Green Chemistry

Molecular simulations inform biomass dissolution in ionic liquids in pursuit of benign solvent-system design

Journal:	*Green Chemistry*
Manuscript ID	GC-ART-06-2023-001981.R1
Article Type:	Paper
Date Submitted by the Author:	31-Jul-2023
Complete List of Authors:	Griffin, Preston; The George Washington University, Chemistry Kostal, Jakub; The George Washington University, Chemistry

ARTICLE

Molecular simulations inform biomass dissolution in ionic liquids in pursuit of benign solvent-system design

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Preston Griffin[a] and Jakub Kostal*[a]

When we look for a poster child of green chemistry 'in action', we do not need to look further than the deconstruction of lignocellulose using ionic liquids (IL) to valorize this renewable resource into useful chemicals. However, there is a caveat: successful development of new chemistries cannot be achieved without systems-based design tools that consider performance in conjunction with potential toxicity. Here, we show that a combination of computational approaches, based on quantum mechanics (QM) calculations and Monte Carlo (MC) simulations, can be leveraged to construct a useful framework for screening existing and designing new ILs capable of safe and selective dissolution of lignocellulosic biomass. With the overwhelming number of IL cation-anion combinations, *in silico* methods are uniquely suited for this challenge so long as they retain mechanistic relevance to the underlying processes. Our computational approach ensures this criterion by relying on well-correlated linear models of interaction energetics between IL and key biomass building blocks. Functional considerations are supplemented with frontier molecular orbital calculations to determine safety toward aquatic species based on previously established and broadly validated guidelines.

Introduction

Climate change is primarily linked to the increased use of fossil fuels. In 1900, with a population of 1.7 billion, the world employed 5,973 terawatt-hours (TWh) of fossil fuels.[1,2] By 2019, with an increased population of 7.8 billion, fossil fuel usage skyrocketed to 136,000 TWh from coal (32%), oil (39%) and gas (32%).[1,2] In the United States, fossil fuels make up over 80% of the fuel composition and are the primary chemical feedstock that contributes to climate change.[1,2] A decreased dependence on petrochemical feedstock could come from effective utilization of biomass; lignocellulosic biomass is particularly attractive, as it is the most abundant and available biomass on Earth.[3] While cellulose and hemicellulose have found numerous industrial applications, lignin usage has been almost negligible by comparison,[4–8] owing to its more variable and complex structure. In 2004, the pulp and paper industry alone generated 50 million tons of lignin, with only 2%-3% used in further applications.[9] While lignin may be difficult to utilize, it has considerable economic potential. The Pacific Northwest National Laboratory, in collaboration with the National Renewable Energy Laboratory, reported that successful implementation of lignin conversion technologies to produce value-added chemicals could enhance current lignin revenue by more than $11 billion.[10]

Lignin is a heterogenous polyaromatic polymer, primarily consisting of paracoumaryl alcohol, coniferyl alcohol and sinapyl alcohol.[3,11] These alcohols are linked through carbon-carbon or ether bonds, with the three most common linkages being the β-1 linkage (ca. 6%–10%), 5–5' linkage (ca. 10%–25%), and β-O-4 linkage (ca. 30%–80%).[3,4,11,12] Other linkages such as β-5, β-β, 4-O-5, and dibenzodioxocin may be present in significant percentages, as composition can vary widely depending on the biomass source.[4,13–15] This variability poses challenges in designing efficient systems for biomass deconstruction.[4,9] In biomass valorization, dissolution is the first key step, where noncovalent interactions between polymer chains are broken. Here, ionic liquids (ILs) offer many advantages over traditional solvents including tunability to optimize performance and a potentially lower risk to human and environmental health, in part due to their low volatility.[16]

In designing new ILs that can dissolve biomass safely and efficiently, we must understand the mode of function (i.e., performance) and the mode of action, MOA (i.e., toxicity). Considering the former, lignin and cellulose dissolution follow similar trends. The IL anion is the driving force for separating polymer strands, where hydrogen bonds with lignin/cellulose replace intermolecular interactions within the polymer, yielding an entropically as well as enthalpically more favorable state.[4,7,17–23] The cation has been postulated to have a greater effect in lignin than in cellulose dissolution owing to π-π interactions.[18,19,21,22,24] The strength of IL-lignin interactions was estimated by Janesko and Zhang et al in computational studies using density function theory (DFT) calculations.[18,21] While DFT helped establish key mechanistic considerations in lignin dissolution, other principle- (as well as data-driven) approaches have been developed to either further our understanding of the deconstruction process or to help predict outcomes for new systems. These techniques, adapted for modeling IL

[a] *Department of Chemistry, George Washington University, 800 22nd St, Ste 4000, Washington, DC, 20052-0066.*
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

thermodynamics, span perturbed-chain statistical associating fluid theory (PC-SAFT),[25] group contribution methods,[26] quantitative structure–property relationships (QSPRs),[27] MC and MD simulations,[28, 29] COnductor-like Screening MOdel for Real Solvents (COSMO-RS),[30] as well as modern statistical tools based on machine learning (ML).[31] However, none provide a broadly predictive framework of both IL performance and safety that can aid in the *rational* design of novel and selective ILs.[16,32]

In considering safety, while many ILs are environmentally benign, some have been shown to be toxic, persistent and bioaccumulative. The sheer number of cation-anion combinations and the diversity of their chemical structures make broad safety statements about this chemical class impossible and (economically and ethically costly) animal testing impractical. Just as mechanistic understanding is key in design strategies for biomass dissolution, it is essential in designing environmentally benign ILs. For this chemical class, narcosis, i.e., interactions (usually of the cation) with cellular membranes, is the most common MOA. To that end, past studies showed that reducing hydrophobicity of alkyl chains was key to lowering IL toxicity.[33-37] In exploring this structure-activity relationship, Coleman and Gathergood proposed that adding polar functional groups, especially protic moieties, to alkyl side chains significantly decreases the likelihood of IL toxicity.[34] Despite the cation's primary role in narcosis, Cho et al and Santos et al showed that some anions, such as fluorinated ones, drive toxicity in imidazolium and cholinium-based ILs, which are generally regarded as safe.[35,37] Critically, a cation, which is known to be nontoxic in one compound, combined with an anion that is nontoxic in another compound, can still yield a toxic IL. Thus, it is important to analyze ILs as pairs, rather than individual ions, as shown in our previous work.[16]

