Green Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

Check for updates

Cite this: *Green Chem.*, 2020, **22**, 3626

Practical guide to designing safer ionic liquids for cellulose dissolution using a tiered computational framework†

Preston Griffin, Selene Ramer, Matthew Winfough and Jakub Kostal *

Received 14th March 2020,
Accepted 13th May 2020

DOI: 10.1039/d0gc00923g

rsc.li/greenchem

The theoretical promise of ionic liquids (ILs) as 'green' designer solvents that can be tuned to facilitate key steps of lignocellulosic biomass processing has not been fully realized due to the sheer number of possible cation–anion combinations and concerns about toxicity of this class of chemicals. Although computational methods are being applied to identify ILs with specific functions, such as dissolution of cellulose, they are not used to iteratively design new ionic liquids with the goal of simultaneously optimizing multiple criteria, such as performance and environmental safety. Here we describe a tiered computational approach to develop new ILs based on mixed quantum and molecular mechanics simulations, which, combined with analysis of physicochemical properties of ILs can be used to guide structural modifications to design both better performing task-specific and safer IL analogs. The increase in computing requirements of the proposed approach over structure-based statistical models is relatively modest; yet, our approach is more robust than these models, and far less costly than highly-accurate but very demanding large-scale molecular simulations.

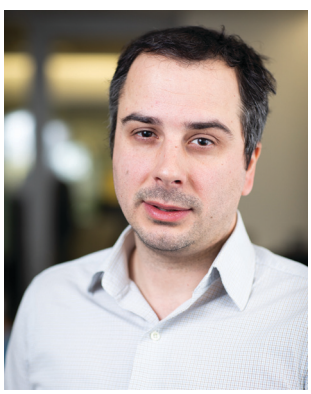
Introduction

Developing efficient and economically-viable processes for generating fuels and platform chemicals from non-food competitive lignocellulose is a critical challenge for sustainable development.[1] The biorefinery concept, which has been proposed as a strategy to vastly increase our ability to efficiently convert lignocellulose to fuels and chemicals, still lacks technology needed for economical and safe processing.[2,3] Specifically, the valorization of cellulose, the main component of lignocellulose (40–50 wt%), requires that the polymer is dissolved prior to deconstruction into constituting monomers.[4] The three-dimensional microfibril structure of cellulose is embedded in a hydrated matrix of covalently coupled hemicellulose and amorphous phenylpropanoid polymer lignin.[5] On the molecular level, cellulose comprises of D-glucose chains that align parallel into flat sheets, which are stacked together to form the microfibril. This structure is held together by a complex network of intra- and inter-chain hydrogen bonds (Hbonds) between hydroxyl groups and van der Waals interactions on contiguous sheets.[6] Thus, dissolving cellulose requires the disruption of these interactions, which typically calls for harsh reaction conditions and corrosive reagents.[7,8]

To address the demand for effective and green solvents in cellulose dissolution, ionic liquids (IL) have emerged as a viable alternative. These molten salts offer many advantages, such as high thermal stability, wide electrochemical window,

Department of Chemistry, George Washington University, 800 22nd St, Ste 4000, Washington, DC, 20052-0066, USA. E-mail: jkostal@gwu.edu

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d0gc00923g



Jakub Kostal

Jakub Kostal is an Assistant Professor of Chemistry at the George Washington University, where his group develops computational methods for the design of safer and efficacious chemicals. Dr Kostal is Founder and Principal of Designing Out Toxicity LLC, which provides consulting services for filling toxicological data gaps using in silico *predictive tools. He holds a PhD from Yale University (PhD '12) under the supervision of William Jorgensen in Theoretical and Biophysical Chemistry and a BA in Chemistry and Biochemistry from Middlebury College (BA '06). Dr Kostal completed postdoctoral work with Paul Anastas and Julie Zimmerman at Sustainability A to Z.*

