



High-throughput computational solvent screening for lignocellulosic biomass processing

Laura König-Mattern^{a,1}, Anastasia O. Komarova^{b,1}, Arpa Ghosh^b, Steffen Linke^a, Liisa K. Rihko-Struckmann^a, Jeremy Luterbacher^{b,*}, Kai Sundmacher^{a,c,**}

^a Max Planck Institute for Dynamics of Complex Technical Systems, Process Systems Engineering, Sandtorstraße 1, Magdeburg, 39106, Germany

^b École Polytechnique Fédérale de Lausanne, Laboratory of Sustainable and Catalytic Processing, Station 6, Lausanne, 1015, Switzerland

^c Otto von Guericke University Magdeburg, Chair for Process Systems Engineering, Universitätsplatz 2, Magdeburg, 39106, Germany

ARTICLE INFO

Keywords:

Lignocellulose
Biomass fractionation
Green solvents
Computer-aided solvent selection
COSMO-RS

ABSTRACT

Lignocellulose is one of the most promising renewable bioresources for the production of chemicals. For sustainable and competitive biorefineries, effective valorization of all biomass fractions is crucial. However, current efforts in lignocellulose fractionation are limited by the use of either toxic or suboptimal solvents that do not always allow producing clean and homogeneous streams. Here, we present a computational screening approach that covers more than 8000 solvent candidates for the processing of lignocellulosic biomass. The automated screening identified highly effective, non-intuitive solvents based on physico-chemical properties, solubilities of the biomass fractions, and environmental, health and safety properties. Solubility experiments for the lignin and cellulose fraction confirmed the applicability of the proposed framework in biomass processing. In addition to the traditional “lignin-first” approaches, we identified solvents applicable for the complete dissolution of biomass. Furthermore, we elucidated particular structural patterns in solvents featuring high lignin solubility. The most promising solvents attained lignin solubilities of more than 33 wt%.

1. Introduction

Climate change and the dependence on fossil resources are driving the development of sustainable processes that are based on renewable resources. The most abundant terrestrial source of renewable carbon is lignocellulosic biomass, consisting of three major fractions: cellulose (30%–50%, dry weight basis), hemicellulose (20%–30%), and lignin (15%–30%) [1,2]. To maximize value and sustainability, the biomass should be used in a holistic manner where each fraction is separated and used for the production of valuable chemicals. In contrast to this vision, the lignin fraction was long seen as a side-product of the pulp and paper industry and was mainly used for heat production due to its high heating value [3]. Currently, the paradigm is shifting towards “lignin-first” approaches where lignin is considered at the

front end of the process, which is designed to facilitate lignin separation and its catalytical conversion into aromatics at high yields and selectivities [4,5]. Carbohydrates can also be fractionated and depolymerized into their constituent monomers: glucose for cellulose and mainly xylose for hemicellulose. These carbohydrates can subsequently serve as precursors for the synthesis of well-known platform chemicals in defined biorefineries [6–8]. Recent advances demonstrate that the hemicellulose fraction can also be upgraded by direct aldehyde functionalization of biomass, leading to acetal-stabilized xylose molecules which can be used as polar aprotic solvents [9,10] and precursors for renewable polyesters [11]. For valorization of each fraction, ensuring the separation of cellulose, hemicellulose, and lignin streams is essential, which can be achieved using solvent-based biomass

Abbreviations: RCF, reductive catalytic fractionation; DMSO, dimethyl sulfoxide; ILs, ionic liquids; DES, deep eutectic solvents; EHS, environmental, health, and safety; QSAR, quantitative structure–activity relationship; QM, quantum mechanical; MP, melting point; BP, boiling point; NMMO, n-methyl morpholine n-oxide; H, hydroxyphenol; G, guaiacyl; S, syringyl; LCC, lignin–carbohydrate complex; QSPR, quantitative structure–property relationship; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; PTFE, polytetrafluoroethylene; GC-FID, gas chromatography with flame-ionization detection; HPLC, high-performance liquid chromatography; RID, refractive index detector; 2-MeTHF, 2-methyltetrahydrofuran; MAL, mild acidolysis lignin

* Corresponding author at: École Polytechnique Fédérale de Lausanne, Laboratory of Sustainable and Catalytic Processing, Station 6, Lausanne, 1015, Switzerland.

** Corresponding author at: Max Planck Institute for Dynamics of Complex Technical Systems, Process Systems Engineering, Sandtorstraße 1, Magdeburg, 39106, Germany.

E-mail addresses: jeremy.luterbacher@epfl.ch (J. Luterbacher), sundmacher@mpi-magdeburg.mpg.de (K. Sundmacher).

¹ Authors contributed equally.

<https://doi.org/10.1016/j.cej.2022.139476>

Available online 1 October 2022

1385-8947/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

fractionation approaches, such as organosolv and reductive catalytic fractionation (RCF). In organosolv processing, the biomass is pretreated in organic solvents at temperatures ranging from 80 °C to 250 °C in the presence of acid catalyst to extract lignin and hemicellulose out of the cellulose fibers [12]. Lignin is then separated from the co-extracted hemicellulose by precipitation while cellulose remains as a solid residue [13,14]. RCF is characterized by concurrent lignin solvolysis and depolymerization [12]. Typically, mixtures of low boiling alcohols and water are used for lignin extraction and, then, heterogeneous redox catalysts enable depolymerization by hydrogenolysis in the presence of a hydrogen source [15–19]. RCF and organosolv processing are similar with RCF resulting in a monomer-rich lignin oil instead of a lignin precipitate [18].

There are two major challenges associated with solvent-based approaches. First, high structural differences between the biomass fractions make it difficult to select one solvent being optimal for processing of all components. This often results in the use of multiple solvents in one process, increasing cost and negative environmental impact. Another challenge is the high complexity of the biomass and the recalcitrance of the material which is resistant to mechanical, enzymatic and chemical attack [20,21]. Due to the recalcitrant biomass, lignocellulose-based biorefineries usually operate at harsh conditions, such as high temperature, acidic or alkaline environment [12,22]. Under such conditions, a balance is sought between preserving the structure of valuable biomolecules and degrading the biomass. This balance is highly dependent on the used process and can be adjusted using the right solvent that (i) makes the cleavable bonds of residues accessible for chemicals; (ii) favors chemical transformations [23].

Several conventional solvents have been proven to promote biomass depolymerization at reasonable extent so far [24]. Three main groups of solvents include cyclic ethers (1,4-dioxane, tetrahydrofuran, γ -valerolactone), polar aprotic solvents (dimethylsulfoxide (DMSO), acetonitrile) and alcohols (ethanol, glycerol) usually mixed with water [25–28]. Ionic liquids (ILs) and deep eutectic solvents (DESs) were also recognized as promising media, but their application is still at early stage [29,30]. Besides the ability of solvents to promote depolymerization, other factors like cost, safety and environmental impact begin playing a crucial role, given the potential scale of biorefinery operations [26]. Moreover, each specific organosolv or RCF technology can require the use of strictly defined classes of solvents (e.g. without free -OH groups due to reactivity) with specified physical parameters within a certain range (e.g. with boiling point higher than 80 °C). This makes solvent selection more challenging while solvent screening tools for biomass treatment are not robust and widely available yet. At the same time, experimental solvent screenings are laborious and are limited by financial and human resources.

A large solvent search space can be screened rapidly by leveraging computational approaches. They drive the search towards the most promising solvents *in silico* facilitating the exploration of unconventional solvent candidates and limiting experimental effort. There exist several studies regarding solvent selection for lignocellulose [31–40] but most of them focus on one biomass fraction (e.g., on lignin for “lignin-first” approaches). Computational studies considering the biomass in a more holistic manner are limited and use a small number of solvent candidates [40]. In addition, these studies do not take into account environmental, health and safety (EHS) criteria which are important for safe process operation and critical for industrial implementation. Computational large-scale solvent screenings have proven to be a powerful tool for rational solvent selection including the estimation of EHS properties [41] and were already applied for biorefinery processes [42].

In the present study, we demonstrate a high-throughput computational solvent screening of more than 8000 candidates for effective dissolution of lignocellulosic biomass with the aim to assist its fractionation into lignin, cellulose, and hemicellulose. We apply state-of-the-art quantitative structure–activity relationship (QSAR) models

for the prediction of EHS properties of solvents and the quantum mechanical (QM)-based method COSMO-RS for solubility predictions of the biomass fractions. Subsequently, we validate the applicability of the framework by performing solubility experiments of the targeted biomass fractions in the most promising solvent candidates. In doing so, we explore a vast number of solvents featuring green EHS criteria for their ability to solubilize lignin from different wood sources and facilitate cellulose hydrolysis. Therefore, this study intends to provide guidance for solvent engineering aimed at complex biomass targets.

2. Methods

A computational solvent screening methodology for biorefineries was recently applied for microalgae processing and its complex separation tasks [42]. In this study, the methodology is extended to the processing of lignocellulosic biomass. In brief, a database containing more than 8000 molecules was screened to select appropriate solvent candidates while requiring non-reactivity with the biomass, melting point (MP) and boiling point (BP) limits as additional criteria (see Fig. 1). Simplified structures of lignocellulosic polymers were used as representative biomolecules and modeled using QM calculations. Their applicability in COSMO-RS solubility predictions was thoroughly studied. Predicted solubilities of the chosen representative biomolecules and predicted EHS properties were used to rank the solvents. Based on the screening results, solvents for different biorefinery scenarios – lignin first or joint lignin and cellulose solubility in one step – were proposed. Finally, the computational approach was experimentally validated by measuring the solubility of hardwood, softwood and herbaceous lignins isolated from birch wood, beech wood, and corn cobs (see ESI for HSQC-NMR spectra of the lignins). In addition, cellulose and cellobiose solubilities were measured in several identified solvents.

2.1. Database

As in [42], the solvent search space consisted of the COSMOthermX19 integrated database, COSMObase13-01, and COSMObaseLL19-01. Green solvents as presented in Moity et al. [43], and the green solvent Cyrene [44] were added to the database. Furthermore, 178 DESs [45–55] and 143 commercially available ILs [56,57] were screened. All DESs were taken from literature and were confirmed to have a eutectic point. After deleting duplicate solvents, we obtained a search space of 8011 molecules. A list of all solvents considered in this work can be found in the ESI.

2.2. Structural constraints

After compiling the database, unfavorable solvents were subsequently removed from the search space. In a first step, single anions and cations were eliminated, since ions can only exist with a counter-ion as combined in ILs. In addition, primary and secondary amines were removed due to their potential reactivity with the biomass. Amines are basic, nucleophilic, and can react with carbonyl compounds. Additionally, aromatic amines are highly reactive in electrophilic aromatic substitution.

2.3. MP/BP screening

In lignocellulose processing, high temperatures are usually applied to attack the recalcitrant biomass. In order to identify solvents being liquid during the biomass pretreatment, we screened for solvents with $MP \geq 70$ °C. The BP was required to be in the range 70 °C \leq BP \leq 200 °C (ambient pressure), in order to ensure its facile downstream recovery and recycle. MPs and BPs are given in the database for most of the solvents. However, if there was no data available, an automated PubChem query was used to collect missing experimental data [58].

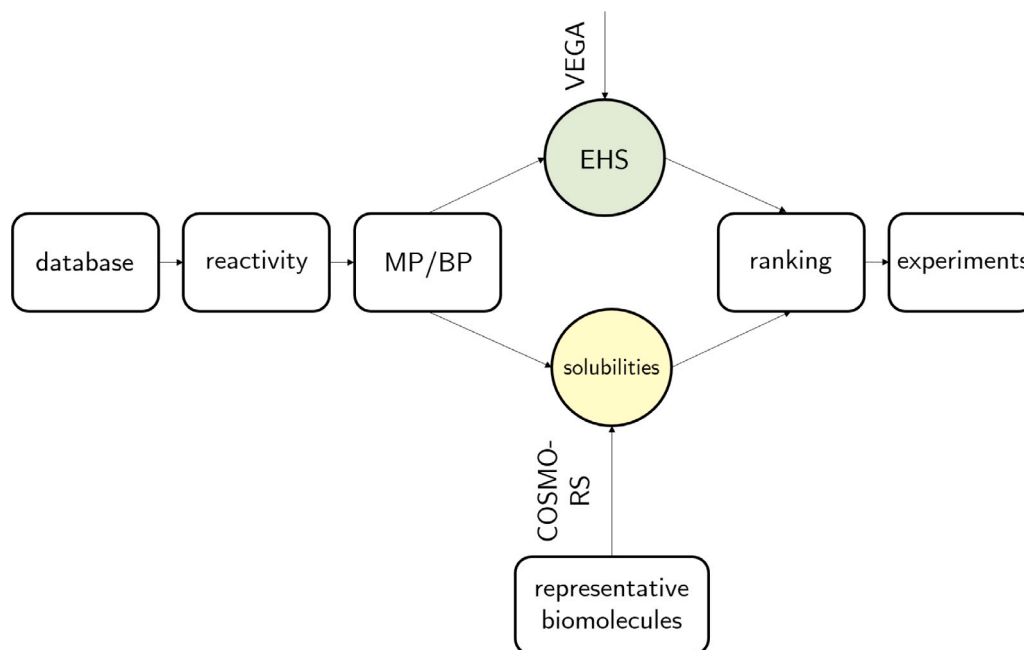


Fig. 1. Approach for the computational solvent screening for lignocellulose processing. A database containing more than 8000 solvents was screened for molecules being unreactive towards the biomass and featuring suitable MP/BP limits. The remaining solvents were ranked according to beneficial EHS properties and solubilities of representative lignocellulose-molecules, as predicted by the VEGA software and COSMO-RS, respectively. Finally, the most promising candidates were evaluated in experiments.

In case no data was stored in the PubChem database, missing BP data was estimated using COSMO-RS at standard conditions. Solvents with missing entries for MP or BP were kept in the screening to prevent false exclusion.

2.4. Modeling of representative lignocellulose molecules

In order to predict the solubilities of the biomass components using COSMO-RS, first, these components must be modeled using the continuum solvation model. Lignocellulose-rich biomass is a complex polymeric mixture, containing mostly cellulose, lignin and hemicellulose as their major components. Depending on the biological source, the proportions of each component may vary, as well as the structure of the macromolecules themselves. Typically, lignocellulosic biopolymers reach high molar weights: > 150 kDa for cellulose, > 30 kDa for hemicelluloses, > 4 kDa for lignin [59,60], hence, modeling their native structures is computationally infeasible. To reduce computational time and complexity, the biomass fractions in this work were modeled individually as fragments of the original polymeric structure, as shown in Fig. 2. We modeled four structures as potential representative molecules for the cellulose fraction, ten structures were assessed for their ability to represent the lignin fraction and two model molecules for the hemicellulose fraction. The structures of all representative molecules can be found in the ESI and were modeled on a quantumchemical level as described in Section 2.8. From these 16 initial structures, we identified the most suitable representative molecules for each fraction by correlating COSMO-RS solubility predictions with experimental solubility data in Section 3.2. In the following, we describe all structures in detail.