Here we extend our previous efforts in designing ILs for cellulose dissolution to lignin in an attempt to propose an integrated computational framework for safe (and selective) biomass processing. Our approach considers both efficacy and safety of ILs, the latter via application of previously-validated guidelines for chemicals with minimal ecotoxicity.[32] Perturbations to the health of aquatic ecosystems, such as microorganisms, invertebrates, plants and fish, are often the first marker of chemical's broader risk to humans and the environment.[33–37] Additionally, acute aquatic toxicity encompasses a wide range of MOAs and can inform chronic effects using adjustment factors.[38–42]

Considering performance, our approach for lignin dissolution is based on Kraft lignin, which represents ca. 85% of the total lignin produced. Kraft lignin is one of the most recalcitrant lignin types due to acid treatment in the pulp paper manufacture. The abundance of Kraft lignin also alleviates some of the underlying structural-variability concerns. The aim was to create practical and quantitative guidelines that would inform design of new ILs at the 'drawing-board' stage of new product development, i.e., when screening structures of potential candidates. Such approach is common in the drug discovery, for example, where initial *in-silico* screening reduces the cost and complexity of downstream analysis, which may include more comprehensive experimental testing. Thus, it is important to view the present analysis as part of a larger protocol versus a standalone tool for IL development.

Methods

Dataset

Data used to develop our model was compiled from four experimental studies, which measured solubility in a total of 37 structurally diverse ILs (Figure 1).[7,43–45] These studies were selected based in (i) measurements of Kraft lignin, (ii) well-documented protocols of the solubility measurements, (iii) reliability of the experimental protocol based on recognized best practices for solubility measurements, and (iv) cation-anion diversity. Despite best efforts, we note the well-documented challenges with data variability, reproducibility and uncertainty related to biomass solubility measurements.[39] A full list of ILs with corresponding experimental and predicted metrics is provided in Table S1. The largest training subset by Liu et al[43] consisted of 18 unique IL pairs based on cholinium cations and amino-acid anions (Table S1), which are particularly intriguing owing to their proposed low ecotoxicity[16] and abundance of experimental solubility data for lignin, hemicellulose, and cellulose. The latter was leveraged here to gauge selectivity of ILs in lignin vs. cellulose dissolution, based on information mined from additional reports.[47-50] For modeling purposes, solubility units were converted to g/kg (used by majority of existing studies).

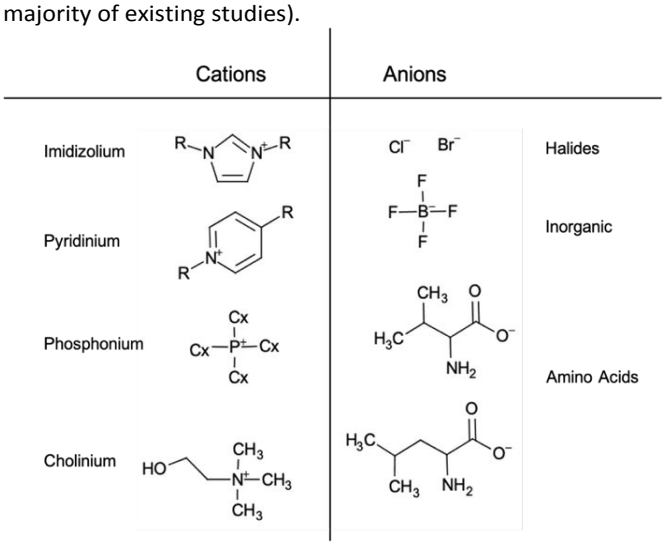

Figure 1. Basic structures of representative IL cations and anions used in this study. The complete list of ILs is available in Table S1.

Model description

In this study, we expanded on the tiered methodology developed and reported to facilitate IL design for cellulose dissolution.[16] In our previous work, the first tier focused on hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, used in conjunction with Monte Carlo simulations to assess energetics of noncovalent interactions between cellulose (treated as the solvent) and IL pairs (treated as the solute to take advantage of QM description). For lignin, our

approach was analogous in the partitioning of the system into solute and solvent, with subsequent evaluation of interface energetics. However, it reflected additional complexity of the polyphenol structure of lignin and the resulting diversity of intermolecular interactions.

A density-functional model developed by Janesko was the first to characterize both IL-cellulose and IL-lignin interactions *in silico*.[21] Their study identified π-stacking and hydrogen bonding interactions with imidazolium-based ILs as important enthalpic drivers of lignin dissolution. More recently, Zhang et al. drew similar conclusions, albeit from a larger model based around the β-O-4 linkage, commonly found in lignin.[18,21] In developing our models, we drew on these computational studies as well as reports by Crestini et al. and Lancefield et al.,[51,52] which examined the molecular structure and formation of Kraft lignin, showing high abundance of β-O-4, stilbenes and secoisolariciresinols interunit bonds. From these three substructures, the β-O-4 linkages are the most prevalent (ca. 45-85% by total mass). Given that observed interactions between lignin and IL are affected by different methoxy and hydroxy substituents on lignin aromatic rings, 4 different models were explored here (Figure 2). Our structures incorporate the β-O-4 linkage, used in combinatorial base-alcohol adaptations of the Zhang computational model, or the guaiacyl glycerol-β-guaiacyl ether (GG). [4,18,21] Due to lack of functionalization, stilbenes in Kraft lignin are not expected to play a significant role in dissolution. As for secoisolariciresinols, moieties key in π-stacking and hydrogen bonding are already well-represented by the GG-derived models in Figure 2. It should be noted, however, that byproducts such as acylglycerol (AG) or enol ether (EE) have been observed in lignin and were not explicitly considered in this study to prevent model overfitting.[51, 52]

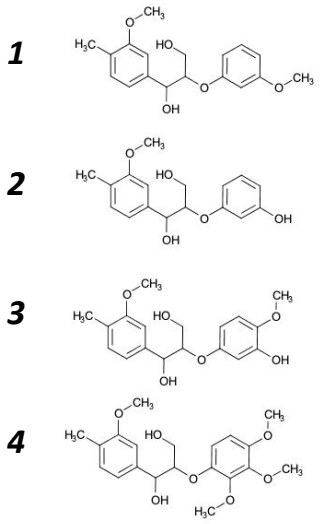

Figure 2. Four computational models used to represent differential functionalization on guaiacyl glycerol-β-guaiacyl ether (GG) in Kraft lignin.