This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

broad liquid range, low vapour pressure and high solvation ability toward various chemicals, which has led to their widespread use in many applications.[9] Most importantly, the physical and chemical properties of ILs can be tuned by varying the structures of cations and anions.[10] Since the seminal report by Rogers *et al.* on cellulose dissolution in 1-butyl-3-methylimidazolium chloride in 2002,[11] extensive body of work on cellulose dissolution by ILs has been developed, providing both a diverse repertoire of suitable structures and studies that elucidate the mechanistic underpinning of the dissolution process.[8] Nevertheless, the extent of cellulose solubility offered by ILs[12] and our mechanistic understanding of ILs' structure–activity relationships still have significant gaps.[13] Most importantly, there is need to develop reliable and efficient design tools that would yield next-generation of these solvent systems based on criteria of high performance and low environmental impact.[8] This is a task that cannot be accomplished experimentally or *ad hoc* as there are estimated 10^{12}–10^{18} cation–anion combinations based on (existing!) structures.[14]

Extensive efforts in the field of ILs over the past two decades have delivered a plethora of computational methods to study and predict cellulose dissolution, and in a few cases, to inform design of more efficient systems. Reported computer models range from structure-based (Q)SAR/(Q)SPR-type approaches[12,15] to quantum-mechanics models[16,17] to sophisticated and computationally-intensive molecular dynamics simulations.[18,19] A comprehensive overview of these models, which explores their utility in the context of IL structures and their conformational equilibria, the role of IL–cellulose interactions in cellulose solubility and co-solvent effects, is provided elsewhere.[8]

Among existing tools showing promise in solvent design, COSMO-based models are effective in predicting thermodynamics of separation processes.[17] While most require costly quantum mechanics calculations, Zhang *et al.* showed a modified group contribution approach paired with COSMO can speed up the discovery of novel ILs.[20] From studies of cellulose dissolution using COSMO, the work by Liu *et al.*[21] should be highlighted for its usefulness and simplicity. Relying on COSMO-RS (COSMO for Real Solvent), an implicit solvation model capable of performing mixture calculations at various temperatures and integrating dominant interactions of H-bonds and van der Waals forces in IL systems,[17] Liu *et al.* performed assessments of 357 ILs (combinations of 17 cations and 21 anions), and validated their approach using experimental measurements of microcrystalline cellulose solubility in 7 of these ILs.[21] Predicted logarithmic activity coefficients showed significant correlation with wt% solubility of cellulose in ILs, with $R^2 \sim 0.35$–0.62, depending on the specific cellulose model used.

While instrumental in advancing the field of cellulose dissolution by ILs, existing computational models suffer from trade-offs: sophisticated modelling approaches have a narrow mechanistic focus, and are too demanding to apply across the vast chemical space of ILs; statistically-heavy approaches trade robustness and mechanistic transparency for high training-set accuracy; and existing methods that attempt to compromise between efficiency and mechanistic detail are yet to show optimal performance metrics. The design approach reported by Liu *et al.* is an example of the latter; while the COSMO-RS model alleviates many limitations and theoretical shortcoming of dielectric continuum models,[22] it remains an implicit model that trades atomic detail for efficiency. In particular, implicit models neglect the specific spatial orientation between donor and acceptor in Hbonds,[23] which play a key role in cellulose interactions with ILs.[8] Thus, there is room for improvement in a design tool for novel ILs by considering explicit solvent systems. Importantly, no existing models combine function and environmental-safety criteria into a single design tool.

Rational design of functional chemicals with minimal toxicity remains a key frontier in green chemistry that is yet to be systematically applied. Prior work by our group has focused on developing design guidelines for chemicals with minimal ecotoxicity,[24] and has explored kinetic and thermodynamic design drivers for a host of human-toxicity endpoints.[25–28] Here, we examined the relevance of the aforementioned computational methods to further the design efforts of Liu *et al.* in developing novel ILs. The goal of our effort was to develop a practical design approach, one that balances accuracy, reliability and mechanistic relevance with computational efficiency, ease of use and interpretability for the non-expert, and can inform both function and environmental performance. We report a two-tier computational framework that links cost-effective yet highly-accurate molecular simulations of cellulose solubility in ILs with physicochemical properties of ILs. The framework combines optimization of performance, *i.e.* ILs' ability to dissolve cellulose, with previously-validated design guidelines for minimal (eco)toxicity[24,29–32] to provide the first step toward holistic approach in new IL development that is consistent with green chemistry principles. Our model achieves good accuracy at moderate computational cost by strategically modelling IL as a solute, rather than a solvent, in a cellulose-like medium. Such system allows us to describe the IL pair more accurately using quantum mechanics, and easily modify its structure to enable rational design. The broader goal of this work is to spur future IL design efforts that promote systems thinking based on the entire lifecycle of these chemicals.