2.4.1. Cellulose

Cellulose is the main constituent of wood and exists as rigid chains of β -1-4-glycosidically linked glucose monomers with a degree of polymerization of up to 10,000 units [61]. Due to its chemical composition and strong intra- and inter-molecular hydrogen bonds, cellulose is insoluble or only partially soluble in most common organic solvents [62]. Solvents and ILs that are known to be able to solubilize cellulose at reasonable extent (n,n-dimethylacetamide-LiCl, DMSO-tetrabutylammonium fluoride, n-methyl morpholine oxide (NMMO))

are either toxic, thermally unstable, expensive and/or difficult to recycle [62–64]. In literature, several representative molecules for cellulose were defined. In the simplest case, a single glucose monomer was used by Casas et al. [39]. However, the solubilization properties of glucose differ significantly from those of cellulose. Chu et al. tested glucose, cellobiose, cellotriose and cellotetraose as representative molecules for the prediction of excess enthalpies for cellulose in ILs [35]. In their study, the predicted excess enthalpies of cellobiose and cellotetraose had the highest correlation with experimental solubility data. Other studies used cellobiose for predictions with COSMO-RS [36,37]. Another common approach in the handling of polymers using COSMO-RS is to run QM-calculations for a computationally feasible polymer fragment and to subsequently truncate its end-groups. In this way, the whole polymer fragment is geometry optimized and mid-groups are positioned as if they were contained in the polymer chain. The influence of the end-groups is usually low in the native polymer structure which is why they should be removed for solubility predictions. This approach was used in different studies on a cellotriose [34] and a cellotetraose molecule [38]. Both studies highlighted the importance of intramolecular hydrogen-bonding which is different from one molecular cellulose-representing conformer to another. According to Yamin, the hydrogen-bonding is better captured when not only one mid-monomer as in a truncated cellotriose molecule is considered, but two mid-monomers as in cellotetraose [38]. In another study of Casas et al. [40], they tried not only to capture intracellular interactions, but also hydrogen bonding with adjacent cellulose chains. However, conformers were not considered, probably due to the time-consuming calculations. With this knowledge, we applied the search for conformers to all modeled biomolecules and we assessed cellobiose, cellotriose, cellotetraose, as well as a capped cellotetraose molecule for their ability as cellulose representatives in solubility predictions.

2.4.2. Lignin

Lignin is an amorphous polymer with molar weights typically between 2500 to 15,000 g/mol. Lignin is formed by radical coupling of three main phenylpropanoid units: *p*-coumaryl, coniferyl, and sinapyl monolignols. As a result, lignin consists of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits [1,60,65]. The amount of each

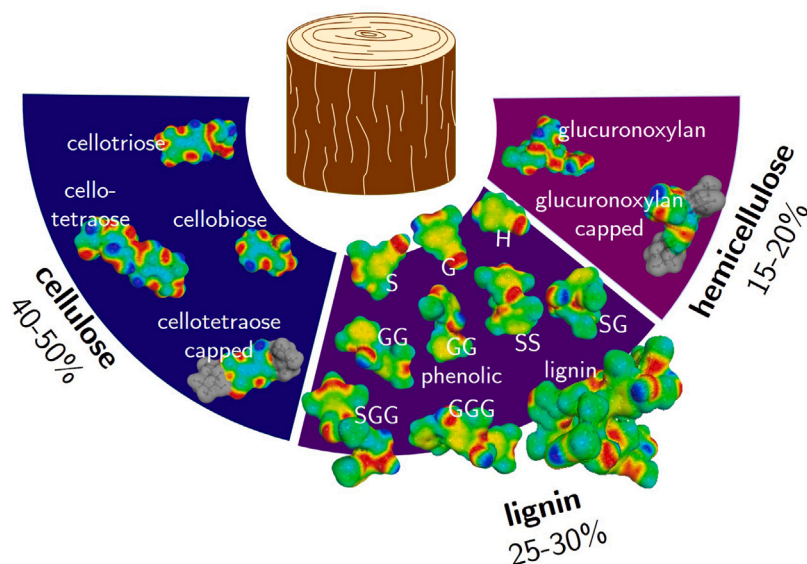


Fig. 2. σ -surfaces of the representative molecules from each biomass fraction. Their chemical structures are presented in the ESI. The amount of each fraction per biomass (dry weight basis) is given in percentage.

subunit differs significantly with biomass source. Lignin from herbaceous biomass is composed of all three subunits (with H-content < 5%). Softwood lignin contains only G units whereas hardwood lignin is composed of both S- and G-units [66]. The bonds that link the lignin subunits together are mostly ether motifs and carbon-carbon bonds. The β -O-4 ether linkage is the most common type of bonds connecting the monolignols and is also most easily cleaved [67,68]. The high amount and the complexity of linkages render lignin a difficult target to solubilize. In previous COSMO-RS studies [31,33,36,37,39] mostly monolignols were used to represent lignin. In addition to monolignols, Casas et al. used pinoresinol and guaiacylglycerol-2-coniferyl ether as representative lignin molecules [40]. Achinivu et al. used S- and G-units connected *via* all major linkage-motifs for solubility predictions [32]. In this study, we modeled S, G and H monomers, as well as dimers and trimers of S- and G-units connected *via* β -O-4 bonds. Additionally, we modeled one conformer of a 1500 g/mol lignin fragment for solubility predictions using COSMO-RS in order to study differences compared to the truncated representatives.

2.4.3. Hemicellulose

Hemicelluloses are a diverse class of polysaccharides found in plant cell walls. The most abundant sugar found in hemicellulose is xylose with minor amounts of mannose, galactose, arabinose, and rhamnose. Uronic acids and acetyl groups are also commonly appended to the chain [69–71]. The degree of polymerization ranges between 50 and 300 [69,71]. Hemicellulose is relatively easy to solubilize in water at high temperature (> 150 °C) due to its hydrophilic and amorphous nature. In the biomass however, hemicellulose and cellulose are bound to lignin, forming a lignin-carbohydrate complex (LCC). Benzyl ether, ester, and phenyl glycosidic bonds are the most typical lignin-carbohydrate linkages. Due to its strong bonding, the LCC hinders enzymatic hydrolysis of biomass and can be partially broken in alkaline media [72]. In recent literature, hemicellulose was represented by a mixture of glucose and xylose monomers [31,36,37]. In this study, we used a glucuronoxylan which is primarily present in dicots to represent the hemicellulose fraction.

2.5. Solubility predictions

After identifying solvents within favorable MP/BP ranges, the solubilities of the representative molecules were predicted. Within COSMOtherm, the solubility of the representative molecules is given as the

logarithmic, molar solubility $\log_{10}(x_{\text{solub}})$. The solubility was calculated as follows:

$$\log_{10}(x_{\text{solub}}) = \frac{[\mu^{(\text{pure})} - \mu^{(\text{solvent})}(x^{\infty}) - \max(0, \Delta G_{\text{fus}})]}{RT \ln(10)} \quad (1)$$

where $\mu^{(\text{pure})}$ denotes the chemical potential of the pure solute, and the chemical potential of the solute at infinite dilution in the solvent is given as $\mu^{(\text{solvent})}(x^{\infty})$. The free enthalpy of fusion ΔG_{fus} is zero for liquid compounds and is estimated for solid compounds based on a quantitative structure-property relationship (QSPR) approach within COSMOtherm [73]. The QSPR approach is valid for standard conditions. Since the solubility was predicted at 85 °C, the temperature dependency of ΔG_{fus} is estimated by Walden's rule, which assumes that $\Delta S_{\text{fus}} = 0.0135 \text{ kcal}/(\text{mol K})$ [73]. If $\log_{10}(x_{\text{solub}}) = 0$, this indicates full dissolution.

In some cases, the solubility for a fraction was calculated as the average solubility of several representative molecules denoted as $\log_{10}(\bar{x}_{\text{solub,avg}})$. In this case, the solubilities of the respective single representative molecules were first predicted according to Eq. (1), and then averaged:

$$\log_{10}(\bar{x}_{\text{solub,avg}}) = \frac{1}{n_{\text{rep}}} \sum_{i=1}^{n_{\text{rep}}} \log_{10}(x_{\text{solub}}^{(i)}) \quad (2)$$

where n_{rep} indicates the number of representative molecules in the lignin fraction. COSMOtherm was chosen as predictive model due to its ability to estimate the thermodynamic properties of complex (bio-)molecules based on quantumchemical information. COSMOtherm delivers fast approximate predictions of the solubility and allows for a qualitative evaluation of large databases.

2.6. EHS properties

The prediction of EHS properties was explained in detail in our previous works [41,42]. In brief, QSAR models implemented in the VEGA software [74] were used to predict solvent EHS properties for mutagenicity, carcinogenicity, toxicity, skin sensitization and ecotoxicity. The QSAR models are training-set dependent and offer fast property prediction for the large database. Song et al. [75], and Linke et al. [41] used VEGA models to predict EHS properties in the context of a solvent screening. Using the model results, a score to express beneficial EHS properties was proposed, the so-called EHS score. The EHS score ranges from 0 to 1, and is also taking the model reliability into account. While

hazardous molecules have a low EHS score, solvents with favorable EHS properties have a high EHS score. For DESs, the EHS score for the hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) were predicted and subsequently averaged according to their molar fractions in the DES. EHS properties of ILs cannot be predicted using VEGA models, since the training set of these models did not include salt-like molecules. As a consequence, their EHS score was set to 1.0, to prevent false exclusion and due to the fact that the safety datasheets of most ILs contain no warnings (which could also be due to lack of experiments). In this study, the EHS score was not used as exclusion criterion, but was used in the solvent ranking (see Section 2.7). After the ranking, EHS properties of potential solvents were manually checked to avoid exclusion of candidates by mistake and to confirm the predicted EHS properties.

2.7. Solvent ranking

The solvent screening was formulated as the following optimization problem:

$$\min_y f(t(y, p), e(y)) \quad (3)$$

$$\text{s.t.} \quad h(y) \leq 0 \quad (\text{structural constraints}) \quad (4)$$

$$g(p) \leq 0 \quad (\text{MP/BP limits}) \quad (5)$$

$$y \in Y \quad (\text{molecular structure}) \quad (6)$$

$$p \in P \quad (\text{process conditions}) \quad (7)$$

The objective function $f(t, e)$ was defined as the distance d from the optimal point $\vec{o}(t, e)$ depending on thermodynamic models t as given in Eqs. (1) and (2), and environmental models e as given by the QSAR models in VEGA described in Section 2.6.

$$f(t, e) = d(\vec{o}(t, e), \vec{x}(t, e)) = \|\vec{o}(t, e) - \vec{x}(t, e)\| \quad (8)$$

where $\vec{x} \in X$, and $X \subset \mathbb{R}^n$, where n denotes the number of dimensions. The vector \vec{x} is defined as $\vec{x} = [x_c, x_l, x_h, \text{EHS score}]^T$. x_c , x_l , and x_h are the molar solubilities acquired by transforming the results obtained from Eq. (1) for cellulose and hemicellulose and Eq. (2) for lignin into molar fractions. As a consequence, values for all dimensions in space lie in the interval [0,1]. The optimal point $\vec{o}(t, e)$ is always known and is defined for different separation tasks commonly encountered in lignocellulosic biorefineries in Section 3.3. Depending on the aim of the process, dimensionality of \mathbb{R}^n varies, since not all dimensions are systematically needed (e.g. hemicellulose solubility is neglected), therefore $n \in \{1, 2, 3, 4\}$ holds. In order to solve the optimization problem, solvent molecules not fulfilling the constraints as given in Eqs. (4) and (5) were removed as explained in Sections 2.2 and 2.3. For the remaining solvents, thermodynamic properties t (solubilities) and EHS property models e as given in Sections 2.5 and 2.6 were used to predict values necessary to evaluate the objective function, thus, to calculate the euclidian distance from the optimal point $d(\vec{o}(t, e), \vec{x}(t, e))$ as defined in Eq. (8). To minimize the objective function $f(t, e)$, solvents were ranked according to minimal euclidian distance to $\vec{o}(t, e)$. After the ranking, the MPs, BPs and EHS criteria of the solvents were compared to literature data to reassure that the constraints imposed on the optimization problem were indeed fulfilled.

2.8. Computational details

Quantum chemical density functional theory calculations were performed for the molecules not contained in the COSMObase13-01 and COSMObaseLL-19-01. Molecular conformers were generated using RD-Kit [76] with a force field according to Eberjer et al. [77]. Subsequently, the resulting structures were further optimized 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 optimized geometries, a single point calculation was performed, using the more accurate def2-TZVPD bases set. Cavities were constructed at the FINE level. A python script (Python 3.7) was used to implement the automated screening procedure. In this procedure, the database described in Section 2.1, was loaded as a pandas dataframe (pandas 0.25.1). Subsequently, depending on structural constraints, as well as the MP and BP limits, entries of unsuitable molecules were deleted. PubChemPy 1.0.4 was used to read MPs and BPs from PubChem. Missing BP and flash points (used for prediction of the EHS criteria) were predicted using CosmoPy 19.10. EHS properties were predicted using VEGA QSAR 1.1.5. For solubility predictions, the python script called COSMOtherm v19 [78–82] via command line. For these calculations, input files were automatically generated and the BP_TZVPD_FINE_19.ctd parameterization was applied. The results of each solubility predictions was stored in SQLite3 databases using sqllite 3.30.0. All COSMO-RS predictions including ILs and DESs were handled using the so-called electroneutral approach [83–85]. In this approach, the single IL and DES constituents were implemented as a stoichiometric mixture. The electroneutral approach is commonly followed 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. Using this setup, the approximate computation time for conformer generation and TURBOMOLE calculations of monomeric lignin structures was 30 min (three meaningful conformers identified) and for the trimeric lignin structures five days (ten conformers identified).

2.9. Experimental validation

The predicted solvent candidates were further narrowed to select chemicals for experimental validation. The final choice of chemicals was based on (i) commercial availability (ii) reasonable price (≤ 100 \$ per 10 g). As a result, 22 solvent candidates were selected for experimental solubility measurements.