Technical approach

It is reasonable to propose that both types of interactions, π-stacking and hydrogen bonding, observed by Zhang et al and Janesko,[18,21] are key to lignin dissolution, and are well-captured by our QM/MM/MC approach. Analogous simulations have been successfully used to characterize cellulose solubility in ILs,[16] as well as quantify often complex molecular interactions in other applications, such as computational drug discovery.[53] Thus, exploring the subtle variations in the phenol substitutions statistically to fit predicted dissolution metrics was considered an effective way to optimize our computational model's performance. This strategy of performance being 'tuned' not by changing descriptors (i.e., how we model IL-biomass interactions) but by optimizing the composition of the biomass components was not previously explored, and rests on the robustness and explicitness of the theoretical approach. To that end, our method, aside of its predictive value, can be potentially valuable in informing the relative functionalization of biomass' building blocks in one source vs. another.

Analogous to our previous study on cellulose dissolution, prior to QM/MM/MC simulations, the geometry of each IL pair was optimized using PM7-SMD based on general IL dielectric.[54–56] During subsequent simulations, translational and rotational degrees of freedom of individual cations and anions were sampled, constrained at a fixed distance apart (reflecting ground-state geometry from PM7-SMD energy minimizations) to prevent dissociation, which is known to occur in highly dilute solutions.[57] Each of the 4 models outlined in Figure 2 was described using the all-atom OPLS force field. The CM3 charges, scaled by a factor of 1.14 in Coulombic terms, were used to assess lignin-IL interactions.[58,59] Simulations were based on an NVT-NPT ensemble at 1 atm and temperature equal to the reported experimental temperature (Table S1), and used 10-Å cutoffs for nonbonded interactions. Solute-solvent systems consisted of the IL treated as a solute in a periodic solvent box of 267 lignin molecules (ca. 35 x 35 x 35 Å in size) built with the BOSS 4.9 software.[59] For each IL-lignin model, a short NVT simulation of 5×10^5 sample configurations was used to quickly equilibrate the lignin 'solvent', followed by an NPT simulation of 2×10^6 configurations for full-system equilibration and 5×10^6 configurations of averaging to compute relevant properties and system energetics.

Ecotoxicity model

To assess aquatic toxicity of ILs, we relied on a previously developed and validated approach,[16,32] which captures both acute and chronic effects across all relevant MOAs (modes of action) in standard test species (i.e., fish, crustaceans, and algae). Briefly, energies of frontier molecular orbitals (FMOs) were used to derive the band gap (a measure of general acid-base reactivity), which was applied along with the octanol-water partition coefficient, log $P_{o/w}$ (a measure of bioavailability) to gauge whether the current ILs were associated with low probability of hazard to aquatic species, according to a protocol reported in Griffin et al.[16] Log $P_{o/w}$ was estimated in linear response calculations from aqueous QM/MM/MC according to a previously published protocol.[16,59]

Statistical model

The Python programming language was used to build multivariate linear models (MLR). In these models, computed interaction energetics between lignin components and ILs based on the Coulomb and Lennard-Jones (LJ) potentials were fitted to the reported solubility metrics. Models were generated using an ordinary least squares regression, fitting a linear model with coefficients optimized to minimize the residual sum of squares between target values in the data set (i.e., solubility) and predicted target values (solubility informed from Coulomb/LJ energies). Individual MLRs were then combined into a single model for lignin solubility in ILs, and coefficients were re-optimized to maximize fit across all four datasets. The logic behind harmonizing the sensitivity of the response was to minimize training-set bias, given that the effect of interaction energetics on solubility should be comparable across studies. This is especially pertinent for small datasets (i.e., Glas et al and Rashid et al),[7,44] and because the composite MLRs cannot be validated. To that end, the current models are predictive of lignin solubility within the confines of the present data sources, establishing mechanistically defensible quantitative relationships, which can be easily augmented and replicated across other studies. Thus, to offer a practical answer the question "What is the predicted solubility of lignin in a new IL using this approach?", one can either select the most appropriate model from the current datasets (i.e., based on IL type and lignin source); compute the mean using all four datasets; or develop their own study-specific MLR following our protocol that 'plugs into' the composite model here.

Results and Discussion

Lignin dissolution model

Prior to developing MLRs, we investigated univariate correlations between computed IL-lignin energetics (for both Coulomb and LJ interactions) and observed solubility. We noted a broad distribution of R^2 values across our 4 different models and experimental studies (Figure 3, model correlations per study in Figure S1).

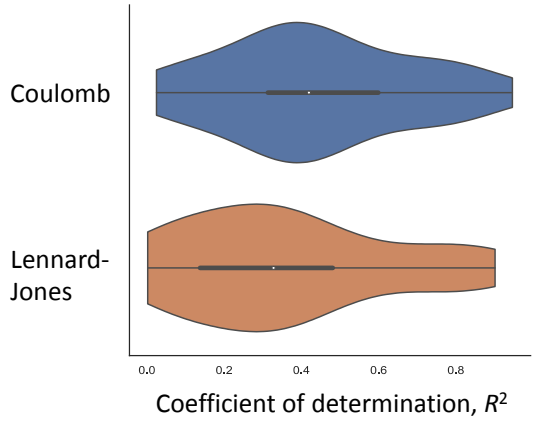

Figure 3. Distribution of R^2 values for univariate correlations between predicted energetics and solubility, across all 4 models and experimental datasets (viz. Methods). Average R^2 value is slightly higher for models correlating Coulomb interactions (0.44) than Lennard-Jones, LJ (0.36). The white dot inside each violin plot represents the median R^2 value, and the grey bar represents the interquartile range (i.e., middle 50%).