Methods

Theoretical approach

The tiered computational framework to facilitate design of novel ILs is outlined in Fig. 1. The first tier focused on IL–cellulose interactions by means of Monte Carlo (MC) simulations of cellulose–IL mixtures. Both IL and cellulose models were represented explicitly, and mixed quantum and molecular mechanics (QM/MM) calculations were used to evaluate the solute and solvent, respectively. Intuitively, to build such a model one would choose to solvate cellulose in IL. While force fields such as OPLS have been developed to describe ILs,[33,34]

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

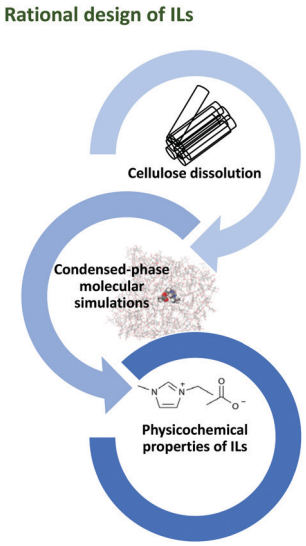


Fig. 1 Schematic representation of a two-tier, reductionist computational approach to design novel ILs *via* rational modifications of IL structures.

adequate sampling of the IL pairs can be problematic,[35,36] particularly in the first solvation shell, and any design changes made to ILs require re-parametrization of the force field – a task that is both onerous and perplexing to the non-expert. Because the phase mixing during cellulose dissolution by ILs is a highly dynamic process,[8] in constructing a computational model to gauge IL–cellulose interactions it is equally reasonable to assume a case where cellulose is the solute in IL solvent as is to assume a model where a single IL pair acts as the solute in cellulose-based medium. This is particularly true when developing a computationally inexpensive model, which must rely on severely-truncated representation of cellulose – in our case, a cellulose monomer (or D-glucose), which was shown useful by Liu and coworkers.[21]

The main advantage of solvating IL in cellulose is the ability to describe the IL pair more accurately using quantum mechanics, which also affords straightforward modifications of the IL pair without the need for force-field parametrization. Additionally, the majority of sampling required to properly describe such a system is delegated to the classically-described cellulose (solvent) model, which can be handled with ease by most available force fields.

In our implementation, IL pairs were represented with the PM7 semiempirical molecular orbital (SMO) method, which is both reasonably accurate and efficient in describing the electronic structure of these systems.[15] We noted that traditional SMO methods such as AM1 or PM3 performed worse in describing geometries of the IL pair; in contrast, higher level density-functional or *ab initio* methods were impractical given sampling requirements. In MC simulations, IL–cellulose interactions used CM3P charges, scaled by a standard factor of 1.12 for IL pairs, for solute atoms in the Coulomb terms[37] and the all-atom OPLS force field for Lennard–Jones interactions.[33] Periodic boundary conditions were invoked in NVT–NPT ensembles at 1 atm and 388 K, which is the experimental temperature reported by Liu *et al.*[21] Cutoffs of 10 Å were used for non-bonded interactions. Solute–solvent systems, which consisted of the IL solute and a solvent box of 267 glucose molecules (*ca.* 35 × 35 × 35 Å in size), were built in the BOSS 4.9 program,[38] with solute geometries minimized prior to MC simulations using Gaussian 16 software.[39] Geometries of IL pairs were optimized with a PM7-SMD approach, which uses an implicit-solvent model based on macroscopic solvent parameters developed by Truhlar *et al.*[40] The distance between the anion and cation was kept fixed during subsequent MC simulation to prevent dissociation, which is known to occur in highly dilute solutions.[41] The MC simulations consisted of sampling 5×10^6 configurations to equilibrate the solvent in the NVT ensemble, followed by 5×10^6 configurations of full-system equilibration in the NPT ensemble and 15×10^6 configurations of averaging to assess system thermodynamics.