2.9.1. Materials

The solvents to measure the solubilities were purchased as following: diethyl methylphosphonate (ACROS Organics, 96%), diethyl sulfide (Fluorochem, 98%), diethyl ethylphosphonate (Sigma-Aldrich, > 98%), dimethyl sulfoxide (Sigma-Aldrich, > 99.5%), 1-methylimidazole (Sigma-Aldrich, > 99%), dimethyl methylphosphonate (STREM Chemicals, 97%), 18-crown-6 ether (Sigma-Aldrich, for synthesis), 2-methyl-2-oxazoline (ABCR, 99%), 4-piperidinopyridine (ABCR, 97%), 4-pyrrolidinopyridine (ABCR, 95%), 4-methoxy-pyridine (ABCR, 97%), 5-Bromo-1-methyl-1H-imidazole (ABCR, 95%), n,n-dimethylmethanesulfonamide (ThermoScientific, > 98%), 4-methylpyridine (ABCR, 98%), pyrazine (ACROS Organics, 99%), n,n-dimethylformamide (Sigma-Aldrich, > 99.8%), pyrazole (ThermoScientific, > 98%), 1,4-dioxane (Carl Roth, > 99.5%), 2-methyltetrahydrofuran (Carl Roth, > 99%), dibutyl ether (ACROS Organics, > 99%), n-heptane (ABCR, > 96%), trimethylamine n-oxide (Cayman). Microcrystalline cellulose (fibers, medium) Sigma-Aldrich, D-glucose (> 99.5%) and D-cellobiose (> 98%) from Sigma were used as carbohydrate materials. Sulfuric acid (95%–97%) purchased from Supelco.

Four lignin samples were investigated in this study. Two of them (FABIOLA™ and Kraft) were obtained as lignin precipitates from our collaboration partners and two samples (from birch wood and corn cobs) were obtained by acidolysis of biomass [86]. Lignin from Rettenmaier beechwood was obtained from TNO (Netherlands) and was originally gained from the FABIOLA™ acetone organosolv process [87]. Kraft lignin was provided by Berner Fachhochschule and was originally isolated from softwood species by Kraft process. Birch wood (*Betula pendula*) was procured from M. Studer of the Bern University of Applied Sciences. The wood chips were sorted to remove residual bark and leaves. The wood chips were then milled using a 6-mm screen and

sieved with a 0.45-mm mesh. The fraction < 0.45 mm was used for lignin extraction by mild acidolysis. The corn cobs were procured from IP-Suisse in Lausanne, Switzerland. They were sorted to remove residual leaves, stems, and corn. The corn cobs were then milled using a 6 mm screen and sieved with a 0.45 mm mesh. The fraction < 0.45 mm was used for lignin extraction by mild acidolysis.

2.9.2. Lignin extraction by mild acidolysis

To validate the applicability of the described computational method, we first isolated lignin from biomass under mild conditions to measure its solubility in the selected solvents. The mild acidolysis lignin extraction method allows to largely retain native-like structure of lignin even if extraction yields are small [88]. Two different lignocellulosic biomass types were used for lignin isolation: a hardwood represented by Birch wood, and herbaceous biomass represented by corn cobs. The procedure for lignin isolation was taken from Das et al. [88] with minor modifications. In brief, grinded biomass (10 g) was dissolved in 120 ml of a dioxane/water mixture (9/1, v/v) containing 0.2 mol/l 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 a dioxane/water mixture (9/1, v/v). The pH of the resulting solution was adjusted to 3–4 using a saturated aqueous NaOH solution. Then, the solution was concentrated to about 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 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 ESI).

2.9.3. Lignin solubility measurements by gas chromatography

We performed a solubility measurement using gas chromatography as opposed to gravimetric methods including solvent evaporation due the difficulty of evaporating high boiling point solvents. In this method, dried lignin (0.1 g) was added to the solvent (0.4 g) in a vial with a magnetic stirring bar. The vials were sealed and placed in an aluminum block holder heated to 85 °C under constant stirring at 400 rpm. The samples were kept under agitation until reaching equilibrium (2 h) and then filtered using 1 ml syringe with an attached polytetrafluoroethylene (PTFE) filter (0.22 μm 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-dioxalane was added to the sample (ca. 0.1 g) to serve as an internal standard. The samples were quantified by gas chromatography with flame-ionization detection (GC-FID). Calibration curves for each solvent with the internal standard were obtained using known amounts of solvent and 1,3-dioxalane dissolved in DMSO or acetone (if the peaks of DMSO and solvent were overlapping). All solubility tests were performed in duplicate. The lignin solubility was calculated using the following equations:

$$\text{lignin solubility wt.\%} = \frac{m_1}{m_1 + m_{\text{solv}}} \cdot 100 \quad (9)$$

By measuring the mass of solvent m_{solv} and the mass of the filtrate m_{filtrate} , the mass of lignin m_1 was obtained using the following equations:

$$m_{\text{solv}} = \frac{F \cdot \text{Area}_{\text{solv}} \cdot m_{\text{IS}}}{\text{Area}_{\text{IS}}} \quad (10)$$

$$m_1 = m_{\text{filtrate}} - m_{\text{solv}} \quad (11)$$

where F 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. Response factors and the GC conditions are given in the ESI.

2.9.4. Lignin solubility measurement by evaporation

To validate the results of the GC-method above, we performed several solubility measurements using the traditional gravimetric evaporation method. We used the procedure from Dick et al. [89], where the filtered lignin solution is dried to remove the solvent and the mass of dissolved lignin is quantified. In brief, dried lignin (0.1 g) was added to the solvent (0.4 g) in a vial with a magnetic stirring bar. The vials were sealed and placed in an aluminum block holder heated to 85 °C under constant stirring at 400 rpm. The samples were kept under agitation until reaching equilibrium (2 h) and then filtered to a tared vial using a 1 ml syringe with attached PTFE filter (0.22 μm pore size) to remove the undissolved solid lignin. The vials containing the filtered solution were placed in the vacuum oven at 45 °C and 20 mbar, and dried overnight. The vials were then 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:

$$\text{lignin solubility [wt.\%]} = \frac{m_{\text{vial with dry lignin}} - m_{\text{vial}}}{m_{\text{vial with lignin solution}} - m_{\text{vial}} - m_{\text{dry lignin}}} \quad (12)$$

This procedure has been applied to six solvents with BP < 150 °C allowing them to be evaporated at 45 °C and 20 mbar.

2.9.5. Cellobiose solubility measurement

Cellobiose, a disaccharide unit of cellulose, was used as a proxy model compound for cellulose owing to the limitation in accurately quantifying solubility of commercially available microcrystalline cellulose. Cellobiose (0.02–0.10 g) was added individually to 0.20–10 g of solvent. For solvents that are liquid at room temperature, the mixture of cellobiose and solvent was stirred vigorously in a vortex for 2 h at 25 °C followed by stirring at 85 °C for 30 min. The mixture was then filtered using a H-PTFE filter (0.22 μm pore size) to collect a homogeneous solution of cellobiose in a glass vial. Solvents solid at room temperature were mixed with cellobiose and heated to 85 or 100 °C (depending on the MP of the solvent) for 30 min and subsequently filtered quickly. The filtrate from the cellobiose solution was then added to 400 mM H₂SO₄ solution and shaken to extract the dissolved cellobiose into the aqueous acid phase. The solubility of cellobiose in each solvent was measured as the concentration of cellobiose extracted in the above 400 mM H₂SO₄ solution and detected by the high-performance liquid chromatography (HPLC). The solubility of cellobiose was determined using Eq. (13) as given below:

$$\text{cellobiose solubility [wt.\%]} = \frac{m_{\text{cellobiose}}}{m_{\text{cellobiose}} + m_{\text{solv}}} \cdot 100 \quad (13)$$

The mass of cellobiose recovered in aqueous acid solution is denoted as $m_{\text{cellobiose}}$ and the mass of solvent in the test solution is given as m_{solv} . As the maximum solubility limits of cellobiose were unknown and could vary widely among solvents [90], the dissolution of cellobiose was studied below the saturation point for individual solvents.

2.9.6. Cellulose hydrolyzability as indicator of swelling/disruption

Due to limitation of direct measurement of solubility of cellulose in the organic solvents, an indirect semi-quantitative method was developed to study the hydrolyzability of cellulose. The swelling/disruption effect of each solvent was studied based on varying degrees of the main product of cellulose hydrolysis (glucose) [91]. Ideally, based on the amount of glucose produced, the amount of cellulose or its oligosaccharide derivatives dissolved in each solvent can be estimated using the stoichiometry of hydrolysis [28,92]. However, as observed in preliminary tests, the filtrate (0.22 μm pore size) of a mixture of microcrystalline cellulose and most of the tested solvents did not contain any dissolved cellulose. Additionally, most common biomass processing technologies based on solvents use elevated temperatures and acidic conditions. Therefore, the mixture of cellulose and each solvent was directly hydrolyzed to glucose without any filtration (see details below) and then analyzed by HPLC. Literature shows that the rates of cellulose hydrolysis is unaffected by the presence of a polar

organic solvent when the concentration of the solvent is below 20 vol% in total solution [93]. Therefore, under such hydrolysis conditions, any changes in hydrolyzability of cellulose after swelling in a solvent could be attributed solely to the swelling/disruption effects of the solvent on cellulose during the pre-hydrolysis swelling. Cellulose (0.02–0.1 g) was mixed individually with solvents (0.3–11.6 g) for 1.5 h and then heated to 85 or 100 °C for 30 min (swelling period) and subsequently transferred into 5 ml of 2 M H₂SO₄ solution for hydrolysis of the swollen, heated cellulose mixture. For solid solvents, only heating at 85 or 100 °C for 30 min under stirring was conducted before hydrolysis. There was no liquid–liquid phase separation between the acidic solution and the solvents. Acid hydrolysis was conducted in Ace glass reactors (6 ml capacity) where the concentration of acid was optimized to ensure that at least 1 mg cellulose in 1 ml of solvent could be fully converted to glucose via hydrolysis reaction. The reactor, tightly sealed, was heated to 125 °C on a hot plate under stirring. The hydrolysis reaction was conducted for 44 min followed by cooling to room temperature. The liquid product of acid hydrolysis was filtered and analyzed by HPLC. An anhydro-correction factor of 0.9 was used to account for the additional mass of water in glucose occurring in hydrolysis reaction:

$$\text{cellulose hydrolyzability [wt.\%]} = \frac{0.9 \cdot m_{\text{gluc,hydrolysis}}}{m_{\text{cellulose,ini}}} \cdot 100 \quad (14)$$

The mass of glucose produced during hydrolysis is expressed as $m_{\text{gluc,hydrolysis}}$ and the initial mass of cellulose is given as $m_{\text{cellulose,ini}}$. An Agilent Infinity 1260 HPLC system equipped with an Aminex HPX-87H Column (300 mm × 7.8 mm; column temperature 60 °C) and a Refractive Index Detector (RID) (G1362 A) was used to analyze the carbohydrates in this study. Water with 5 mM H₂SO₄ at a flow rate of 0.6 ml/min was used as the mobile phase with injection volume of 20 μl. In HPLC, 400 mM H₂SO₄ solution was used as matrix for calibration of cellobiose and glucose. The calibration was performed using D-glucose and cellobiose standards (concentration range of 0.1–10.0 mg/ml) to identify and quantify the sugars in the above cellulose hydrolysis or cellobiose solutions. Each liquid sample carbohydrate solution was analyzed by HPLC twice and we reported the average result for each solvent tested.

3. Results and discussion

Accurate selection of suitable representative biomolecules is crucial for reliable solubility predictions. Therefore, we analyzed the σ -profiles of each molecule and correlated experimental solubility data with COSMO-RS predictions in Sections 3.1 and 3.2. The selected representative molecules were subsequently used for solubility predictions in the computational high-throughput screening. The identified solvents were ranked with respect to different process objectives (e.g. complete dissolution of the biomass or “lignin-first” approach) in Sections 3.3.1–3.3.3. The approach was validated in solubility experiments for the lignin and cellulose fraction presented in Sections 3.4 and 3.5.

3.1. Initial assessment of representative molecules by σ -profile analysis

The σ -profiles for different representative biomolecules give insight about their polarity, acidity and basicity. This is important for the solvent screening as the properties of solvent and solute should be compatible. We plotted σ -profiles for all representative molecules (see Fig. 3) and analyzed them to obtain a first overview of the physico-chemical solute properties. Briefly, a σ -profile represents the probability $p(\sigma)$ of a molecular surface segment to have a specific screening charge density σ . The screening charge density determines the interaction energy with other surface segments in the liquid phase. In COSMO-RS theory, σ -profiles are the basis for predicting the chemical potential, and hence, basis for all following thermodynamic predictions. σ -profiles can be divided into a HBD region ranging from -0.01 to -0.03 e/Å, a rather

neutral region from -0.01 to 0.01 e/Å, and subsequently until 0.03 e/Å a HBA region. The σ -profiles of all cellulose representatives (see Fig. 3(a)) show strong HBA- and HBD-behavior, as visible by the peaks at 0.017 and -0.018 e/Å, respectively. This is not only indicative of interactions with potential solvents via hydrogen bonding, but also for intramolecular hydrogen bond formation. Additionally, the cellulose fraction shows a peak in the neutral region at -0.007 e/Å arising from the sugar carbons. For the σ -profile analysis of cellulose, we also included glucose, as its water-soluble monomer. With increasing chain length, peak height increases, especially in the neutral region which indicates decreasing water-solubility for the cellulose oligomers. The 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. For glucose, the peak in the neutral region is low compared to its peaks in the HBA/HBD region which correspond to its high water solubility in contrast to the other representatives. Overall, a solvent capable of cellulose dissolution should form strong intermolecular hydrogen bonds and additionally have neutral screening charges. The σ -profiles of the hemicellulose representatives (see Fig. 3(c)) have high similarities to that of cellulose leading to strong abilities for hydrogen bond formation. All lignin representatives (see Fig. 3(b)) show two distinct peaks in the neutral region at -0.005 and -0.0025 e/Å caused by the slightly electropositive hydrogen atoms of the aromatic rings and its carbon atoms, respectively. The π -face of the aromatic ring leads to the peak at 0.008 e/Å, which is consequently the highest for the trimeric lignin representatives. The oxygen atoms of the β -O-4 bonds and the free hydroxy groups lead to HBA-behavior. Additionally, a small peak in the HBD region is visible, originating from the hydrogen atoms of the free hydroxy groups. Therefore, inter- and intramolecular hydrogen bonding is possible, but with less intensity compared to the cellulose and hemicellulose representatives. Compared to the σ -profiles of the cellulose representatives, the peaks in the neutral region of lignin are broader. This peak behavior indicates different solubilization mechanisms and the possibility for solvent-based separation of lignin from cellulose and hemicellulose.

