. . . . .	167
D.4	$\sigma$ -profiles of the solvents identified by the screening approach. . . . .	169
D.5	COSMO-RS predicted solubilities for PLs, NLs, chlorophylls and carotenoids for 2-BuOH/water mixtures. Reprinted from König-Mattern et al. <sup>455</sup> . . . . .	170
D.6	Fatty acid composition of the <i>P. tricornutum</i> biomass used in this work (dried biomass, extracted with chloroform/methanol, 1/1, v/v). Reprinted from König-Mattern et al. <sup>455</sup> . . . . .	170
D.7	GC-FID chromatogram of a lipid extract obtained from dried <i>P. tricornutum</i> using chloroform/methanol (1/1, v/v). The lipids were converted to FAMES prior to GC analysis. . . . .	172

---

D.8 Chlorophyll and carotenoid composition of the <i>P. tricornutum</i> biomass used in this work (dried biomass, extracted with EtOH <sub>96 vol.%</sub> ). Reprinted from König-Mattern et al. <sup>455</sup> . . . . .	172
D.9 HPLC chromatogram of a pigment extract obtained from dried <i>P. tricornutum</i> biomass using EtOH <sub>96 vol.%</sub> . . . . .	173
D.10 Microscopy images of <i>P. tricornutum</i> . a) <i>P. tricornutum</i> in M+M medium in the final days of cultivation. b) <i>P. tricornutum</i> treated with ultrasound probe (2 min) in water. c) Thawed <i>P. tricornutum</i> biomass 500x magnified, and d) 1000x magnified. e) Thawed biomass after extraction with water after 90 min incubation similar to extraction experiments 500 x magnified and f) 1000x magnified. Reprinted from König-Mattern et al. <sup>455</sup> . . . . .	174

## List of Tables

5.1	QSAR models from VEGA applied for the prediction of selected EHS properties.	50
6.1	Ranking of ILs for the joint dissolution of cellulose, lignin and hemicellulose. The distance from the optimal point $\mathbf{o}(\mathbf{t}, \mathbf{e}) = [x_C, x_L, x_H, \text{EHS score}]^T = [1, 1, 1, 1]^T$ is given as $d(\mathbf{o}, \mathbf{i})$ . Solubilities for each biomass fraction are given in logarithmic scale. The $T_m$ was obtained from lolitec, <sup>410</sup> $T_b$ was unknown. RT indicates room temperature. EHS properties were taken from PubChem. <sup>414</sup>	69
6.2	Solvent ranking for joint lignin and cellulose dissolution. EHS criteria and hemicellulose solubilities were neglected in the ranking. The distance from the optimal point $\mathbf{o}(\mathbf{t}) = [x_C, x_L]^T = [1, 1]^T$ was zero for all solvents presented in the table. Unless otherwise stated, all properties were obtained from the respective safety data sheets.	70
6.3	Solvent ranking for selective lignin extraction. The distance from the optimal point $\mathbf{o}(\mathbf{t}) = [x_C, x_L]^T = [1, 0]^T$ is given as $d(\mathbf{o}, \mathbf{i})$ . Only solvent candidates with $d(\mathbf{o}, \mathbf{i}) \leq 0.2$ are shown. Unless otherwise indicated, $T_m$ , $T_b$ , and EHS properties were obtained from the respective SDSs.	71
7.1	Representative molecules according to their fraction in the algal biomass of <i>P. tricornutum</i> and their reference solvents. Later in the screening, the solubilities of representative molecules in the solvent candidates is compared to the reference solvents. A solvent candidate will be eliminated in case of lower solubility than the reference.	105
7.2	Overview of the PWM solvents identified using the computational screening approach. The $T_b$ was taken from the NIST Chemistry WebBook. <sup>440</sup> The solubilities $\log_{10}(x_{\text{frac}})$ of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem. <sup>414</sup>	110