In the second-tier analysis, we carried out linear-response calculations based on MC simulations in aqueous phase to evaluate a host of physicochemical properties of ILs.[38] Here, we relied on the TIP4P model,[42] which was used to construct solvent boxes of 500 water molecules with dimensions of *ca.* 25 × 25 × 25 Å. All simulations were carried out at 298 K with periodic boundary conditions and 10 Å cutoffs for all non-bonded interactions. A 2.0×10^5 configurations of solvent equilibration were sampled in the NVT ensemble, followed by 5.0×10^6 configurations of full-system equilibration in the NPT ensemble and 25×10^6 configurations of averaging. The purpose of this analysis was to identify drivers of IL–cellulose interactions, and by extension cellulose solubility, which could be used to rationally design new ILs by means of structural modifications of IL components. It should be noted that properties determined in this report using linear-response calculations in BOSS 4.9,[38] which are based on QM calculations in conjunction with explicit MC simulations, are far more robust and widely applicable than typical structure-based estimation methods.[43–45]

In addition to cellulose dissolution, in the second tier, we used physicochemical properties to also assess ecotoxicity of ILs. To that end, energies of frontier molecular orbitals (FMOs), which were used along with octanol–water partition coefficient ($\log P_{o/w}$) to assess ecotoxicity of studied ILs according to a previously developed approach,[24] were computed using the mPW1PW91 hybrid density functional[46] and the MIDIX+ basis set.[47] The mPW1PW91/MIDIX+ level of theory was found to provide excellent performance-to-cost ratio for orbital energy calculations, and can be readily used to describe larger systems in reasonable timeframes.[48] Assuming dilute IL solutions in ecotoxicity testing, a QM-based implicit-solvent model (SMD)[49] was employed in conjunction with density-functional FMO calculations to estimate the effects of orbital polarization by the aqueous medium.

Dataset

For a mechanistically-driven approach that requires no or few empirical parameters, the size of an experimental dataset is

 This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

less relevant in model training than the quality and diversity of its underlying structural and activity data.[19,50] To that end, and to benchmark our method against one of the most-robust design effort for cellulose dissolution by ILs to date, we chose the set by Liu *et al.*,[21] which used 1-*R*-3-methylimidazolium and *N*-*R*-pyridinium cations with methyl, ethyl, 2-hydroxylethyl, 2-methoxyethyl and acryloyloxypropyl substituents, paired with chloride, bromide, acetate and diethylphosphate anions (Fig. 2). Crucially, the ILs selected by Liu *et al.* cover a broad solubility range (15–46 g mol^{-1} IL), and constitute a structurally diverse mixture of cations and anions, which include chemically-interesting and atypical substituents on imidazolium and pyridinium rings.

Solubility values measured in this dataset are consistent with existing mechanistic knowledge of the dissolution process. For example, the acetate anion was found to be particularly effective due bridging Hbonds within cellulose chains.[51,52] In a different study, aromatic IL cations with electron-donating groups were noted to promote cellulose dissolution by increasing interaction of the side chain with cellulose, while electron-withdrawing groups hindered interactions of ILs with cellulose.[53] Solubility values reported by Liu and co-workers reflect these fundamental observations, showing [Emim]Ac as the most effective IL in the set, and unsaturated side-chain Amim and Apy as more effective than electron-withdrawing HOEtmim, HOEtpy and EtOMmim. Lastly, using a single dataset to validate any computational approach ensures a unified protocol for solubility testing, which is a concern in model development, particularly for statistical approaches that require relatively large datasets pooled from various experimental sources.[12]

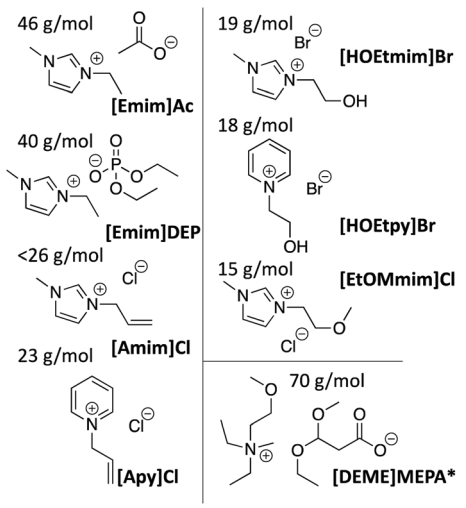


Fig. 2 Structures of ionic liquids (ILs) used in the first tier of computational framework for the design of ILs with high cellulose solubility. Corresponding solubility (g mol^{-1}) values were obtained from ref. 21. * [DEME]MEPA (ref. 12) was used to illustrate our approach on a structurally-different solvent system.