3.2. Verification of representative molecules

The structure of lignin depends on its biological source and the applied extraction method leading to different solubility behavior [94]. Moreover, lignin can be tailored towards desired solubilities [89]. Therefore, lignin structures for modeling should be selected carefully to better represent the target lignin. Sameni et al. reported solubilities for 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) [94]. In order to assess how COSMO-RS predictions correlate with experimental data, we compared experimental data for lignins with three compilations of COSMO-RS prediction results by linear regression analysis: averaged COSMO-RS solubilities for monolignols, dimers, trimers; an average of all monolignols, dimers and trimers altogether; and a single conformer of the 1500 g/mol lignin fragment. We removed experimental data points for DMSO and pyridine, as the lignin saturation was not reached in the experiments. For the remaining 9 data points, we converted COSMO and experimental solubility to wt.% as given in Eq. (9) and performed linear regression. The correlation coefficients R^2 are shown in Table 1.

We obtained the most robust and highest correlations using the averaged solubilities of monolignols, dimers and trimers denoted as average_{m,d,t} (Table 1). Correlation coefficients varied strongly among the wood species and the representative molecules. We therefore highly recommend a detailed analysis of representative molecules tailored towards the used type of wood. Surprisingly, we obtained higher correlations using lignin monomers compared to the 1500 g/mol lignin fragment for Kraft and Soda Protobind lignins. For alkaline lignin, the

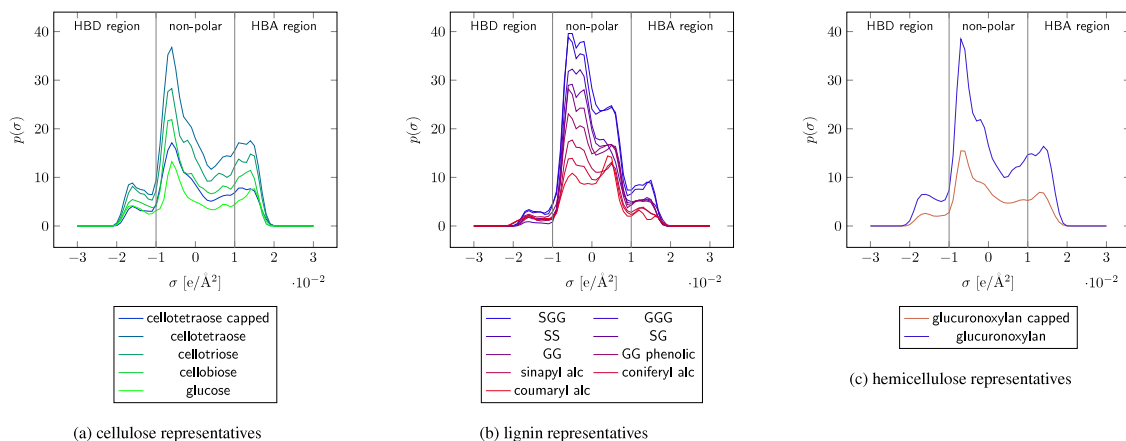


Fig. 3. σ -profiles of the representative molecules from each biomass fraction.

Table 1

Correlation coefficients R^2 from linear regressions between solubilities for lignin from different sources and COSMO-RS predictions for lignin monomers, dimers, trimers, the 1500 g/mol lignin fragment and the average of all monomers, dimers, and trimers. Experimental data was taken from [94].

Lignin type	Monolignols	Dimers	Trimers	Lignin fragment	Average $_{m,d,t}$
Lignin hardwood/non-wood	0.2513	0.0694	0.6738	0.6801	0.7074
Kraft lignin (Eucalyptus)	0.6926	0.0931	0.3845	0.2398	0.6901
Indulin AT	0.0574	0.5589	0.7705	0.5462	0.4747
Protobind	0.6552	0.7791	0.497	0.3260	0.7749

1500 g/mol lignin fragment was found to be the best representative of the real structure, while for other lignins trimers led to improved correlation. This result was probably due to the fact that alkaline lignin has the highest H-content while our representative molecules contain mostly S and G units. Another possible cause is that alkaline lignin has a more than 2-fold higher molecular weight (13,488 g/mol) compared to others [94]. Nevertheless, despite the detailed modeling, the 1500 g/mol lignin fragment was not unreservedly the best representative molecule. The poor performance of the fragment might be caused by the fact that lignin is an amorphous polymer which changes its conformation based on the solvent environment. Especially in unfavorable solvents, lignin forms globules in order to minimize solvent exposure [95]. As a result, only certain parts of the surface area of such a fragment, or its σ -surface would be accessible to the solvent. In our study, we only modeled one conformer for the large lignin fragment and a broader range of possible conformations was not captured.

In contrast to the flexible lignin polymer, cellulose is rigid and recalcitrant. Only few solvents with the ability to dissolve cellulose are known, e.g. several ILs [96,97] and NMMO which is used in the lyocell process [98]. We used experimental cellulose solubilities from Vitz et al. [96] and Zhao et al. [97], where cellulose was dissolved in several ILs at temperatures of 100 °C and 110 °C, respectively. We used the representatives molecules for cellulose as given in the ESI for COSMO-RS solubility predictions. We converted the given data points and COSMO-RS predictions to wt.% as given in Eq. (13) and performed linear regression. We excluded all ILs containing Cl^- ions from the analysis, as there are known to be systematic deviations in COSMO-RS predictions [34,35]. As a result, we obtained a correlation coefficient of $R^2 = 0.7053$ for the capped cellotetraose molecule. Correlation coefficients for cellotetraose, cellotriose and cellulose were 0.3691, 0.3752, and 0.2476, respectively.

Based on these results, we decided to use the capped cellotetraose molecule as a representative molecule for the cellulose fraction as it gave the highest correlation coefficient. For lignin, we chose representation by the average values of monolignols, dimers, and trimers average $_{m,d,t}$. By analogy with cellulose, we used the capped glucuronoxylan molecule to represent the hemicellulose fraction. Overall,

the comparison of experimental solubility data with COSMO-RS predictions for lignin and cellulose highlighted the importance of identifying meaningful representative molecules and confirmed the suitability of COSMO-RS for qualitative solvent comparison for lignocellulose processing.

3.3. Screening results

Of the initial 8011 molecules, 84% met the structural constraints (see Section 2.2). After screening for favorable MPs and BPs, 3525 potential solvents remained for solubility predictions as shown in Fig. 4. The COSMO-RS predicted solubilities of all biomass fractions in the identified solvents and their EHS score is visualized in Fig. 5(a). Solvents with high solubilities of all three biomass fractions with high EHS scores are located in the proximity of $[\log_{10}(x_c), \log_{10}(x_l), \log_{10}(x_h)]^T = [0, 0, 0]^T$. As the number of identified solvents was small, we then relaxed the requirements for high hemicellulose solubilities and high EHS scores. Therefore, we considered solvents in the proximity of $[\log_{10}(x_c), \log_{10}(x_l)]^T = [0, 0]^T$ (Fig. 5(b)). Interestingly, there was no solvent with the ability to exclusively extract cellulose within the search space, as all solvents that are predicted to have a high cellulose solubility co-extract lignin. However, we identified several solvents, which were predicted to be selective towards lignin. These solvents are especially interesting for use in “lignin-first” biomass fractionation approaches. In the following, we formulated three main objectives and ranked solvents in each objective: (i) solvents with high EHS score for the joint dissolution of all lignocellulose fractions, (ii) solvents for the joint dissolution of lignin and cellulose, (iii) solvents for selective lignin extraction. Complete lists of solvents for each objective were added to the ESI.

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

Up to now, few solvents are known to dissolve all biomass fractions of lignocellulose. High EHS score of the solvent candidate would be an essential feature for its implementation in industry. The discovery of such solvents would open new perspectives in lignocellulose processing. Therefore, we chose $\vec{d}(t, e) = [x_c, x_l, x_h, \text{EHS score}]^T = [1, 1, 1, 1]$ as

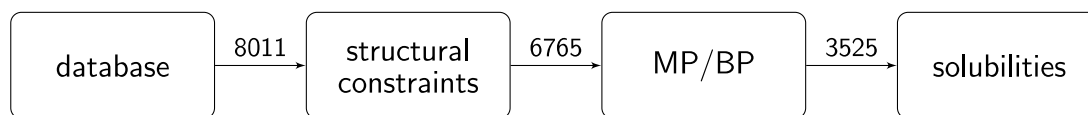


Fig. 4. Screening steps of the computational solvent screening. Numbers above the arrows indicate the amount of eligible solvent candidates after each step.

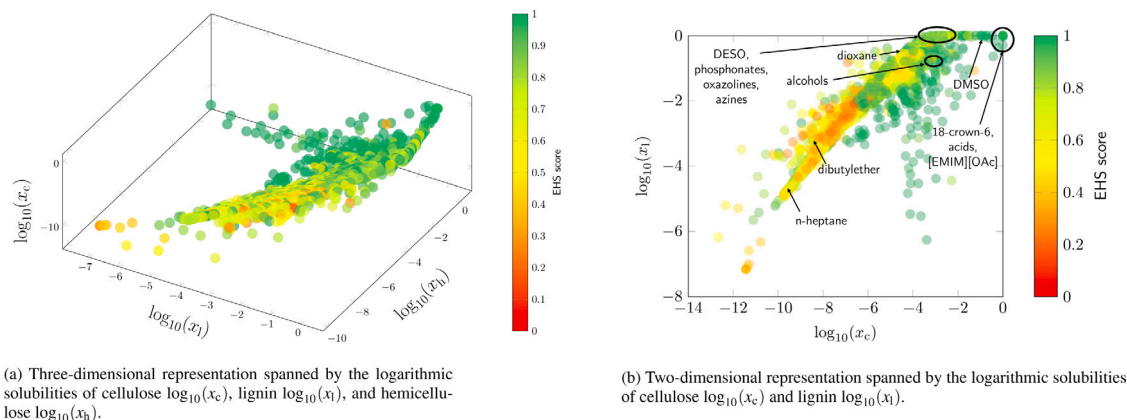


Fig. 5. COSMO-RS predicted solubilities of the biomass fractions and their corresponding EHS score. Note that solubilities are given in log-units, therefore, zero denotes the highest solubility.

Table 2

ILs for the joint dissolution of cellulose, lignin and hemicellulose. The distance from the optimal point $\vec{a}(t, e) = [x_c, x_l, x_h, \text{EHS score}]^T = [1, 1, 1, 1]$ is given as $d(\vec{a}, \vec{x})$. Solubilities for each biomass fraction are given in logarithmic scale. MPs were obtained from Iolitec [56], no BPs were stated. RT indicates room temperature. EHS properties were taken from the PubChem database [58].^a

IL	$d(\vec{a}, \vec{x})$	$\log_{10}(x_c)$	$\log_{10}(x_l)$	$\log_{10}(x_h)$	MP [°C]	EHS
[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

^a [Chol], choline; [OH], hydroxide; [P666,14], trihexyltetradecylphosphonium; [BTMP], bis(2,4,4-trimethylpentyl)phosphinate; [BMIM], 1-butyl-3-methylimidazolium; [OAc], acetate; [EMIM], 1-ethyl-3-methylimidazolium; [MMIM], 1,3-dimethylimidazolium; [DMP], dimethyl phosphate; [DEP], diethyl phosphate; [DBP], dibutyl phosphate; [Dec], decanotate.

optimal point (see Fig. 5(a)). After raking the solvents, we identified several ILs that lie at or near the optimal point (Table 2). The two best performing ILs were [Chol][OH] and [P666,14][BTMP]. [Chol][OH] is an alkaline IL which was already proposed for the valorization of agricultural waste [99]. In combination with urea, it forms a DES which can dissolve 9.5 wt% cellulose [100]. To the best of our knowledge, for [P666,14][BTMP] there exists no experimental data on lignocellulose solubility in literature. However, phosphonium-based ILs are known to dissolve lignocellulosic biomass well [101,102]. In agreement with experimental data, our approach identified [EMIM][OAc] which is able to completely dissolve wood chips [103]. Zavrel et al. dissolved different wood species in [EMIM][OAc] and [MMIM][DMP] [104]. According to Vitz et al. [EMIM][OAc] and [BMIM][OAc] dissolve 8 and 14 wt% cellulose, respectively.

The DESs included in the screening are predicted to lead to a limited cellulose solubility, which is reflected by their distance from the defined optimal point. Nevertheless, choline chloride in combination with imidazole was ranked the highest among the DESs and was reported to dissolve 2.48 wt% of cellulose [105]. An organic solvent that was located at the defined optimal point was 18-crown-6 ether (see Table 3


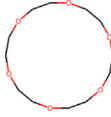

for chemical structure). According to its safety datasheet, there are no EHS concerns and our experiments showed that 18-crown-6 ether is indeed a powerful solvent for lignin and cellulose (see Sections 3.4 and 3.5). Interestingly, the distance from the optimal point decreased rapidly after the first four ranks in general, mainly caused by low cellulose solubilities. Overall, of the 8011 potential solvent candidates only 8 ILs and the 18-crown-6 ether were close to the defined optimal point. Hence, we decided to focus the search on high lignin and cellulose solubilities only and neglect the hemicellulose fraction that can normally be separated in the aqueous stream. The EHS score is a useful metric to quantify beneficial EHS properties, however, it also restricts the search space. Therefore, we also removed the EHS score for the following objectives to obtain more potential solvent candidates in order to gain a better understanding of chemical structures of suitable solvents.

3.3.2. Objective 2: Solvents for the joint dissolution of lignin and cellulose

In order to promote the discovery of solvents, we relaxed the search criteria and defined the optimal point as $\vec{a}(t, e) = [x_c, x_l]^T = [1, 1]^T$ (without hemicellulose solubility and EHS score). In agreement

Table 3

Identified solvents for joint lignin and cellulose dissolution. EHS criteria and hemicellulose solubilities were neglected in the ranking. The distance from the optimal point $\vec{d}(t, e) = [x_c, x_l]^T = [1, 1]^T$ was zero for all solvents presented in the table. Unless otherwise stated, all properties were obtained from the respective safety data sheets.

Solvent	MP [°C]	BP [°C]	EHS	Structure
1-methylpiperidine-1-oxide	n.a.	n.a.	n.a.	
18-crown-6 ether	42–45	118 (0.07 mbar) [107]	irritant	
Dimethylselenoxide	n.a.	n.a.	acutely toxic, health hazard, environmental hazard	

with Section 3.3.1, we identified the ILs given in Table 2 and 18-crown-6 ether. Table 3 lists the best solvents for joint dissolution of lignin and cellulose located at the defined optimal point. We did neither find experimental data for lignocellulose processing using 1-methylpiperidine-1-oxide, nor was it commercially available. However, 1-methylpiperidine-1-oxide has close structural similarities to the solvent NMMO which is known to have excellent properties for cellulose dissolution but suffers from low thermal stability and relatively high price. Another effective solvent for joint lignin/cellulose dissolution was triethylamine-n-oxide with slightly lower cellulose solubility compared to the solvents presented in Table 3. The n-oxide-based structures could serve as a basis for directional synthesis of these or similar solvents for biomass, as the extraordinary polarity of N-O bond could significantly facilitate dissolution due to its strong hydrogen bonding ability. Dimethyl selen oxide is highly toxic. It is the selenium analog of DMSO but has even higher HBA properties [106], which can potentially increase lignocellulose solubility. Similar to objective 1, no DES could compete with the solvents from Tables 2 and 3.