7.3	Overview of the NWM solvents identified using the computational screening approach. The $T_b$ was taken from the NIST Chemistry WebBook. <sup>440</sup> The solubilities $\log_{10}(x_{\text{frac}})$ of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem. <sup>414</sup> . . . . .	112
7.4	Overview of the WM solvents identified using the computational screening approach. The $T_b$ was taken from the NIST Chemistry WebBook. <sup>440</sup> The solubilities $\log_{10}(x_{\text{frac}})$ of the NL, PL, and carotenoid (Car.) fraction were predicted using COSMO-RS. The EHS properties were obtained from PubChem. <sup>414</sup> . . . . .	113
C.1	Chemical structures of the representative molecules evaluated for the computational screening. . . . .	143
C.2	Correlation coefficients $R^2$ from linear regressions between solubilities for lignin from different sources and COSMO-RS predictions for lignin monomers, dimers, trimers, the $1,500 \text{ g mol}^{-1}$ lignin fragment and the average of all monomers, dimers, and trimers. Experimental data was taken from <sup>101</sup> . . . . .	149
C.3	Hyperparameter test for <i>PSEvolve</i> . Selected values are marked as bold numbers.	150
C.4	Hyperparameters and constraints applied in the solvent design for high lignin solubilities and AAF using <i>PSEvolve</i> . . . . .	151
C.5	Atom and bond features considered in the GNN. . . . .	153
C.6	Parameter test for GNN solubility predictions. The selected parameters are marked with bold numbers. . . . .	153
C.7	COSMO-RS lignin solubility predictions for selected solvents at $T = 25 \text{ }^\circ\text{C}$ and $70 \text{ }^\circ\text{C}$ . . . . .	155
C.8	Experimental solubility data for Kraft, FABIOLA, as well as birch and corn cob MAL obtained by the GC method. Values in brackets correspond to additional datapoints obtained by the evaporation method. The asterisk indicates that the solubility limit in the corresponding solvent was reached. N/M corresponds to „not measured“. Solvents for which saturated lignin solutions were obtained are marked by an asterisk (*). . . . .	160
C.9	Biomass composition of the birch wood used for AAF experiments. . . . .	161
C.10	Experimental data of DPX and lignin monomer yields obtained from AAF experiments. N/M corresponds to „not measured“. . . . .	161
D.1	Representative molecules for <i>P. tricornutum</i> and their molecular structures. . . . .	163
D.2	Solvent candidates that were falsely excluded in the EHS property screening step and were manually re-added to the solvent list. . . . .	168

---

D.3	Biomass composition of the <i>P. tricornutum</i> applied in the extraction experiments. All values are given on a dry matter basis. . . . .	171
D.4	Economic value of potential biorefinery products of <i>P. tricornutum</i> . The values were taken from Ruiz et al <sup>165</sup> unless otherwise stated. . . . .	175
D.5	Experimentally measured $\Phi_{p,c}$ -factors to calculate the mass flows of the developed 2-BuOH-based biorefinery process. . . . .	177



## Afterword

Several journal articles were published in the course of preparing and writing this thesis. The contributions of each individual article is provided in the following list:

- The description of the biomass composition of lignocellulosic and microalgal biomass in Chapter 2 and 3 is partially based on the publications of König-Mattern et al.<sup>394</sup> and König-Mattern et al.<sup>395</sup>, respectively.
- The solvent screening methodology presented in Chapter 5 is based on König-Mattern et al.<sup>394,395</sup>.
- The solvent design methodology presented in Chapter 5 was developed by König-Mattern et al.<sup>396</sup>.
- The application of the screening methodology to lignocellulosic biomass in Chapter 6 and to microalgal biomass in Chapter 7 is based on König-Mattern et al.<sup>395</sup>, König-Mattern et al.<sup>394</sup>, and König-Mattern et al.<sup>455</sup> respectively.
- The tailor-made solvent design method for dissolution-based lignin upgrading and aldehyde-assisted fractionation in Chapter 6 is based on König-Mattern et al.<sup>396</sup>.
- The experimental lignin solubility measurements included in Chapter 6 are based on König-Mattern et al.<sup>396</sup>. The experiments investigating the applicability of selected solvents in aldehyde-assisted fractionation were performed as part of König-Mattern et al.<sup>396</sup>.
- The experiments on microalgal biomass, and the development of a lab-scale biorefinery process for wet microalgal biomass included in Chapter 7 are part of König-Mattern et al.<sup>455</sup> which is published as a preprint and is submitted for peer review.
- Chapter 8 provides an outlook to connect the solvent selection with the overall process level for optimal biorefinery design *via* superstructure optimisation which was published in Kopton et al.<sup>496</sup> This publication stems from a former master student who prepared his thesis under my supervision.