Results and discussion

Informing design of ILs for high cellulose solubility

IL–cellulose interactions. In the first tier of our computational approach, we explored the relationship between IL–glucose interactions and experimental solubility of cellulose in ILs. A sample model system, [Emim]Ac, is shown in Fig. 3. To examine the effect of hydrogen and van der Waals bonding, which are both implicated in cellulose dissolution,[8] we considered separately contributions from electrostatic (Coulomb) and van der Waals (Lennard–Jones, LJ) interactions:

$$\Delta E_{\text{IL-glucose}} = U_{\text{elec}} + U_{\text{vdW}} = \sum_i \sum_j \left\{ \frac{q_i q_j e^2}{r_{ij}} \right\} + \sum_i \sum_j \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\} \quad (1)$$

From eqn (1), ΔE is the potential energy between IL and glucose molecules, which contain interaction sites i and j with associated charges, q, and LJ parameters, σ and ε.

Fig. 4 shows coefficients of determination, R^2, corresponding to univariate correlations between cellulose solubility and the number of IL–glucose interactions and their energies as assessed for the cation, anion and the IL pair. Fig. 4 also shows the correlation between total energetics (Coulomb + LJ) of glucose–glucose interactions and cellulose dissolution.

Table S1† lists all underlying data used to develop these models. It should be noted that correlations reported in Fig. 4 are based on solubility expressed as grams of glucose per mole of IL, which yielded more accurate and reliable predictive models over mass percent (wt%) values in this work and in previous studies.[54]

From Fig. 4, 6 models showed meaningful correlations (R^2 > 0.5) between IL–glucose interactions and solubility, with LJ-interaction energy (E_{LJ}) having the highest correlation (R^2 ~

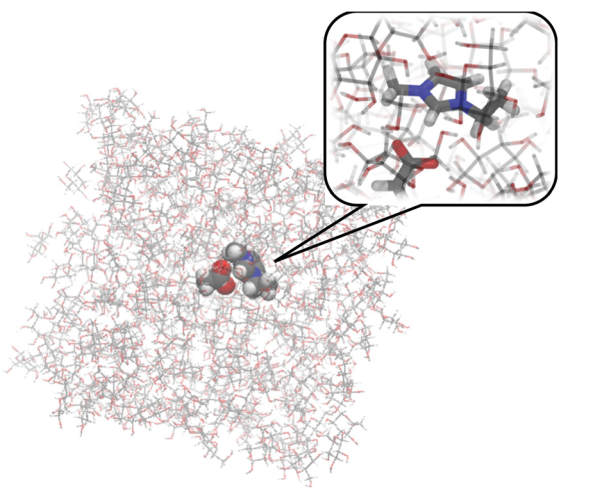


Fig. 3 Equilibrated [Emim]Ac-glucose box as obtained from PM7/OPLS-AA/MC simulations. A close-up window depicts the interacting IL pair.

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

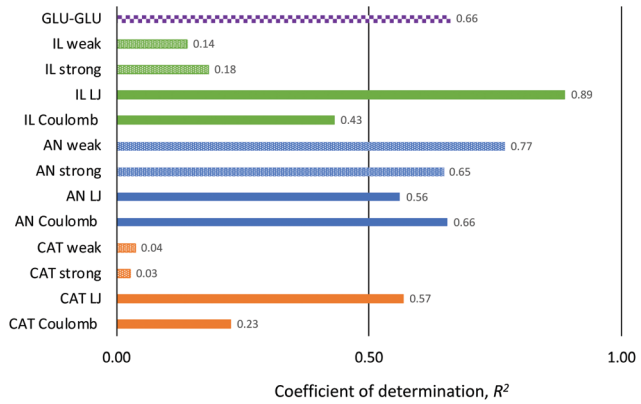


Fig. 4 Coefficients of determination (R^2) for univariate correlations between solute–solvent/solvent–solvent energetics and solubility of cellulose in a series of ILs, as computed from PM7/OPLS-AA/MC simulations. IL = ionic-liquid pair, AN = anion, CAT = cation; solid bars reflect correlations with Coulomb and LJ solute–solvent energetics; dotted bars' "weak" and "strong" labels reflect correlations with the total number of Lennard Jones (LJ) and Coulomb pairwise interactions between the solute and solvent; square-patterned purple bar corresponds to total glucose–glucose, *i.e.* solvent–solvent, interaction energies.