From Figure 3, Coulomb energetics correlated marginally better with solubility than LJ interactions, with a higher median and interquartile range of R^2 values. This finding is consistent with Coulomb forces being reflective of the largely electrostatic π-stacking and hydrogen bonding interactions in lignin dissolution. While we noted some differences in the experimental protocol across the 4 studies, it is more practical to posit that lignin source and different ILs used in each study are reflective of the distributions observed in Figure 3. To that end, it was necessary to develop MLR models that would capture variability of lignin composition, while accounting for different types of interactions dominating different solvent systems, i.e., electrostatic as well as van der Waals interactions, the latter being important for ILs with larger hydrophobic regions.

The resultant MLR is a composite model that captures variability in lignin across experiments, utilizing computed Coulomb and LJ energetics of the four computational structures (Figure 2, breakdown by study provided in Figure S2). As noted in Methods, representative model structures are weighted within each experimental subset to optimize the fit between computed energetics of IL-lignin interactions and observed solubility. This approach allowed us to develop separate MLR for each study (i.e., lignin) type (Figure S2), which were then combined into a composite linear model for either Coulomb (Figure 4, top) or LJ interactions (Figure 4, bottom).

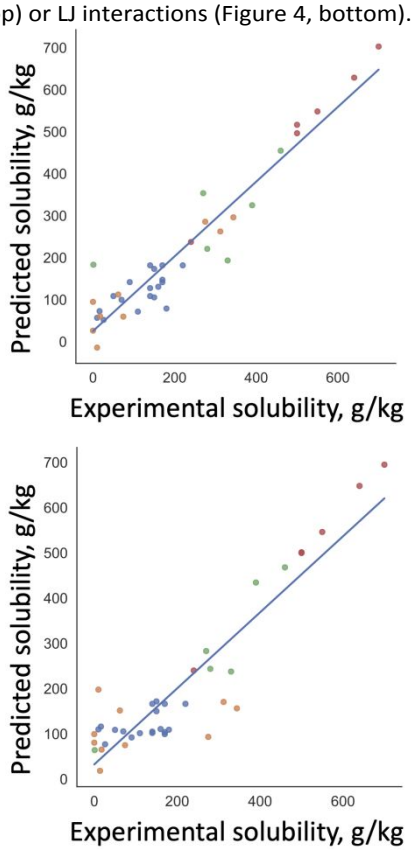

Figure 4. Linear correlations between the predicted solubility values from the **Coulomb model (TOP)** y = 0.91x + 19.2711, R^2 = 0.91, *SE* = 58.19, *p*-value = 4.88×10^{-21}, *RMSE* = 56.7; Liu et al (Sol = 0.95Ca – 0.41Cb + 0.66Cc – 0.16Cd), Pu et al (Sol = 2.40 Ca + 3.4Cb – 5.93Cc + 0.48Cd), Glas et al (Sol = 0.97Ca – 7.07Cb + 0.51Cc + 12.73Cd), Rashid et al (Sol= -3.13Ca + 3.63Cb +3.31Cc -12.87Cd). **The LJ model (BOTTOM)** y = 0.84x + 32.55, R^2 = 0.84, SE = 27.5, *p*-value = 1.068×10^{-16}, *RMSE* = 54.68; Liu et al (Sol = 0Ca + 0.37Cb + 2.37Cc – 0.58Cd),

Pu et al (Sol = -6.46Ca – 19.68Cb +10.18Cc + 9.53Cd), Glas et al (Sol = -5.31Ca + 51.0Cb +6.36Cc – 49.59Cd), Rashid et al (Sol = 14.37Ca – 10.61Cb – 2.07Cc – 0.03Cd). Ca = Lignin Model 1, Cb = Lignin Model 2, Cc = Lignin Model 3, Cd = Lignin Model 4 (Figure 2, top to bottom); Liu et al = blue dots, Pu et al = orange dots, Glas et al = green dots, Rashid et al = red dots.

From Figure 4, both composite MLRs are reasonably well correlated: Coulomb (R^2 = 0.91) and LJ (R^2 = 0.84). A *K*-fold internal validation showed consistency between predictive power and robustness of the initial fit, with average q^2 of 0.88 (Coulomb) and 0.83 (LJ) for K = 40 and 0.87 (Coulomb) and 0.80 (LJ) for K = 10. To further ensure the model is not derived due to chance, a 10-fold Y-randomization was performed, yielding an average correlation coefficient of 0.283, and the highest of 0.475. The applicability domain (AD) of the two compositive MLRs was evaluated using the leverage approach,[60] constructing a Williams plot of standardized residuals vs. leverage values (Figures S3). From Figure S3, all compounds in the dataset are within the AD of the Coulomb and LJ MLRs, without undue structural influence on the models. The highest leverage values observed were 0.24 and 0.27 for the Coulomb and LJ models, respectively, which are below the 0.35 threshold value for the two models.[60] It is notable that the compounds with highest leverage also have low standard residual values, and so are important for the model fit but are not outliers.

From Figure 4, the high RMSE (above 50 g/kg for both Coulomb and LJ models) can be interpreted through the lens of fitting high-resolution energy information onto low-resolution solubility data, which is exacerbated by the composite MLR optimization. Additionally, computational-model limitations may be a contributing factor to the large residuals. Aside of omitting acylglycerol (AG) or enol ether (EE) byproducts, the current model does not account for differences in IL viscosity as ILs are modelled quantum mechanically as individual solutes in the biomass 'solvent'. The LJ (van der Waals) MLR is particularly susceptible to this, as can be observed from the relatively poor correlations with dissolution metrics in the dataset by Pu et al (Figure 4, orange dots), which features large temperature variations. For example, lignin solubility in both [Hmim][CF_3SO_3] and [Mmim][$MeSO_4$] is erroneously predicted to decrease with increasing temperature in the LJ MLR (Table S1). In contrast, the Coulomb MLR is less affected due to the relatively greater strength of these interactions for most ILs. The greater general robustness of this model is also supported by our understanding of role that short-range electrostatics play in the lignin dissolution process. Lastly, the Coulomb MLR is sensible from a modelling standpoint, as LJ parameters need to be developed for new solvents, whereas atomic charges, which drive Coulomb interactions, can be computed on the fly using QM. ILs with larger hydrophobic regions, where van der Waals interactions play an important role, may be an exception, as supported by the slightly better fitted LJ model for the Glas et al dataset, which contains some of the largest and most polarizable ILs in the present study (green dots in Figure 4).