3.3.3. Objective 3: Solvents for selective lignin extraction

Solubility of lignin alone has been a matter of extensive studies over the years [32,108,109]. High lignin solubility is essential for producing clean and homogeneous lignin streams that can be further valorized into valuable aromatic materials. Efforts have been put into finding suitable solvents that also meet green chemistry principles. Therefore, for this objection, we defined the optimal point as $\vec{d}(t, e) = [x_c, x_l]^T = [0, 1]^T$ (low cellulose, high lignin solubility). Solvents near the optimal point selectively target lignin, while cellulose remains as an easily separable solid, which is the case for common fractionation approaches. The EHS score was neglected at this stage but addressed later manually which allowed us to discover a higher number of molecules. A large number of identified solvents allows to gain insights into the underlying structural patterns causing high lignin solubilities. We identified in total 104 organic solvents, one IL, and 4 tetrabutyl ammonium-based DESs with a distance of ≤ 0.5 from the optimal point. Commercially available solvents identified for this objective are presented in Table 4. From the list, we identified many structural similarities among the most promising solvent candidates. The identified solvents belong to four major classes: azines (pyridines, pyrazines, pyrimidines, pyridazines, triazines), sulfoxides, oxazolines, and phosphonates. Alcohols, such as ethanol, 2-propanol or 1-butanol, commonly applied in RCF and in industrial organosolv showed comparably low lignin solubilities. 1,4-dioxane, another commonly used solvent in lignocellulose processing, was predicted to have a high lignin solubility. However, 1,4-dioxane is a known carcinogen linked to organ toxicity and known as environmental contaminant in contrast to many other identified solvents that have more benign EHS properties. These predictions underline how distinctive our results are from the state-of-the-art and give directions for original process design featuring newly discovered solvents. These solvents might have not only better solubility profiles but also more benign EHS characteristics since the EHS score for most of the identified chemicals from Table 4 is > 0.8 .





















3.4. Experimental lignin solubilities in the predicted solvents

To test commercially available substances that were identified specifically for lignin extraction we selected 21 chemicals from objective 3 and measured solubility of four types of lignin in them (Fig. 6 and ESI). Different lignin targets were chosen due to (i) high abundance (Kraft lignin accounts for about 85% of the total lignin production in the world), (ii) high potential for commercialization, (iii) high similarity to native lignin structure (mild acidolysis and FABIOLA™ lignins), (iv) diversity of biomass species (birch wood, beech wood, corn cobs) [110]. For better comparison, in addition to the chemicals with lowest distance to the optimal point $d(\vec{d}, \vec{x})$, we also tested chemicals from the middle of the ranking (pyrazole, 1,4-dioxane) and very end (dibutyl ether, n-heptane).

Lignin solubility is usually determined by two methods: by measuring the mass of non-dissolved lignin after filtration of the initial lignin solution or by measuring the mass of solubilized lignin left after evaporation of the filtered lignin solution. However, both methods are not suitable for solvents that are solids at room temperature and/or high-boiling due to their low volatility (e.g. 18-crown-6 ether, pyrazine, 4-piperidinopyridine). UV-Vis methods that measure absorbance of filtered lignin solution diluted with other solvents (e.g. DMSO) lack sufficient sensitivity. In this study, we used the previously described original method for measuring lignin solubilities using GC-FID (Section 2.9.3). Only 6 solvents out of the selected 21 had boiling point < 150 °C (2-methyl-2-oxazoline, 4-methylpyrimidine, 1,4-dioxane, 2-methyltetrahydrofuran, dibutyl ether, n-heptane), allowing them to be evaporated relatively easily. For them, we repeated the measurements using the solvent evaporation method (Section 2.9.4) to validate the results of the GC-based method. As shown in Fig. 6, the solubility of all 4 types of lignin was very high in the predicted solvents (≥ 20 wt%), while the low ranking candidates (dibutyl ether, n-heptane) showed almost 0 wt% lignin solubility. When preparing the initial lignin solutions at concentrations higher than 20–25 wt%, their filtration was hindered due to the high viscosity of the solvent/lignin mixture. Therefore, we could not reach the solubility limit for most of the selected solvents. Nevertheless, for a few of them, we were able to prepare and filter solutions with solubilized lignin at > 33 wt% (DMSO, crown ether, diethyl sulfoxide, diethyl methylphosphonate), although the solubility limit was still not reached. An automated filtering system with several membranes of different sizes designed for high-viscosity samples could make it possible to determine the exact solubility values of lignin in the selected solvents. However, the obtained data is sufficient to prove high lignin solubilities in the high ranking solvents and reliability of the developed computational method. In general, the solubility of lignins in the predicted solvents was comparable with that in DMSO, which is a known good solvent for lignin. Interestingly, 2-methyltetrahydrofuran (2-MeTHF) which is a common sustainable solvent for organosolv processes [12,111], demonstrated mediocre lignin solubility (around 15 wt% in average).

Table 4

Identified solvents for selective lignin extraction. The distance from the optimal point $\vec{\alpha}(t, e) = [x_c, x_i]^T = [1, 0]^T$ is given as $d(\vec{\alpha}, \vec{x})$. In the table only solvent candidates with $d(\vec{\alpha}, \vec{x}) \leq 0.2$ are shown. Unless otherwise indicated, MPs, BPs, and EHS criteria were obtained from the respective safety data sheets.

Solvent	$d(\vec{\alpha}, \vec{x})$	$\log_{10}(x_c)$	$\log_{10}(x_i)$	MP [°C]	BP [°C]	EHS	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 [58]	
Diethyl-methylphosphonate	0.00	-2.74	0.00	n.a.	194	irritant	
Diethylsulfoxide	0.00	-2.51	0.00	14 [58]	n.a.	n.a.	
4-piperidino-pyridine	0.01	-2.33	0.00	78	n.a.	toxic, corrosive	
Dimethyl-methylphosphonate	0.01	-2.22	0.00	< 50 [58]	181 [58]	flammable, irritant, health hazard [58]	
4-pyrrolidino-pyridine	0.01	-2.31	0.00	54	n.a.	corrosive, acute toxic	
Dimethylsulfoxide	0.02	-1.66	0.00	16	189	irritant	
4-methoxypyridine	0.04	-2.84	-0.02	n.a.	n.a.	irritant	
Dimethyl ethylphosphonate	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.	
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-methyl-pyrimidine	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	

In general, COSMO-RS predictions were qualitatively in line with the experimental results (Fig. 6). For low lignin solubilities such as in n-heptane, dibutyl ether and 2-MeTHF, COSMO-RS predicted correct absolute lignin solubilities. For higher solubilities, COSMO-RS predictions deviated from the experimental results because (i) saturation in the solubility experiments was not reached for most of the solvents, (ii) the approximate COSMO-RS predictions converge to full dissolution for very high solubilities [73]. COSMO-RS correctly identified solvents with low solubilities which is crucial for the presented framework and allows to selectively exclude unsuitable solvents.

Most of the solvents shown in Table 4 and Fig. 6 have a high EHS score (> 0.8) and can be promising potential candidates for inclusion in the existing lignin treatment processes or for innovative process design. Opportunities are especially interesting for solvent candidates with “unusual” physical properties, such as high MP. Lignin treatment in such solvents will require reformulation (potentially for the better) or preparation of new procedures for recovery and recycling of these solvents. Inspiration can be taken from the processes with DESs as solvents, where addition of antisolvents (e.g. acetone), extraction, crystallization and other techniques are used to recover the main solvent [112].

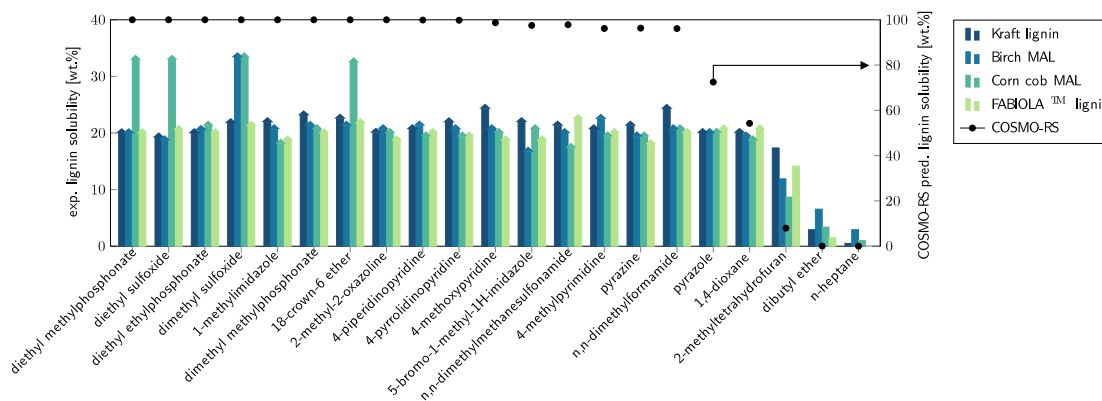
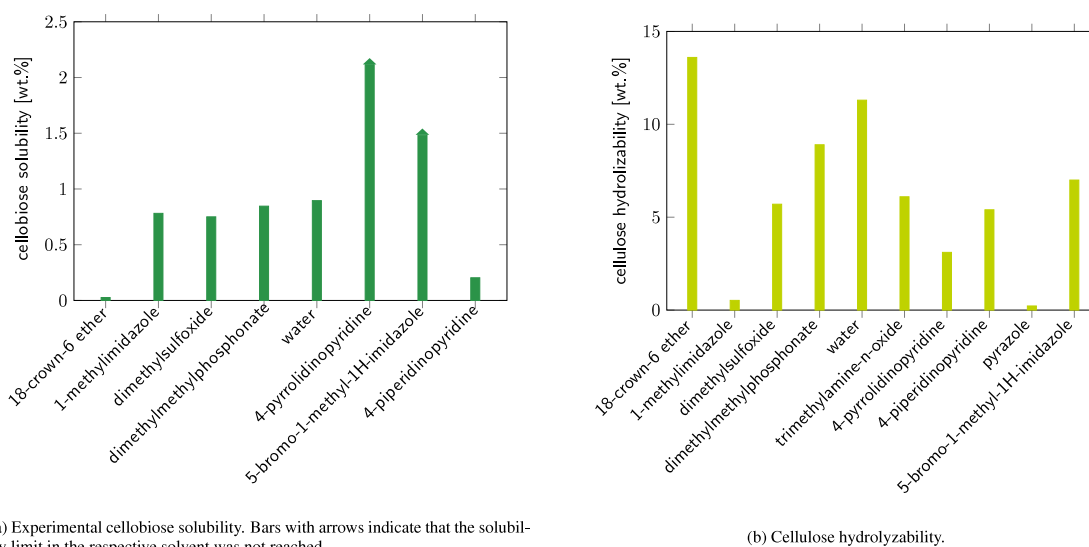


Fig. 6. Experimental solubilities in different solvents for Kraft lignin, Birch MAL, Corn cob MAL, and FABIOLA™ lignin (bars) and COSMO-RS solubility predictions (dots). The maximum solubility could not be reached for several solvents due to high viscosity of the solvent/lignin mixtures as indicated by bars with arrows. Please note the different axis scaling for COSMO-RS predicted lignin solubility and experimental values.



(a) Experimental cellobiose solubility. Bars with arrows indicate that the solubility limit in the respective solvent was not reached.

(b) Cellulose hydrolyzability.

Fig. 7. Cellobiose solubility and cellulose hydrolyzability were measured in various solvents. During cellobiose solubility measurements, rapid crystallization occurred in pyrazole and triethylamine-*n*-oxide, therefore, these solvents were excluded.

3.5. Experimental cellobiose solubility and cellulose hydrolyzability in the predicted solvents

Cellulose solubility measurement in organic solvents is challenging due to its highly recalcitrant structure arising from crystallinity and H-bonding network within the biopolymer inhibiting its dissolution. For this reason, this study mainly focused on investigating the solubility of cellobiose in the selected solvents. In addition, solvent effects on cellulose disruption and swelling, measured through its subsequent influence on acid hydrolysis of cellulose to glucose, was also analyzed as an indirect indication of cellulose swelling/disruption in different solvents. After mixing cellulose with each solvent followed by heating, the extent of swelling/disruption of cellulose in that solvent was characterized based on the amount of hydrolyzed cellulose, i.e. higher amount of cellulose hydrolysis indicates stronger disruption/swelling effects in the tested solvents.

Fig. 7(a) shows that cellobiose solubility was most effective in 4-pyrrolidinopyridine and 5-bromo-1-methyl-1H-imidazole followed by good solubility in water, dimethyl methylphosphonate, DMSO and 1-methylimidazole. Generally, this behavior of cellobiose solubility is in line with predicted cellulose solubilities. However, some exceptions are notable. Water, in which the capped cellotetraose model has a predicted solubility of 8.73 wt%, could easily solubilize cellobiose to a reasonable

extent in our tests. Similarly, 5-bromo-1-methyl-1H-imidazole showed a good ability to dissolve cellobiose (1.48 wt%) despite lower predicted cellulose solubility. On the other hand, 18-crown-6 ether and 4-piperidinopyridine, exhibited a significantly lower ability to dissolve cellobiose compared to other solvents despite higher predicted cellulose solubility. The measured cellobiose solubility, cellulose hydrolyzability and COSMO-RS predictions for the solubility of the cellulose fraction can be found in the ESI.