0.9) with experiment (Fig. 5). The trends are consistent in that the greater the interactions, the higher was the apparent solubility of cellulose in the IL. Notably, four single-descriptor models shown in Fig. 4 fitted experimental data better than the COSMO-RS model reported by Liu *et al.*:[21] energetics of the LJ interactions of the IL pair (IL LJ), energetics of the Coulomb interactions of the anion (AN Coulomb), and the number of both strong and weak interactions of the anion (AN strong and weak) with glucose. Consistent with a study by Zavrel *et al.*,[55] aromatic structure of IL cations afforded strong interactions with glucose (Table S1†); however, no useful trend impacting solubility was observed for individual cations.

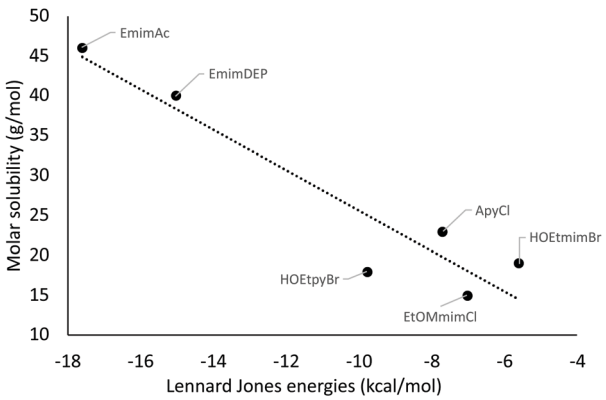


Fig. 5 Experimental solubility of cellulose in ILs plotted as a function of Lennard Jones interaction energy as computed from PM7/OPLS-AA/MC simulations of IL pairs in a glucose solvent box. Solubility (g mol^{-1}) = $-2.536 \times E_{LJ} + 0.304$. $R^2 = 0.89$; $R^2_{adj} = 0.86$; standard error (SE) = 1.8. AmimCl was omitted from the analysis due to the lack of an exact experimental value, *viz*. ref. 21.

In agreement with previous reports, our analysis showed that differences in cellulose dissolution were dominated by the anion as the hydrogen-bond acceptor, with cation playing a secondary role.[8,55] An important correlation noted in Fig. 4 is that between total glucose–glucose interaction energies (Coulomb + LJ) and solubility (GLU–GLU, $R^2 = 0.66$). As the LJ energetics become more favourable between the IL and glucose, driving the observed increase in cellulose dissolution, the glucose–glucose interaction energies diminish, reflecting disruption in the nonbonded network of this cellulose model. This relationship is dominated by decrease in Coulomb interactions, which represent the Hbonds between neighbouring glucose molecules (Table S1†). In fact, removing the sole outlier, [EtOMmim]Cl, increases the correlation between glucose–glucose interaction energies and solubility of cellulose to an impressive R^2 of 0.89 for total interactions and R^2 of 0.95 for coulombic interactions only.

Among debated issues in the field of cellulose dissolution, there is controversy as to whether ILs dissociate in the dissolution process or act as an ion pair.[8,56] While the answer is surly "it depends" (*e.g.* on the stage in the dissolution process, IL concentration and electronic properties, as well as experimental conditions such as temperature), our results, which indicate stronger correlation between solubility and IL pair energetics than solvation metrics obtained for individual ions, support the latter hypothesis. To examine this issue further, we briefly considered the energetics of the cation–anion separation in the glucose medium for [Emim]Ac, the most effective IL, and [EtOMmim]Cl, the least effective IL in dissolving cellulose in Liu's dataset. To that end, we carried out free energy perturbation (FEP) calculations in glucose, where free energy changes were computed with statistical perturbation theory in a windowing format with double-wide sampling.[57] In this combined MC/FEP approach, each FEP window entailed 2.0×10^6 configurations of glucose equilibration in the NVT ensemble, followed by 5.0×10^6 configurations of full-system equilibration and 10×10^6 configurations of averaging in the NPT ensemble. Potentials of mean force (PMF) were computed by perturbing the distance between the cation and anion of the IL pairs at 0.05 Å increments per FEP window, starting from the interacting IL-pair equilibrium state to *ca.* 6.5 Å of cation–anion separation (Fig. 6). In [Emim]Ac, this distance was measured between the oxygen (O) of the acetate and the carbon (C) between the two nitrogen (N) atoms of the imidazolium; in [EtOMmim]Cl, the perturbed distance corresponded to that between chloride (Cl^-) and the same carbon on the imidazolium. From Fig. 6, the process is noted to be favourable, as is known for highly dilute IL solutions.[41] Importantly, the separation is *ca.* 2.1 kcal mol^{-1} more favourable for [EtOMmim]Cl, which is the solvent least capable of dissolving cellulose in our dataset. This result indicates that the ion-pair separation is likely not the driving force for cellulose dissolution. Our hypothesis is further supported by examining the correlations between IL-glucose LJ-interaction energy and solubility at large separation of the cation and anion for all 7 ILs in the dataset. At anion–cation distance of 6.5 Å, this corre-