Integrating cellulose and lignin models

Our previously developed models for cellulose dissolution in ILs[16] and our current MLRs for lignin indicate that whereas van der Waals interactions appear to better differentiate the former, the latter is more robustly predicted by Coulomb energetics. To that end, a partition coefficient can be expressed, log $P_{L/C}$ (Equation 1), which denotes the relative solubility of lignin vs cellulose in any particular IL at a given temperature:

$$log\ P_{L/C} = log\left(\frac{[Lignin]}{[Cellulose]}\right)_{IL-T} \quad (1)$$

Log $P_{L/C}$ allows us to identify ILs that are selective for either biomass component, which is valuable for the design of new solvent systems. Here, predictive models are critical in filling experimental data gaps, as measured values in the literature are typically available for either lignin or cellulose but not always for both.[17,46] Furthermore, due to the variability and low resolution of these measurements *in chemico*,[46] one might prefer to adopt an *in-silico* model validated for the specific lignin source vs. relying on external studies. Table S1 shows how calculated solubility values can be used to supplement literature values to derive log $P_{L/C}$ metrics.

From Table S1, general trends in log $P_{L/C}$ confirm existing knowledge: imidazolium based ILs paired with an anion capable of hydrogen bonding either favor cellulose or are near-equivocal in their partition coefficients (viz. the dataset by Pu et al). Conversely, high log $P_{L/C}$ values were observed for cholinium and pyridinium ILs, where the former can favorably interact with alcohols in lignin (viz. guaiacyl glycerol-β-guaiacyl ether models in Figure 2), and the latter can form π–π interactions with lignin's aromatic base alcohols. The lowest preference for lignin among cholinium ILs was noted for glutamate and aspartate, which are both small and capable of strong hydrogen bonding. Interestingly, the two highest log $P_{L/}$ values correspond to the [BMPyr] (i.e., 1-butyl-1-methylpyrrolidinium) cation, which is likely too large to effectively dissolve cellulose.

We should note that the current exploration does not comprehensively capture log $P_{L/C}$'s dependence on temperature, though a general trend of increasing relative preference for lignin over cellulose with increasing temperature can be observed (Table S1). For example, [Bmim][$MeSO_4$] appears to switch preference in dissolving cellulose at 25 °C to favoring lignin at 50 °C, though the difference is small. Similarly, [Hmim][CF_3SO_3] prefers cellulose at 50 °C and is near equivocal at 70 °C. Both [Mmim][$MeSO_4$] and [Py][Ac] favor lignin but the relative preference for lignin over cellulose dissolution increases with higher temperate. It is reasonable to postulate that the more disordered state of lignin facilitates the phase change at higher temperatures, increasing the thermodynamic driving force regardless of the IL.

Design guidelines for (selective) biomass dissolution

Our previous work showed a strong correlation between LJ interactions and cellulose dissolution, as well strong correlations between certain physiochemical properties (polarizability and aqueous solubility) of ILs and computed IL-cellulose LJ energetics, which allowed us to link design-ready

properties with solubility. It is well known that hydrogen bonding and the size of the IL play an important role in both lignin and cellulose dissolution mechanisms,[61–67] where notably large cations (e.g., tBMA, BA, DDA, TBA, etc.) are unable to dissolve cellulose, regardless of the anion present. However, size as a measure of polarizability can also increase solubility.[16,67] What is assumed in most theoretical studies is that the composition and structure remain constant, which is not the case for lignin. As such, physiochemical properties observed in one study might be less valuable depending on the difference in lignin and/or ILs used. To that end, we observed no meaningful correlations between physicochemical properties and log $P_{L/C}$ (and none that would correlate to lignin dissolution metrics) here. While the apparent lack of (atomistic) structure-property relationships upsets facile design of new solvents through *rational* modifications of existing systems,[68] it highlights the complexity of the lignin-dissolution process and the need for molecular interaction-based models, such as the source-specific MLRs developed here. Furthermore, we envision that future studies, relying on the current linear models, can target the development of useful QSPRs to aid in rational design, by quantifying the substituent effect (around common IL cores) on computed Coulomb and LJ energetics and, by extension, on the solubility of lignin in ILs.

To incorporate a criterion of safety into our models for biomass dissolution, we applied the previously developed cutoffs for minimal ecotoxicity, which posit that ILs with log $P_{o/w}$ < 3 and the band gap, $\Delta E_{\text{HOMO-LUMO}}$ > 6 are likely safe toward aquatic species (Table S1).[16] The log $P_{o/w}$ criterion, a proxy for general bioavailability, is especially useful here, since ILs with low lipophilicity are unlikely to passively diffuse through membranes, and so their acid-base reactivity ($\Delta E_{\text{HOMO-LUMO}}$) is of lesser concern. From Table S1, consistent with toxicological studies, cholinium/amino-acid based ILs fall into this category. Conversely, aromatic (i.e., imidazolium and pyridinium) ILs showed greater likelihood of ecotoxicity based on either lower HOMO-LUMO gaps or higher log $P_{o/w}$. Substituents on the aromatic rings do matter, and the larger the alkyl groups, the greater the likelihood of toxicity due to higher log $P_{o/w}$ (e.g., [Mmim] vs. [Bmim] methylsulfate in Table S1). While the present cutoffs were derived and validated for a binary outcome (i.e., safer chemical space vs. 'the rest'), our previous work using the US EPA's categories of concern indicates that increase in hazard probability with respect to log $P_{o/w}$/$\Delta E_{\text{HOMO-LUMO}}$ is incremental, and so relative changes in either metric are informative in designing safer chemicals.[32] For aforementioned reasons, this is particularly true for log $P_{o/w}$, where compounds with low values (i.e., readily water soluble) and low $\Delta E_{\text{HOMO-LUMO}}$ (i.e., reactive) are of lesser concern than those with both higher $\Delta E_{\text{HOMO-LUMO}}$ and log $P_{o/w}$, which may be less reactive but possibly activated to electrophilic toxicants via metabolic transformations upon absorption.[32]