The effect of solvents on cellulose disrupting/swelling was generally significantly different than their effects on cellobiose solubility. Fig. 7(b) shows that 18-crown-6 ether and 4-piperidinopyridine both had a strong effect on cellulose disruption/swelling, as observed by hydrolyzed cellulose content in these solvents. Interestingly, these two solvents performed poorly in dissolution of cellobiose, demonstrating that the three-dimensional structure of cellulose has a big impact on its interaction with solvents. This also shows importance of developing new approaches towards plausible representation of cellulose for computational tools. Dimethyl methylphosphonate, DMSO, trimethylamine-*n*-oxide and 5-bromo-1-methyl-1H-imidazole affected cellulose hydrolysis to a relatively high extent (> 5 wt-%) likely due to their known ability to form strong hydrogen bonds. Trimethylamine-*n*-oxide self-associates similarly to surfactants, which may be necessary factors for disrupting cellulose [113]. Other solvents had relatively

moderate effects on cellulose disruption/swelling (≤ 5 wt% hydrolyzed cellulose) or low cellulose hydrolyzability with 1-methylimidazole and pyrazole < 1 wt% hydrolyzed cellulose. Pyrazole, as predicted by the model, had a poor effect on cellulose (nearly 0 wt% hydrolyzed). On the other hand, despite low predicted solubility, water showed high disruption/swelling (11.8 wt% hydrolyzed cellulose), confirming the same trend that we observed in the cellobiose experiment. In our computational approach, we cannot account for the highly specific and unusual behavior that water molecules normally exhibit [114]. Considering the challenges of accurately predicting and experimentally measuring cellulose solubility or even disruption and swelling in organic solvents, this model can be used as an initial indicator of solvents that might be effective in solubilizing cellulose. However, caution must be followed as exceptions exist that need further experimental validation. Experimental solubility tests with advanced techniques such as microscopic images, NMR using cellulose as well as its representative cello-oligosaccharides can be conducted to further investigate the accuracy of this model.

4. Conclusions

We developed a QM-based solvent screening framework for the processing of lignocellulosic biomass. The computational framework utilizes thoroughly assessed representative biomolecules for solubility predictions of the biomass fractions and QSAR models for the prediction of EHS properties. Notably, the tool allowed identification of non-intuitive organic solvents as well as ILs and DESs, featuring high lignin solubilities, sometimes exceeding 33 wt% as confirmed in the experiments. The screening approach elucidated functional groups beneficial for high lignin solubility, which were sulfoxides, azines, oxazolines and phosphonates. The identified solvents have favorable EHS properties and, thus, clear advantages over solvents being commonly used in lignocellulose processing which often suffer from either low biomass solubilities, such as alcohols, or high toxicity, such as 1,4-dioxane. Most of the identified chemicals are not frequently used or have very specific practical application in industry or laboratory, and are not commonly used as solvents. This paves a way for new opportunities for the utilization of these chemicals. In addition to solvents that might be useful in current “lignin-first” methods, our screening procedure identified solvents for the joint dissolution of all biomass fractions. Namely, the IL [EMIM][OAc] and 18-crown-6 ether were predicted to have high solubilities for all biomass fractions and additionally have green EHS criteria. The screening tool also revealed many interesting candidates among DESs. Although their ability to dissolve biomass was predicted to be lower than that for ILs and organic solvents, they are still promising green media since they are synthesized from renewables. The newly identified solvents present alternatives to the state-of-the-art organosolv and RCF processing and provide improvements in terms of solubility and EHS properties. These results open new possibilities for lignocellulose fractionation, process design, and deliver useful guidance for targeted lignocellulose solvent design. The proposed framework is applicable for other biomass sources and could facilitate rational solvent selection for biorefinery processes. As a future perspective, the viability and sustainability of the identified solvents in biorefinery processes including fractionation and downstream processing should be determined. Analyses of the identified chemicals in terms of stability, recovery, recycling, cost, and product selectivity are required for their implementation in a specific biorefinery process.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Experimental data is available in the ESI. Code can be made available on request, however, a COSMOtherm license is required.

Acknowledgments

LKM acknowledges funding by the Christiane Nüsslein-Volhard foundation, Germany. Open Access funding was provided by the Max Planck Society, Germany. This work was also supported by the Swiss National Science Foundation through the NCCR Catalysis (Grant no: 51NF40_180544) as well as Grants 200021_182605 and CRSIII_180258, as well as by the Swiss Competence Center for Energy Research: Biomass for a Swiss Energy Future through the Swiss Commission for Technology and Innovation Grant KTI.2014.0116, and by EPFL, Switzerland.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ccej.2022.139476>.

References

- J. Zakzeski, P.C.A. Bruijninx, A.L. Jongerius, B.M. Weckhuysen, The catalytic valorization of lignin for the production of renewable chemicals, *Chem. Rev.* 110 (6) (2010) 3552–3599, <http://dx.doi.org/10.1021/cr900354u>, URL <https://pubs.acs.org/doi/10.1021/cr900354u>.
- J. Ralph, K. Lundquist, G. Brunow, F. Lu, H. Kim, P.F. Schatz, J.M. Marita, R.D. Hatfield, S.A. Ralph, J.H. Christensen, W. Boerjan, Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids, *Phytochem. Rev.* 3 (1–2) (2004) 29–60, <http://dx.doi.org/10.1023/B:PHYT.0000047809.65444.a4>, URL <http://link.springer.com/10.1023/B:PHYT.0000047809.65444.a4>.
- M.M. Abu-Omar, Guidelines for performing lignin-first biorefining, *Environ. Sci.* (2021) 31.
- A.J. Ragauskas, G.T. Beckham, M.J. Biddy, 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, C.E. Wyman, Lignin valorization: improving lignin processing in the biorefinery, *Science* 344 (6185) (2014) 1246843, <http://dx.doi.org/10.1126/science.1246843>, URL <https://www.science.org/doi/10.1126/science.1246843>.
- R. Rinaldi, R. Jastrzebski, M.T. Clough, J. Ralph, M. Kennema, P.C.A. Bruijninx, B.M. Weckhuysen, Paving the way for lignin valorisation: recent advances in bioengineering, biorefining and catalysis, *Angew. Chem. Int. Ed.* 55 (29) (2016) 8164–8215, <http://dx.doi.org/10.1002/anie.201510351>, URL <https://onlinelibrary.wiley.com/doi/10.1002/anie.201510351>.
- A.K. Chandel, V.K. Garlapati, A.K. Singh, F.A.F. Antunes, S.S. silvada Silva, The path forward for lignocellulose biorefineries: bottlenecks, solutions, and perspective on commercialization, *Bioresour. Technol.* 264 (2018) 370–381, <http://dx.doi.org/10.1016/j.biortech.2018.06.004>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0960852418307831>.
- A.S. Mamman, J.-M. Lee, Y.-C. Kim, I.T. Hwang, N.-J. Park, Y.K. Hwang, J.-S. Chang, J.-S. Hwang, Furfural: hemicellulose/xyloederived biochemical, *Biofuels, Bioprod. Bioref.* 2 (5) (2008) 438–454, <http://dx.doi.org/10.1002/bbb.95>, URL <https://onlinelibrary.wiley.com/doi/10.1002/bbb.95>.
- E.I. Gürbüz, S.G. Wettstein, J.A. Dumesic, Conversion of hemicellulose to furfural and levulinic acid using biphasic reactors with alkylphenol solvents, *ChemSusChem* 5 (2) (2012) 383–387, <http://dx.doi.org/10.1002/cssc.201100608>, URL <https://onlinelibrary.wiley.com/doi/10.1002/cssc.201100608>.
- A.O. Komarova, G.R. Dick, J.S. Luterbacher, Diformylxylose as a new polar aprotic solvent produced from renewable biomass, *Green Chem.* 23 (13) (2021) 4790–4799, <http://dx.doi.org/10.1039/D1GC00641J>, URL <http://xlink.rsc.org/?DOI=D1GC00641J>.
- Y.M. Questell-Santiago, R. Zambrano-Varela, M. Talebi Amiri, J.S. Luterbacher, Carbohydrate stabilization extends the kinetic limits of chemical polysaccharide depolymerization, *Nat. Chem.* 10 (12) (2018) 1222–1228, <http://dx.doi.org/10.1038/s41557-018-0134-4>, URL <http://www.nature.com/articles/s41557-018-0134-4>.
- L. Manker, G. Dick, A. Demongeot, M. Hérou, C. Rayroud, T. Rambert, M. Jones, I. Sulaeva, Y. Leterrier, A. Potthast, F. Maréchal, V. Michaud, H.-A. Klok, J. Luterbacher, Sustainable Polyesters via Direct Functionalization of Lignocellulosic Sugars, *Chemistry*, 2021, <http://dx.doi.org/10.26434/chemrxiv-2021-9xwlv>, Preprint, URL <https://chemrxiv.org/engage/chemrxiv/article-details/61727d74c04e877c18bee11>.