## List of publications

1. **L. König-Mattern**, S. Linke, L. Rihko-Struckmann, and K. Sundmacher, „Computer-aided solvent screening for the fractionation of wet microalgae biomass“, *Green Chemistry*, 10.1039.D1GC03471E, 2021.

L. König-Mattern developed the screening approach, performed the experiments, analysed the data, visualised the data, and wrote the manuscript.

2. **L. König-Mattern\***, A. O. Komarova\*, A. Ghosh, S. Linke, L. K. Rihko-Struckmann, J. Luterbacher, and K. Sundmacher, „High-throughput computational solvent screening for lignocellulosic biomass processing“, *Chemical Engineering Journal*, vol. 452, p. 139 476, 2023.

\* Authors contributed equally. L. König-Mattern developed the screening approach, analysed the data, visualised the data, and wrote the manuscript.

3. J. Kopton, L. K. Rihko-Struckmann, **L. König-Mattern**, and K. Sundmacher, „Superstructure optimization of a microalgal biorefinery design with life cycle assessment-based and economic objectives“, *Biofuels, Bioproducts and Biorefining*, vol. 17, no. 6, pp. 1515–1527, 2023.

L. König-Mattern supervised the development of the optimisation approach, analysed the data, wrote parts of the manuscript, and revised the manuscript.

4. **L. König-Mattern**, E. I. Sanchez Medina, A. O. Komarova, S. Linke, L. Rihko-Struckmann, J. Luterbacher, and K. Sundmacher, „Machine learning-supported solvent design for lignin-first biorefineries and lignin upgrading“, *Chemical Engineering Journal*, vol. 495, p. 153 524, 2024.

L. König-Mattern developed the solvent design methodology, performed parts of the solubility experiments, analysed the data, visualised the data, and wrote the manuscript.

5. **L. König-Mattern**, L. Rihko-Struckmann, and K. Sundmacher, „Systematic solvent selection enables the fractionation of wet microalgal biomass“, *Separation and Purification Technology*, 2024, accepted.

L. König-Mattern performed the solvent screening, performed the experiments, analysed the data, visualised the data, and wrote the manuscript.



6. E. Matthies, K. Beer, M. Böcher, K. Sundmacher, **L. König-Mattern**, J. Arlinghaus, A. Blöbaum, M. Jaeger-Erben, and K. Schmidt, "Framework conditions for the transformation toward a sustainable carbon-based chemical industry – a critical review of existing and potential contributions from the social sciences," *Journal of Cleaner Production*, vol. 470, p. 143 279, 2024.

L. König-Mattern wrote parts of the manuscript.



## Colophon

The writing style of this dissertation benefited from the adoption of two language processing tools, ChatGPT (GPT-3.5, <https://chat.openai.com/>) developed by OpenAI, and DeepL Write (<https://www.deepl.com/write>). These tools aided in enhancing the structure, language, and grammar of the originally authored paragraphs. Afterwards, each paragraph underwent human proofreading by the author followed by subsequent editing. This thesis was typeset with  $\LaTeX$ . The MastersDoctoralThesis style developed by Sunil Patel and published under the License CC BY-NC-SA 3.0 was used and modified. Download the template from <http://www.sunilpatel.co.uk/thesis-template/>.



## Declaration of Honour

I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used. In the case of co-authorship, especially in the context of a cumulative dissertation, the own contribution is correctly and completely stated. I did not make use of any commercial consultant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis. All sources of information are clearly marked, including my own publications. In particular I have not consciously:

- Fabricated data or rejected undesired results
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data
- Plagiarized data or publications
- Presented the results of other researchers in a distorted way

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities.

I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism.

This work has not yet been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not yet been published as a whole.

Magdeburg, September 6, 2024

Laura König-Mattern