Conclusions

The design of high-performing chemicals with minimal toxicity remains a formidable challenge, yet it is a vital component of systems-based solutions in green chemistry. Point in case, while deconstruction of biomass using ILs is a green chemistry 'success story', showcasing the use of non-volatile solvents and renewable materials as chemical feedstocks, progress is hindered unless we understand mechanistic drivers of toxicity and performance that can inform design of tomorrow's systems. Here and in our previous work,[16] we showed that by approaching biomass dissolution computationally, via both explicit modeling of molecular interactions and physicochemical property calculations, we can elucidate these drivers, and develop useful design tools. One caveat worth repeating is that no *in silico* model should be held to a higher standard than the quality of its underlying data, and in the case of biomass dissolution, available data is of variable quality with low resolution and dubious reproducibility. To that end, principle-driven models relying on modeling of mechanistically-relevant molecular interactions may have the upper hand over data-driven models, which require large (i.e., less curated) datasets and thus are more likely to optimize performance by including artifacts of data variability and uncertainty in model training.

Author Contributions

Preston Griffin: Data curation, Calculations, Statistical methods, Data analysis, Writing- Original draft preparation **Jakub Kostal**: Conceptualization, Methodology, Statistical modelling, Supervision, Writing- Reviewing and Editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Gratitude is expressed to Prof. William L. Jorgensen (Yale University) for providing the BOSS 4.9 program; to Prof. Orlando Acevedo (University of Miami) for sharing his OPLS force-field parameters for ionic liquids; and to Prof. Adelina Voutchkova-Kostal (GW) for her guidance with statistical models. This work was supported by the National Science Foundation (NSF1805080).

References

1 Vaclav Smil. *Energy Transitions: Global and National Perspectives*, 2nd ed.; Energy Transitions: Global and National Perspectives. & BP Statistical Review of World Energy, 2017.
2 US Energy Information Administration. *Fossil fuels have made up at least 80% of U.S. fuel mix since 1900*, 2015, https://www.eia.gov/todayinenergy/detail.php?id=21912 (accessed 2023-5-23).
3 Lourenço, A.; Pereira, H. Compositional Variability of Lignin in Biomass. In Lignin - Trends and Applications, *IntechOpen*, 2018.