- [12] W. Schutyser, T. Renders, S. Van den Bosch, S.-F. Koelewijn, G.T. Beckham, B.F. Sels, Chemicals from lignin: An interplay of lignocellulose fractionation, depolymerisation, and upgrading, *Chem. Soc. Rev.* 47 (3) (2018) 852–908, <http://dx.doi.org/10.1039/C7CS00566K>, URL <http://xlink.rsc.org/?DOI=C7CS00566K>.
- [13] X. Zhao, K. Cheng, D. Liu, Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis, *Appl. Microbiol. Biotechnol.* 82 (5) (2009) 815–827, <http://dx.doi.org/10.1007/s00253-009-1883-1>, URL <https://link.springer.com/10.1007/s00253-009-1883-1>.
- [14] M. Talebi Amiri, G.R. Dick, Y.M. Questell-Santiago, J.S. Luterbacher, Fractionation of lignocellulosic biomass to produce uncondensed aldehyde-stabilized lignin, *Nat. Protoc.* 14 (3) (2019) 921–954, <http://dx.doi.org/10.1038/s41596-018-0121-7>, URL <http://www.nature.com/articles/s41596-018-0121-7>.
- [15] Q. Song, F. Wang, J. Cai, Y. Wang, J. Zhang, W. Yu, J. Xu, Lignin depolymerization (LDP) in alcohol over nickel-based catalysts via a fragmentation-hydrogenolysis process, *Energy Environ. Sci.* 6 (3) (2013) 994, <http://dx.doi.org/10.1039/c2ee23741e>, URL <http://xlink.rsc.org/?DOI=c2ee23741e>.
- [16] P. Ferrini, R. Rinaldi, Catalytic biorefining of plant biomass to non-pyrolytic lignin bio-oil and carbohydrates through hydrogen transfer reactions, *Angew. Chem. Int. Ed.* 53 (33) (2014) 8634–8639, <http://dx.doi.org/10.1002/anie.201403747>, URL <https://onlinelibrary.wiley.com/doi/10.1002/anie.201403747>.
- [17] M.V. Galkin, J.S.M. Samec, Selective route to 2-propenyl aryls directly from wood by a tandem organosolv and palladium-catalysed transfer hydrogenolysis, *ChemSusChem* 7 (8) (2014) 2154–2158, <http://dx.doi.org/10.1002/cssc.201402017>, URL <https://onlinelibrary.wiley.com/doi/10.1002/cssc.201402017>.
- [18] 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, B.F. Sels, Reductive lignocellulose fractionation into soluble lignin-derived phenolic monomers and dimers and processable carbohydrate pulps, *Energy Environ. Sci.* 8 (6) (2015) 1748–1763, <http://dx.doi.org/10.1039/C5EE00204D>, URL <http://xlink.rsc.org/?DOI=C5EE00204D>.
- [19] T. Parsell, S. Yohe, J. Degenstein, T. Jarrell, I. Klein, E. Gencer, B. Hewetson, M. Hurt, J.I. Kim, H. Choudhari, B. Saha, R. Meilan, N. Mosier, F. Ribeiro, W.N. Delgass, C. Chapple, H.I. Kenttämaa, R. Agrawal, M.M. Abu-Omar, A synergistic biorefinery based on catalytic conversion of lignin prior to cellulose starting from lignocellulosic biomass, *Green Chem.* 17 (3) (2015) 1492–1499, <http://dx.doi.org/10.1039/C4GC01911C>, URL <http://xlink.rsc.org/?DOI=C4GC01911C>.
- [20] M.E. Himmel, S.-Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, T.D. Foust, Biomass recalcitrance: engineering plants and enzymes for biofuels production, *Science* 315 (5813) (2007) 804–807, <http://dx.doi.org/10.1126/science.1137016>, URL <https://www.science.org/doi/10.1126/science.1137016>.
- [21] J.D. DeMartini, S. Pattathil, J.S. Miller, H. Li, M.G. Hahn, C.E. Wyman, Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass, *Energy Environ. Sci.* 6 (3) (2013) 898, <http://dx.doi.org/10.1039/c3ee23801f>, URL <http://xlink.rsc.org/?DOI=c3ee23801f>.
- [22] J. Behaghel de Bueren, F. Héroguel, C. Wegmann, G.R. Dick, R. Buser, J.S. Luterbacher, Aldehyde-assisted fractionation enhances lignin valorization in endocarp waste biomass, *ACS Sustain. Chem. Eng.* 8 (45) (2020) 16737–16745, <http://dx.doi.org/10.1021/acssuschemeng.0c03360>, URL <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c03360>.
- [23] L. Shuai, J. Luterbacher, Organic solvent effects in biomass conversion reactions, *ChemSusChem* 9 (2) (2016) 133–155, <http://dx.doi.org/10.1002/cssc.201501148>, URL <https://onlinelibrary.wiley.com/doi/10.1002/cssc.201501148>.
- [24] P.P. Thoresen, L. Matsakas, U. Rova, P. Christakopoulos, Recent advances in organosolv fractionation: towards biomass fractionation technology of the future, *Bioresour. Technol.* 306 (2020) 123189, <http://dx.doi.org/10.1016/j.biortech.2020.123189>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0960852420304600>.
- [25] M.A. Mellmer, C. Sener, J.M.R. Gallo, J.S. Luterbacher, D.M. Alonso, J.A. Dumesic, Solvent effects in acid-catalyzed biomass conversion reactions, *Angew. Chem. Int. Ed.* 53 (44) (2014) 11872–11875, <http://dx.doi.org/10.1002/anie.201408359>, URL <https://onlinelibrary.wiley.com/doi/10.1002/anie.201408359>.
- [26] L. Soh, M.J. Eckelman, Green solvents in biomass processing, *ACS Sustain. Chem. Eng.* 4 (11) (2016) 5821–5837, <http://dx.doi.org/10.1021/acssuschemeng.6b01635>, URL <https://pubs.acs.org/doi/10.1021/acssuschemeng.6b01635>.
- [27] F. Cheng, X. Zhao, Y. Hu, Lignocellulosic biomass delignification using aqueous alcohol solutions with the catalysis of acidic ionic liquids: A comparison study of solvents, *Bioresour. Technol.* 249 (2018) 969–975, <http://dx.doi.org/10.1016/j.biortech.2017.10.089>, URL <https://linkinghub.elsevier.com/retrieve/pii/S096085241731920X>.
- [28] A. Ghosh, M.R. Haverly, J.K. Lindstrom, P.A. Johnston, R.C. Brown, Tetrahydrofuran-based two-step solvent liquefaction process for production of lignocellulosic sugars, *React. Chem. Eng.* 5 (9) (2020) 1694–1707, <http://dx.doi.org/10.1039/D0RE00192A>, URL <http://xlink.rsc.org/?DOI=D0RE00192A>.
- [29] A. Kulshrestha, I. Pancha, S. Mishra, A. Kumar, Deep eutectic solvents and ionic liquid assisted hydrolysis of microalgal biomass: A promising approach towards sustainable biofuel production, *J. Mol. Liq.* 335 (2021) 116264, <http://dx.doi.org/10.1016/j.molliq.2021.116264>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0167732221009880>.
- [30] D.J.G.P. oschvan Osch, L.J.B.M. Kollau, A. bruinhorstvan den Bruinhorst, S. Asikainen, M.A.A. Rocha, M.C. Kroon, Ionic liquids and deep eutectic solvents for lignocellulosic biomass fractionation, *Phys. Chem. Chem. Phys.* 19 (4) (2017) 2636–2665, <http://dx.doi.org/10.1039/C6CP07499E>, URL <http://xlink.rsc.org/?DOI=C6CP07499E>.
- [31] C. Balaji, T. Banerjee, V.V. Goud, COSMO-RS based predictions for the extraction of lignin from lignocellulosic biomass using ionic liquids: effect of cation and anion combination, *J. Solut. Chem.* 41 (9) (2012) 1610–1630, <http://dx.doi.org/10.1007/s10953-012-9887-3>, URL <http://link.springer.com/10.1007/s10953-012-9887-3>.
- [32] E.C. Achinivu, M. Mohan, H. Choudhary, L. Das, K. Huang, H.D. Magurudeniya, V.R. Pidatala, A. George, B.A. Simmons, J.M. Gladden, A predictive toolset for the identification of effective lignocellulosic pretreatment solvents: A case study of solvents tailored for lignin extraction, *Green Chem.* 23 (18) (2021) 7269–7289, <http://dx.doi.org/10.1039/D1GC01186C>, URL <http://xlink.rsc.org/?DOI=D1GC01186C>.
- [33] Y. Chu, X. He, MoDooP: an automated computational approach for COSMO-RS prediction of biopolymer solubilities in ionic liquids, *ACS Omega* 4 (1) (2019) 2337–2343, <http://dx.doi.org/10.1021/acsomega.8b03255>, URL <https://pubs.acs.org/doi/10.1021/acsomega.8b03255>.
- [34] J. Kahlen, K. Masuch, K. Leonhard, Modelling cellulose solubilities in ionic liquids using COSMO-RS, *Green Chem.* 12 (12) (2010) 2172, <http://dx.doi.org/10.1039/c0gc00200c>, URL <http://xlink.rsc.org/?DOI=c0gc00200c>.
- [35] Y. Chu, X. Zhang, M. Hillestad, X. He, Computational prediction of cellulose solubilities in ionic liquids based on COSMO-RS, *Fluid Phase Equilib.* 475 (2018) 25–36, <http://dx.doi.org/10.1016/j.fluid.2018.07.026>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0378381218302978>.
- [36] M. Mohan, C. Balaji, V.V. Goud, T. Banerjee, Thermodynamic insights in the separation of cellulose/hemicellulose components from lignocellulosic biomass using ionic liquids, *J. Solut. Chem.* 44 (3–4) (2015) 538–557, <http://dx.doi.org/10.1007/s10953-015-0295-3>, URL <http://link.springer.com/10.1007/s10953-015-0295-3>.
- [37] M. Mohan, P. Viswanath, T. Banerjee, V.V. Goud, Multiscale modelling strategies and experimental insights for the solvation of cellulose and hemicellulose in ionic liquids, *Mol. Phys.* 116 (15–16) (2018) 2108–2128, <http://dx.doi.org/10.1080/00268976.2018.1447152>, URL <https://www.tandfonline.com/doi/full/10.1080/00268976.2018.1447152>.
- [38] P. Yamin, COSMO-RS-Based Methods for Improved Modelling of Complex Chemical Systems (Ph.D. thesis), RWTH Aachen, Aachen, 2019.
- [39] A. Casas, J. Palomar, M.V. Alonso, M. Oliet, S. Omar, F. Rodriguez, Comparison of lignin and cellulose solubilities in ionic liquids by COSMO-RS analysis and experimental validation, *Ind. Crops Prod.* 37 (1) (2012) 155–163, <http://dx.doi.org/10.1016/j.indcrop.2011.11.032>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0926669011004596>.
- [40] A. Casas, S. Omar, J. Palomar, M. Oliet, M.V. Alonso, F. Rodriguez, Relation between differential solubility of cellulose and lignin in ionic liquids and activity coefficients, *RSC Adv.* 3 (10) (2013) 3453, <http://dx.doi.org/10.1039/c2ra22800a>, URL <http://xlink.rsc.org/?DOI=c2ra22800a>.
- [41] S. Linke, K. McBride, K. Sundmacher, Systematic green solvent selection for the hydroformylation of long-chain alkenes, *ACS Sustain. Chem. Eng.* (2020) <http://dx.doi.org/10.1021/acssuschemeng.0c02611>, <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c02611>.
- [42] L. König-Mattern, S. Linke, L. Rihko-Struckmann, K. Sundmacher, Computer-aided solvent screening for the fractionation of wet microalgae biomass, *Green Chem.* (2021) <http://dx.doi.org/10.1039/D1GC03471E>, URL <http://xlink.rsc.org/?DOI=D1GC03471E>.
- [43] L. Moity, M. Durand, A. Benazzouz, C. Pierlot, V. Molinier, J.-M. Aubry, Panorama of sustainable solvents using the COSMO-RS approach, *Green Chem.* 14 (4) (2012) 1132–1145, <http://dx.doi.org/10.1039/c2gc16515e>, URL <http://xlink.rsc.org/?DOI=c2gc16515e>.
- [44] J.E. Camp, Bio-available solvent cyrene: synthesis, derivatization, and applications, *ChemSusChem* 11 (18) (2018) 3048–3055, <http://dx.doi.org/10.1002/cssc.201801420>, URL <http://doi.wiley.com/10.1002/cssc.201801420>.
- [45] A. Paiva, R. Craveiro, I. Aroso, M. Martins, R.L. Reis, A.R.C. Duarte, Natural deep eutectic solvents – solvents for the 21st century, *ACS Sustain. Chem. Eng.* 2 (5) (2014) 1063–1071, <http://dx.doi.org/10.1021/sc500096j>.
- [46] Q. Zhang, K. De Oliveira Vigier, S. Royer, F. Jérôme, Deep eutectic solvents: Syntheses, properties and applications, *Chem. Soc. Rev.* 41 (21) (2012) 7108–7146, <http://dx.doi.org/10.1039/C2CS35178A>, URL <http://dx.doi.org/10.1039/C2CS35178A>.
- [47] 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, <http://dx.doi.org/10.1016/B978-0-12-815577-6.00016-5>, URL <https://linkinghub.elsevier.com/retrieve/pii/B9780128155776000165>.
- [48] E. Slupek, P. Makoš, J. Gebicki, Theoretical and economic evaluation of low-cost deep eutectic solvents for effective biogas upgrading to bio-methane, *Energies* 13 (13) (2020) 3379, <http://dx.doi.org/10.3390/en13133379>, URL <https://www.mdpi.com/1996-1073/13/13/3379>.

- [49] K. Jeong, M. Yang, Y. Jin, E. Kim, J. Ko, J. Lee, Identification of major flavone C-glycosides and their optimized extraction from cymbidium kanran using deep eutectic solvents, *Molecules* 22 (11) (2017) 2006, <http://dx.doi.org/10.3390/molecules22112006>, URL <http://www.mdpi.com/1420-3049/22/11/2006>.
- [50] K. Radošević, N. Čurko, V. Gaurina Srček, M. Cvjetko Bubalo, M. Tomašević, K. Kovačević Ganić, I. Radojčić Redovniković, Natural deep eutectic solvents as beneficial extractants for enhancement of plant extracts bioactivity, *LWT-Food Sci. Technol.* 73 (2016) 45–51, <http://dx.doi.org/10.1016/j.lwt.2016.05.037>, URL <https://linkinghub.elsevier.com/retrieve/pii/S002364381630295X>.
- [51] T. Křížek, M. Bursová, R. Horsley, M. Kuchař, P. Tůma, R. Čabala, T. Hložek, Menthol-based hydrophobic deep eutectic solvents: towards greener and efficient extraction of phytocannabinoids, *J. Clean. Prod.* 193 (2018) 391–396, <http://dx.doi.org/10.1016/j.jclepro.2018.05.080>.
- [52] M.W. Nam, J. Zhao, M.S. Lee, J.H. Jeong, J. Lee, Enhanced extraction of bioactive natural products using tailor-made deep eutectic solvents: Application to flavonoid extraction from flos sophorae, *Green Chem.* 17 (3) (2015) 1718–1727, <http://dx.doi.org/10.1039/C4GC01556H>, URL <http://xlink.rsc.org/?DOI=C4GC01556H>.
- [53] C. Florindo, L.C. Branco, I.M. Marrucho, Quest for green-solvent design: from hydrophilic to hydrophobic (deep) eutectic solvents, *ChemSusChem* 12 (8) (2019) 1549–1559, <http://dx.doi.org/10.1002/cssc.201900147>, URL <https://onlinelibrary.wiley.com/doi/10.1002/cssc.201900147>.
- [54] W. Tang, Y. Dai, K.H. Row, Evaluation of fatty acid/alcohol-based hydrophobic deep eutectic solvents as media for extracting antibiotics from environmental water, *Anal. Bioanal. Chem.* 410 (28) (2018) 7325–7336, <http://dx.doi.org/10.1007/s00216-018-1346-6>, URL <http://link.springer.com/10.1007/s00216-018-1346-6>.
- [55] H. Vanda, Y. Dai, E.G. Wilson, R. Verpoorte, Y.H. Choi, Green solvents from ionic liquids and deep eutectic solvents to natural deep eutectic solvents, *C. R. Chim.* 21 (6) (2018) 628–638, <http://dx.doi.org/10.1016/j.crci.2018.04.002>, URL <https://linkinghub.elsevier.com/retrieve/pii/S1631074818301048>.
- [56] Iolitec, Iolitec catalogue, 2020, URL https://iolitec.de/products/ionic_liquids/catalogue (last Accessed: 20 April 2020).
- [57] SigmaAldrich, *ChemFiles* vol. 5 no. 6, 2005.
- [58] PubChem, 2022, URL <https://pubchem.ncbi.nlm.nih.gov/> (last Accessed: 9 July 2022).
- [59] B. Pereira, V. Arantes, Nanocelluloses from sugarcane biomass, in: *Advances in Sugarcane Biorefinery*, Elsevier, Amsterdam, Netherlands, 2018, p. 181, <http://dx.doi.org/10.1016/C2015-0-02033-0>, URL <https://linkinghub.elsevier.com/retrieve/pii/C20150020330>.
- [60] A. Tolbert, H. Akinoshio, R. Khunupat, A.K. Naskar, A.J. Ragauskas, Characterization and analysis of the molecular weight of lignin for biorefining studies, *Biofuels, Bioprod. Bioref.* 8 (6) (2014) 836–856, <http://dx.doi.org/10.1002/bbb.1500>, URL <https://onlinelibrary.wiley.com/doi/10.1002/bbb.1500>.
- [61] D. Klemm, B. Heublein, H.-P. Fink, A. Bohn, Cellulose: fascinating biopolymer and sustainable raw material, *Angew. Chem. Int. Ed.* 44 (22) (2005) 3358–3393, <http://dx.doi.org/10.1002/anie.200460587>, URL <https://onlinelibrary.wiley.com/doi/10.1002/anie.200460587>.
- [62] C. Olsson, G. Westm, Direct dissolution of cellulose: background, means and applications, in: T.G. Van De Ven (Ed.), *Cellulose - Fundamental Aspects*, InTech, 2013, pp. 143–178, <http://dx.doi.org/10.5772/52144>, URL <http://www.intechopen.com/books/cellulose-fundamental-aspects/direct-dissolution-of-cellulose-background-means-and-applications>.
- [63] S. Köhler, T. Heinze, New solvents for cellulose: dimethyl sulfoxide/ammonium fluorides, *Macromol. Biosci.* 7 (3) (2007) 307–314, <http://dx.doi.org/10.1002/mabi.200600197>, URL <https://onlinelibrary.wiley.com/doi/10.1002/mabi.200600197>.
- [64] A. Potthast, T. Rosenau, H. Sixta, P. Kosma, Degradation of cellulosic materials by heating in DMAc/liCl, *Tetrahedron Lett.* 43 (43) (2002) 7757–7759, [http://dx.doi.org/10.1016/S0040-4039\(02\)01767-7](http://dx.doi.org/10.1016/S0040-4039(02)01767-7), URL <https://linkinghub.elsevier.com/retrieve/pii/S0040403902017677>.
- [65] S. Bertella, J.S. Luterbacher, Lignin functionalization for the production of novel materials, *Trends Chem.* 2 (5) (2020) 440–453, <http://dx.doi.org/10.1016/j.trechm.2020.03.001>, URL <https://linkinghub.elsevier.com/retrieve/pii/S2589597420300551>.
- [66] R. Vanholme, K. Morreel, C. Darrach, P. Oyarce, J.H. Grabber, J. Ralph, W. Boerjan, Metabolic engineering of novel lignin in biomass crops, *New Phytol.* 196 (4) (2012) 978–1000, <http://dx.doi.org/10.1111/j.1469-8137.2012.04337.x>, URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2012.04337.x>.
- [67] M.V. Galkin, J.S.M. Samec, Lignin valorization through catalytic lignocellulose fractionation: a fundamental platform for the future biorefinery, *ChemSusChem* 9 (13) (2016) 1544–1558, <http://dx.doi.org/10.1002/cssc.201600237>, URL <https://onlinelibrary.wiley.com/doi/10.1002/cssc.201600237>.
- [68] R. Parthasarathi, R.A. Romero, A. Redondo, S. Gnanakaran, Theoretical study of the remarkably diverse linkages in lignin, *J. Phys. Chem. Lett.* 2 (20) (2011) 2660–2666, <http://dx.doi.org/10.1021/jz201201q>, URL <https://pubs.acs.org/doi/10.1021/jz201201q>.
- [69] P. Sannigrahi, A.J. Ragauskas, G.A. Tuskan, Poplar as a feedstock for biofuels: A review of compositional characteristics, *Biofuels, Bioprod. Bioref.* 4 (2) (2010) 209–226, <http://dx.doi.org/10.1002/bbb.206>, URL <https://onlinelibrary.wiley.com/doi/10.1002/bbb.206>.
- [70] F. Gírio, C. Fonseca, F. Carvalheiro, L. Duarte, S. Marques, R. Bogel-Lukasik, Hemicelluloses for fuel Ethanol: A review, *Bioresour. Technol.* 101 (13) (2010) 4775–4800, <http://dx.doi.org/10.1016/j.biortech.2010.01.088>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0960852410001744>.
- [71] H.V. Scheller, P. Ulvskov, Hemicelluloses, *Annu. Rev. Plant Biol.* 61 (1) (2010) 263–289, <http://dx.doi.org/10.1146/annurev-arplant-042809-112315>, URL <https://www.annualreviews.org/doi/10.1146/annurev-arplant-042809-112315>.
- [72] D. Tarasov, M. Leitch, P. Fatehi, Lignin-carbohydrate complexes: Properties, applications, analyses, and methods of extraction: A review, *Biotechnol. Biofuels* 11 (1) (2018) 269, <http://dx.doi.org/10.1186/s13068-018-1262-1>, URL <https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-018-1262-1>.
- [73] Dassault Systèmes company, *Cosmotherm reference manual*, 2019.
- [74] E. Benefenati, A. Manganaro, G. Gini, VEGA-qsar: AI inside a platform for predictive toxicology, in: *CEUR Workshop Proc.*, (1107) 2013, pp. 21–28.
- [75] Z. Song, X. Hu, H. Wu, M. Mei, S. Linke, T. Zhou, Z. Qi, K. Sundmacher, Systematic screening of deep eutectic solvents as sustainable separation media exemplified by the CO₂ capture process, *ACS Sustain. Chem. Eng.* 8 (23) (2020) 8741–8751, <http://dx.doi.org/10.1021/acssuschemeng.0c02490>, URL <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c02490>.
- [76] Rdkit: open-source cheminformatics, 2022, URL <https://www.rdkit.org> (last Accessed: 19 July 2022).
- [77] J.-P. Ebejer, G.M. Morris, C.M. Deane, Freely available conformer generation methods: how good are they? *J. Chem. Inf. Model.* 52 (5) (2012) 1146–1158, <http://dx.doi.org/10.1021/ci2004658>, URL <https://pubs.acs.org/doi/10.1021/ci2004658>.
- [78] COSMOtherm, Release 19; © 2019, COSMOlogic GmbH & Co. KG, a Dassault Systèmes Company.
- [79] F. Eckert, A. Klamt, Fast solvent screening via quantum chemistry: COSMO-RS approach, *AIChE J.* 48 (2) (2002) 369–385, <http://dx.doi.org/10.1002/aic.690480220>, URL <http://doi.wiley.com/10.1002/aic.690480220>.
- [80] A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.* 99 (7) (1995) 2224–2235, <http://dx.doi.org/10.1021/j100007a062>, URL <https://pubs.acs.org/doi/10.1021/j100007a062>.
- [81] A. Klamt, F. Eckert, COSMO-RS: A novel and efficient method for the a priori prediction of thermophysical data of liquids, *Fluid Phase Equilib.* 172 (1) (2000) 43–72, [http://dx.doi.org/10.1016/S0378-3812\(00\)00357-5](http://dx.doi.org/10.1016/S0378-3812(00)00357-5), URL <https://linkinghub.elsevier.com/retrieve/pii/S0378381200003575>.
- [82] A. Klamt, V. Jonas, T. Bürger, J.C.W. Lohrenz, Refinement and parametrization of COSMO-RS, *J. Phys. Chem. A* 102 (26) (1998) 5074–5085, <http://dx.doi.org/10.1021/jp980017s>, URL <https://pubs.acs.org/doi/10.1021/jp980017s>.
- [83] F. Bezold, M.E. Weinberger, M. Minceva, Computational solvent system screening for the separation of tocopherols with centrifugal partition chromatography using deep eutectic solvent-based biphasic systems, *J. Chromatogr. A* 1491 (2017) 153–158, <http://dx.doi.org/10.1016/j.chroma.2017.02.059>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0021967317303175>.
- [84] H.F. Hizaddin, A. Ramalingam, M.A. Hashim, 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, *J. Chem. Eng. Data* 59 (11) (2014) 3470–3487, <http://dx.doi.org/10.1021/je5004302>, URL <https://pubs.acs.org/doi/10.1021/je5004302>.
- [85] M. Diedenhofen, A. Klamt, COSMO-RS as a tool for property prediction of IL mixtures—a review, *Fluid Phase Equilib.* 294 (1–2) (2010) 31–38, <http://dx.doi.org/10.1016/j.fluid.2010.02.002>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0378381210000695>.
- [86] A. Smit, W. Huijgen, Effective fractionation of lignocellulose in herbaceous biomass and hardwood using a mild acetone organosolv process, *Green Chem.* 19 (22) (2017) 5505–5514, <http://dx.doi.org/10.1039/C7GC02379K>, URL <http://xlink.rsc.org/?DOI=C7GC02379K>.
- [87] A. Smit, M. Verges, P. Schulze, A. zomerenvan Zomeren, H. Lorenz, Laboratory-to pilot-scale fractionation of lignocellulosic biomass using an acetone organosolv process, *ACS Sustain. Chem. Eng.* (in press) (2022) <http://dx.doi.org/10.1021/acssuschemeng.2c01425>.
- [88] A. Das, A. Rahimi, A. Ulbrich, M. Alherrech, A.H. Motagamwala, A. Bhalla, L. costa sousada Costa Sousa, V. Balan, J.A. Dumesic, E.L. Hegg, B.E. Dale, J. Ralph, J.J. Coon, S.S. Stahl, Lignin conversion to low-molecular-weight aromatics via an aerobic oxidation-hydrolysis sequence: comparison of different lignin sources, *ACS Sustain. Chem. Eng.* 6 (3) (2018) 3367–3374, <http://dx.doi.org/10.1021/acssuschemeng.7b03541>, URL <https://pubs.acs.org/doi/10.1021/acssuschemeng.7b03541>.
- [89] G.R. Dick, A.O. Komarova, J.S. Luterbacher, Controlling lignin solubility and hydrogenolysis selectivity by acetal-mediated functionalization, *Green Chem.* 24 (3) (2022) 1285–1293, <http://dx.doi.org/10.1039/D1GC02575A>, URL <http://xlink.rsc.org/?DOI=D1GC02575A>.

- [90] J. Heng, Z. Zhang, E. Proctor, M. Tyufekchiev, N.A. Deskins, M.T. Timko, Cellobiose as a model carbohydrate for predicting solubilities in nonaqueous solvents, *Ind. Eng. Chem. Res.* 60 (4) (2021) 1859–1871, <http://dx.doi.org/10.1021/acs.iecr.0c04963>, URL <https://pubs.acs.org/doi/10.1021/acs.iecr.0c04963>.
- [91] X. Zhang, T. Qu, N.S. Mosier, L. Han, W. Xiao, Cellulose modification by recyclable swelling solvents, *Biotechnol. Biofuels* 11 (1) (2018) 191, <http://dx.doi.org/10.1186/s13068-018-1191-z>, URL <https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-018-1191-z>.
- [92] P.A. Johnston, R.C. Brown, Quantitation of sugar content in pyrolysis liquids after acid hydrolysis using high-performance liquid chromatography without neutralization, *J. Agric. Food Chem.* 62 (32) (2014) 8129–8133, <http://dx.doi.org/10.1021/jf502250n>, URL <https://pubs.acs.org/doi/10.1021/jf502250n>.
- [93] M.A. Mellmer, D. Martin Alonso, J.S. Luterbacher, J.M.R. Gallo, J.A. Dumesic, Effects of γ -valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides, *Green Chem.* 16 (11) (2014) 4659–4662, <http://dx.doi.org/10.1039/C4GC01768D>, URL <http://xlink.rsc.org/?DOI=C4GC01768D>.
- [94] J. Sameni, S. Krigstin, M. Sain, Solubility of lignin and acetylated lignin in organic solvents, *BioResources* 12 (1) (2017) 1548–1565, <http://dx.doi.org/10.15376/biores.12.1.1548-1565>, URL <http://ojs.cnr.ncsu.edu/index.php/BioRes/article/view/10881>.
- [95] J.V. Vermaas, M.F. Crowley, G.T. Beckham, Molecular lignin solubility and structure in organic solvents, *ACS Sustain. Chem. Eng.* 8 (48) (2020) 17839–17850, <http://dx.doi.org/10.1021/acssuschemeng.0c07156>, URL <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c07156>.
- [96] J. Vitz, T. Erdmenger, C. Haensch, U.S. Schubert, Extended dissolution studies of cellulose in imidazolium based ionic liquids, *Green Chem.* 11 (3) (2009) 417, <http://dx.doi.org/10.1039/b818061j>, URL <http://xlink.rsc.org/?DOI=b818061j>.
- [97] H. Zhao, G.A. Baker, Z. Song, O. Olubajo, T. Crittle, D. Peters, Designing enzyme-compatible ionic liquids that can dissolve carbohydrates, *Green Chem.* 10 (6) (2008) 696, <http://dx.doi.org/10.1039/b801489b>, URL <http://xlink.rsc.org/?DOI=b801489b>.
- [98] S. Jadhav, A. Lidhure, S. Thakre, V. Ganvir, Modified lyocell process to improve dissolution of cellulosic pulp and pulp blends in NMMO solvent, *Cellulose* 28 (2) (2021) 973–990, <http://dx.doi.org/10.1007/s10570-020-03580-1>, URL <http://link.springer.com/10.1007/s10570-020-03580-1>.
- [99] S. Arora, N. Gupta, V. Singh, Choline based basic ionic liquid (BIL)/acidic DES mediated cellulose rich fractionation of agricultural waste biomass and valorization to 5-HMF, *Waste Biomass Valor.* 11 (7) (2020) 3345–3354, <http://dx.doi.org/10.1007/s12649-019-00603-2>, URL <http://link.springer.com/10.1007/s12649-019-00603-2>.
- [100] Y. Zhong, J. Wu, H. Kang, R. Liu, Choline hydroxide based deep eutectic solvent for dissolving cellulose, *Green Chem.* (2022) 12.
- [101] K. Glińska, J. Gitalt, E. Torrens, N. Plechkova, C. Bengoa, Extraction of cellulose from corn stover using designed ionic liquids with improved reusing capabilities, *Process Saf. Environ. Prot.* 147 (2021) 181–191, <http://dx.doi.org/10.1016/j.psep.2020.09.035>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0957582020317523>.
- [102] F.P. Burns, P.-A. Themens, K. Ghandi, Assessment of phosphonium ionic liquid-dimethylformamide mixtures for dissolution of cellulose, *Compos. Interfaces* 21 (1) (2014) 59–73, <http://dx.doi.org/10.1080/15685543.2013.831669>, URL <http://www.tandfonline.com/doi/full/10.1080/15685543.2013.831669>.
- [103] P. Mäki-Arvela, I. Anugwom, P. Virtanen, R. Sjöholm, J. Mikkola, Dissolution of lignocellulosic materials and its constituents using ionic liquids—a review, *Ind. Crops Prod.* 32 (3) (2010) 175–201, <http://dx.doi.org/10.1016/j.indcrop.2010.04.005>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0926669010000841>.
- [104] M. Zavrel, D. Bross, M. Funke, J. Büchs, A.C. Spiess, High-throughput screening for ionic liquids dissolving (ligno-)cellulose, *Bioresour. Technol.* 100 (9) (2009) 2580–2587, <http://dx.doi.org/10.1016/j.biortech.2008.11.052>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0960852408010638>.
- [105] Y.-L. Chen, X. Zhang, T.-T. You, F. Xu, Deep eutectic solvents (DESs) for cellulose dissolution: A mini-review, *Cellulose* 26 (1) (2019) 205–213, <http://dx.doi.org/10.1007/s10570-018-2130-7>, URL <http://link.springer.com/10.1007/s10570-018-2130-7>.
- [106] A.S. Filatov, E. Block, M.A. Petrukhina, Dimethyl selenoxide, *Acta Crystallogr. C Cryst. Struct. Commun.* 61 (10) (2005) o596–o598, <http://dx.doi.org/10.1107/S0108270105027587>, URL <http://scripts.iucr.org/cgi-bin/paper?S0108270105027587>.
- [107] NIST webbook, 2022, URL <https://webbook.nist.gov/chemistry/> (last Accessed: 9 July 2022).
- [108] J. Sameni, S. Krigstin, M. Sain, Characterization of lignins isolated from industrial residues and their beneficial uses, *BioResources* 11 (4) (2016) 8435–8456, <http://dx.doi.org/10.15376/biores.11.4.8435-8456>, URL <http://ojs.cnr.ncsu.edu/index.php/BioRes/article/view/9919>.
- [109] A. Duval, F. Vilaplana, C. Crestini, M. Lawoko, Solvent screening for the fractionation of industrial kraft lignin, *Holzforschung* 70 (1) (2016) 11–20, <http://dx.doi.org/10.1515/hf-2014-0346>, URL <https://www.degruyter.com/document/doi/10.1515/hf-2014-0346/html>.
- [110] A. Salaghi, L. Zhou, P. Saini, F. Kong, M. Konduri, P. Fatehi, Lignin production in plants and pilot and commercial processes, in: *Biomass, Biofuels, Biochemicals*, Elsevier, 2021, pp. 551–587, <http://dx.doi.org/10.1016/B978-0-12-821888-4.00002-2>, URL <https://linkinghub.elsevier.com/retrieve/pii/B9780128218884000022>.
- [111] P.M. Grande, J. Viell, N. Theyssen, W. Marquardt, P. Domínguez de María, W. Leitner, Fractionation of lignocellulosic biomass using the OrganoCat process, *Green Chem.* 17 (6) (2015) 3533–3539, <http://dx.doi.org/10.1039/C4GC02534B>, URL <http://xlink.rsc.org/?DOI=C4GC02534B>.
- [112] A. Isci, M. Kaltschmitt, Recovery and recycling of deep eutectic solvents in biomass conversions: A review, *Biomass Conv. Bioref.* 12 (S1) (2022) 197–226, <http://dx.doi.org/10.1007/s13399-021-01860-9>, URL <https://link.springer.com/10.1007/s13399-021-01860-9>.
- [113] J. Ma, I.M. Pazos, F. Gai, Microscopic insights into the protein-stabilizing effect of trimethylamine N-oxide (TMAO), *Proc. Natl. Acad. Sci. USA* 111 (23) (2014) 8476–8481, <http://dx.doi.org/10.1073/pnas.1403224111>, URL <https://pnas.org/doi/full/10.1073/pnas.1403224111>.
- [114] P. Chen, J. Wohler, L. Berglund, I. Furó, Water as an intrinsic structural element in cellulose fibril aggregates, *J. Phys. Chem. Lett.* 13 (24) (2022) 5424–5430, <http://dx.doi.org/10.1021/acs.jpcclett.2c00781>, URL <https://pubs.acs.org/doi/10.1021/acs.jpcclett.2c00781>.