4 Melro, E.; Alves, L.; Antunes, F. E.; Medronho, B. A Brief Overview on Lignin Dissolution. *J Mol Liq*, 2018, 265, 578-584.
5 Sun, J.; Dutta, T.; Parthasarathi, R.; Kim, K. H.; Tolic, N.; Chu, R. K.; Isern, N. G.; Cort, J. R.; Simmons, B. A.; Singh, S. Rapid Room Temperature Solubilization and Depolymerization of Polymeric Lignin at High Loadings. *Green Chem*, 2016, 18 (22), 6012-6020.
6 Sameni, J.; Krigstin, S.; Sain, M. Solubility of lignin and acetylated lignin in organic solvents, 2017, *BioRes.* 12(1), 1548-1565.
7 Rashid, T.; Kait, C. F.; Regupathi, I.; Murugesan, T. Dissolution of Kraft Lignin Using Protic Ion-ic Liquids and Characterization. Ind Crops Prod, 2016, 84, 284-293.
8 Achinivu, E. C.; Howard, R. M.; Li, G.; Gracz, H.; Henderson, W. A. Lignin Extraction from Biomass with Protic Ionic Liquids. *Green Chem*, 2014, 16 (3), 1114-1119.
9 Zakzeski, J.; Bruijnincx, P. C. A.; Jongerius, A. L.; Weckhuysen, B. M. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem Rev*, 2010, 110 (6).
10 Holladay, J.; White, J.; Bozell, J.; Johnson, D. Top Value-Added Chemicals from Biomass Volume II—Results of Screening for Potential Candidates from Biorefinery Lignin, 2007, https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-16983.pdf (accessed 2023-5-23).
11 Boerjan, W.; Ralph, J.; Baucher, M. Lignin Biosynthesis. *Annu Rev Plant Biol*, 2003, 54 (1), 519-546.
12 Hanson, S. K.; Baker, R. T. Knocking on Wood: Base Metal Complexes as Catalysts for Selective Oxidation of Lignin Models and Extracts. *Acc Chem Res*, 2015, 48 (7), 2037–2048.
13 Wen, J.-L.; Yuan, T.-Q.; Sun, S.-L.; Xu, F.; Sun, R.-C. Understanding the Chemical Transformations of Lignin during Ionic Liquid Pretreatment. *Green Chem*, 2014, 16 (1), 181-190.
14 Sathitsuksanoh, N.; Holtman, K. M.; Yelle, D. J.; Morgan, T.; Stavila, V.; Pelton, J.; Blanch, H.; Simmons, B. A.; George, A. Lignin Fate and Characterization during Ionic Liquid Biomass Pre-treatment for Renewable Chemicals and Fuels Production. *Green Chem*, 2014, 16 (3), 1236-1247.
15 Costa, G.; Plazanet, I. Plant Cell Wall, a Challenge for Its Characterisation. *Adv Biol Chem,* 2016, 6 (3), 70-105.
16 Griffin, P.; Ramer, S.; Winfough, M.; Kostal, J. Practical Guide to Designing Safer Ionic Liquids for Cellulose Dissolution Using a Tiered Computational Framework. *Green Chem,* 2020, 22 (11), 3626-3637.
17 Rashid, T.; Kait, C. F.; Murugesan, T. Effect of Temperature on Molecular Weight Distribution of Pyridinium Acetate Treated Kraft Lignin. *Procedia Eng,* 2016, 148, 1363-1368.
18 Zhang, Y.; He, H.; Dong, K.; Fan, M.; Zhang, S. A DFT Study on Lignin Dissolution in Imidazolium-Based Ionic Liquids. *RSC Adv*, 2017, 7 (21), 12670-12681.
19 Ji, H.; Lv, P. Mechanistic Insights into the Lignin Dissolution Behaviours of a Recyclable Acid Hydrotrope, Deep Eutectic Solvent (DES), and Ionic Liquid (IL). *Green Chem,* 2020, 22 (4), 1378-1387.
20 Casas, A.; Palomar, J.; Alonso, M. V.; Oliet, M.; Omar, S.; Rodriguez, F. Comparison of Lignin and Cellulose Solubilities in Ionic Liquids by COSMO-RS Analysis and Experimental Validation. *Ind Crops Prod,* 2012, 37 (1), 155-163.
21 Janesko, B. G. Modeling Interactions between Lignocellulose and Ionic Liquids Using DFT-D. *Phys Chem Chem Phys,* 2011, 13 (23), 11393-11401.
22 Zavrel, M.; Bross, D.; Funke, M.; Büchs, J.; Spiess, A. C. High-Throughput Screening for Ionic Liquids Dissolving (Ligno-)Cellulose. *Bioresour Technol,* 2009, 100 (9), 2580-2587.
23 Fegyverneki, D.; Orha, L.; Láng, G.; Horváth, I. T. Gamma-Valerolactone-Based Solvents. *Tetrahedron*, 2010, 66 (5), 1078-1081.
24 Ji, W.; Ding, Z.; Liu, J.; Song, Q.; Xia, X.; Gao, H.; Wang, H.; Gu, W. Mechanism of Lignin Dissolution and Regeneration in Ionic Liquid. *Energy Fuels,* 2012, 26 (10), 6393–6403.
25 Paduszynski, K.; Domanska, U. Thermodynamic modeling of ionic liquid systems: Development and detailed overview of novel methodology based on the PC-SAFT. *J. Phys. Chem. B,* 2012, 116, 5002–5018.
26 Valderrama, J.; Robles, P. Critical properties, normal boiling temperatures, and acentric factors of fifty ionic liquids. *Ind. Eng. Chem. Res.,* 2007, 46, 1338–1344.
27 Eike, D. M.; Brennecke, J. F.; Maginn, E. J. Predicting infinite-dilution activity coefficients of organic solutes in ionic liquids. *Ind. Eng. Chem. Res.,* 2004, 43, 1039–1048.
28 Shah, J. K.; Maginn, E. J. Monte Carlo simulations of gas solubility in the ionic liquid 1-n-butyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. B*, 2005, 109, 10395–10405.
29 Shi, W.; Maginn, E. J. Molecular simulation and regular solution theory modeling of pure and mixed gas absorption in the ionic liquid 1-n-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl) amide ([hmim][Tf2N]). *J. Phys. Chem. B,* 2008, 112, 16710–16720.
30 Yu, K.; Ding, W.-L.; Lu, Y.; Wang, Y.; Liu, Y.; Liu, G.; Huo, F.; He, H. Ionic liquids screening for lignin dissolution: COSMO-RS simulations and experimental characterization, *J. Mol. Liq.*, 2022, 348, 118007.
31 Ding, W.-L. et al. Machine learning screening of efficient ionic liquids for targeted cleavage of the β–O–4 bond of lignin. *J. Phy. Chem. B*, 2022, 126, 3693–3704.
32 Kostal, J.; Voutchkova-Kostal, A.; Anastas, P. T.; Zimmerman, J. B. Identifying and Designing Chemicals with Minimal Acute Aquatic Toxicity. *PNAS,* 2015, 112 (20), 6289-6294.
33 Brooks, B. W.; Sabo-Attwood, T.; Choi, K.; Kim, S.; Kostal, J.; LaLone, C. A.; Langan, L. M.; Margiotta-Casaluci, L.; You, J.; Zhang, X. Toxicology Advances for 21st Century Chemical Pollution. *One Earth,* 2020, 2 (4), 312–316.
34 Coleman, D.; Gathergood, N. Biodegradation Studies of Ionic Liquids. *Chem Soc Rev,* 2010, 39 (2), 600-637.
35 Kebaili, H.; Pérez de los Ríos, A.; Salar-García, M. J.; Ortiz-Martínez, V. M.; Kameche, M.; Her-nández-Fernández, J.; Hernández-Fernández, F. J. Evaluating the Toxicity of Ionic Liquids on Shewanella Sp. for Designing Sustainable Bioprocesses. *Front Mater*, 2020, 7, 578411.
36 Cho, C.-W.; Phuong Thuy Pham, T.; Jeon, Y.-C.; Yun, Y.-S. Influence of Anions on the Toxic Ef-fects of Ionic Liquids to a Phytoplankton Selenastrum Capricornutum. *Green Chem*, 2008, 10 (1), 67–72.
37 Santos, J. I.; Gonçalves, A. M. M.; Pereira, J. L.; Figueiredo, B. F. H. T.; e Silva, F. A.; Coutinho, J. A. P.; Ventura, S. P. M.; Gonçalves, F. Environmental Safety of Cholinium-Based Ionic Liq-uids: Assessing Structure–Ecotoxicity Relationships. *Green Chem,* 2015, 17 (9), 4657–4668.
38 Ford, A. T.; Ågerstrand, M.; Brooks, B. W.; Allen, J.; Bertram, M. G.; Brodin, T.; Dang, Z.; Du-quesne, S.; Sahm, R.; Hoffmann, F.; Hollert, H.; Jacob, S.; Klüver, N.; Lazorchak, J. M.; Ledes-ma, M.; Melvin, S. D.; Mohr, S.; Padilla, S.; Pyle, G. G.; Scholz, S.; Saaristo, M.; Smit, E.; Steevens, J. A.; van den Berg, S.; Kloas, W.; Wong, B. B. M.; Ziegler, M.; Maack, G. The Role of Behavioral Ecotoxicology in Environmental Protection. *Environ Sci Technol,* 2021, 55 (9), 5620–5628.
39 May, M.; Drost, W.; Germer, S. et al. Evaluation of acute-to-chronic ratios of fish and Daphnia to predict acceptable no-effect levels. *Environ Sci Eur*, 2016, 28 (1), 1-9.
40 Voutchkova-Kostal, A. M.; Kostal, J.; Connors, K. A.; Brooks, B. W.; Anastas, P. T.; Zimmerman, J. B. Towards Rational Molecular Design for Reduced Chronic Aquatic Toxicity. *Green Chem,* 2012, 14 (4), 1001-1008.

41 Ankley, G. T.; Villeneuve, D. L. The Fathead Minnow in Aquatic Toxicology: Past, Present and Future. *Aquatic Toxicology,* 2006, 78 (1), 91–102.
42 Steele, W. B.; Kristofco, L. A.; Corrales, J.; Saari, G. N.; Haddad, S. P.; Gallagher, E. P.; Ka-vanagh, T. J.; Kostal, J.; Zimmerman, J. B.; Voutchkova-Kostal, A.; Anastas, P.; Brooks, B. W. Comparative Behavioral Toxicology with Two Common Larval Fish Models: Exploring Relationships among Modes of Action and Locomotor Responses. *Sci Tot Environ*, 2018, 640–641, 1587–1600.
43 Liu, Q.-P.; Hou, X.-D.; Li, N.; Zong, M.-H. Ionic Liquids from Renewable Biomaterials: Synthesis, Characterization and Application in the Pretreatment of Biomass. *Green Chem*, 2012, 14 (2), 304-307.
44 Glas, D.; van Doorslaer, C.; Depuydt, D.; Liebner, F.; Rosenau, T.; Binnemans, K.; de Vos, D. E. Lignin Solubility in Non-Imidazolium Ionic Liquids. *J Chem Technol & Biotechnol*, 2015, 90 (10), 1821-1826.
45 Pu, Y.; Jiang, N.; Ragauskas, A. J. Ionic Liquid as a Green Solvent for Lignin. *J Wood Chem Technol*, 2007, 27 (1), 23-33.
46 Brandt, A; Gräsvik, J; Hallett, J. P.; Welton, T. Deconstruction of lignocellulosic biomass with ionic liquids. *Green Chem,* 2013, 15, 550-583.
47 Xu, A.; Wang, J.; Wang, H. Effects of Anionic Structure and Lithium Salts Addition on the Dissolution of Cellulose in 1-Butyl-3-Methylimidazolium-Based Ionic Liquid Solvent Systems. *Green Chem*, 2010, 12 (2), 268–275.
48 Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. Dissolution of Cellulose with Ionic Liquids. *J Am Chem Soc,* 2002, 124 (18), 4974–4975.
49 Liu, Y.-R.; Thomsen, K.; Nie, Y.; Zhang, S.-J.; Meyer, A. S. Predictive Screening of Ionic Liquids for Dissolving Cellulose and Experimental Verification. *Green Chem,* 2016, 18 (23), 6246–6254.
50 Isik, M.; Sardon, H.; Mecerreyes, D. Ionic Liquids and Cellulose: Dissolution, Chemical Modification and Preparation of New Cellulosic Materials. *Int J Mol Sci,* 2014, 15 (7), 11922–11940.
51 Crestini, C.; Lange, H.; Sette, M.; Argyropoulos, D. S. On the Structure of Softwood Kraft Lignin. *Green Chem,* 2017, 19 (17), 4104-4121.
52 Lancefield, C. S.; Wienk, H. L. J.; Boelens, R.; Weckhuysen, B. M.; Bruijnincx, P. C. A. Identification of a Diagnostic Structural Motif Reveals a New Reaction Intermediate and Condensation Pathway in Kraft Lignin Formation. *Chem Sci,* 2018, 9 (30), 6348-6360.
53 Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science*, 2004, 303 (5665), 1813–1818.
54 Marenich, A. v.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J Phys Chem B*, 2009, 113 (18), 6378–6396.
55 Bernales, V. S.; Marenich, A. v.; Contreras, R.; Cramer, C. J.; Truhlar, D. G. Quantum Mechanical Continuum Solvation Models for Ionic Liquids. *J Phys Chem B,* 2012, 116 (30), 9122–9129.
56 Gaussian 16, Revision B.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.
57 Jain, P.; Kumar, A. Concentration-Dependent Apparent Partition Coefficients of Ionic Liquids Possessing Ethyl- and Bi-Sulphate Anions. *Phys Chem Chem Phys*, 2016, 18 (2), 1105-1113.
58 Sambasivarao, S. v.; Acevedo, O. Development of OPLS-AA Force Field Parameters for 68 Unique Ionic Liquids. *J Chem Theory Comput,* 2009, 5 (4), 1038–1050.
59 Jorgensen, W. L.; Tirado-Rives, J. Molecular Modeling of Organic and Biomolecular Systems Using BOSS And MCPRO. *J Comput Chem*, 2005, 26 (16), 1689-1700.
60 Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.*, 2007, 26 (5), 694-701.
61 Casas, A.; Omar, S.; Palomar, J.; Oliet, M.; Alonso, M. V.; Rodriguez, F. Relation between Differential Solubility of Cellulose and Lignin in Ionic Liquids and Activity Coefficients. *RSC Adv,* 2013, 3 (10), 3453-3460.
62 Zahn, S.; Kirchner, B. Uncovering Molecular Secrets of Ionic Liquids, Chemical Modeling Applications and Theory, 2012, 9.
63 Benedetto, A.; Bodo, E.; Gontrani, L.; Ballone, P.; Caminiti, R. Amino Acid Anions in Organic Ionic Compounds. An Ab Initio Study of Selected Ion Pairs. *J Phys Chem B,* 2014, 118 (9), 2471–2486.
64 Gabl, S.; Schröder, C.; Steinhauser, O. Computational Studies of Ionic Liquids: Size Does Matter and Time Too. *J Chem Phys,* 2012, 137 (9), 094501.
65 Izgorodina, E. I.; Seeger, Z. L.; Scarborough, D. L. A.; Tan, S. Y. S. Quantum Chemical Methods for the Prediction of Energetic, Physical, and Spectroscopic Properties of Ionic Liquids. Chem Rev 2017, 117 (10), 6696–6754.
66 Hiroyuki Ohno; Kenta Fukumoto. Amino Acid Ionic Liquids. *Acc Chem Res,* 2007, 40 (11), 1122–1129.
67 Ossowicz, P.; Klebeko, J.; Roman, B.; Janus, E.; Rozwadowski, Z. The Relationship between the Structure and Properties of Amino Acid Ionic Liquids. *Molecules*, 2019, 24 (18), 3252.
68 Kostal, J.; Voutchkova-Kostal, A. Going All In: A Strategic Investment in In Silico Toxicology. *Chem Res Toxicol,* 2020, 33 (4), 880–888.