 This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

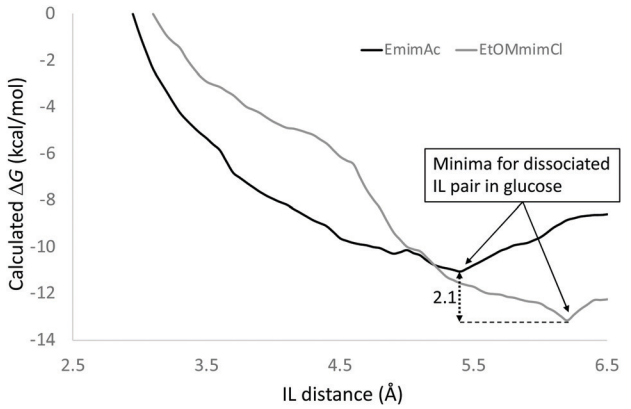


Fig. 6 Computed potentials of mean force (PMF) for separating the cation and anion of EmimAc (black curve) and EtOMmimCl (gray curve). A 2.1 kcal mol^{-1} difference in free energies of the dissociated-pair minima is noted.

lation decreases to $R^2 = 0.49$, which is considerably lower than $R^2 = 0.89$ noted for the interacting ion pair. Furthermore, no meaningful correlation was observed between shorter-range Coulomb interactions and experimental solubility at large distances.

To briefly demonstrate the utility of the first tier of our computational approach in informing design of ILs with high ability to dissolve cellulose, we considered *N*,*N*-diethyl-2-methoxy-*N*-methylethanaminium methoxyethoxypropionate, or [DEME]MEPA (Fig. 2), which is one of the few ILs more effective in dissolving cellulose than [Emim]Ac, with reported solubility of 70.4 g mol^{-1}.[12] An alkylammonium salt, this IL is very different from the imidazolium- and pyridinium-based ILs in the set by Liu *et al.*[21] Using our linear model for LJ interactions and solubility (Fig. 5), we estimated solubility at 75.5 mol g^{-1} based on computed LJ energy, −29.6 kcal mol^{-1}. Our analysis correctly predicts [DEME]MEPA to have the highest solubility of the series in Fig. 2. The small disparity between the estimated and experimental value is not surprising given the use of different experimental protocols in this study[12] *versus* that of Liu *et al.*[21]

Physicochemical properties of ILs. Because the strength of IL–glucose LJ interactions cannot be intuitively affected by structural manipulations to achieve desired outcome in cellulose dissolution, in the second tier of our computational framework (Fig. 1), we explored the role of physicochemical properties of ILs as drivers of IL–glucose interactions. We focused, in particular, on LJ energies between IL and glucose (E_{LJ}), which showed highest correlations with experimental solubility values. A comprehensive list of 18 properties were calculated or predicted from QM/MM simulations, and correlated with E_{LJ} (Table S2†). In this purely computational tier, we expanded our dataset to 31 ILs (Fig. S1†), incorporating a diverse mix of cations and anions, which represent the breath of the IL chemical space and have been assessed for their environmental toxicity, a consideration that is instrumental to our later discussion of safety-design criteria.

Two physicochemical properties with significant correlations to E_{LJ} that could be used to inform design of new ILs emerged: aqueous solubility, $\log S_w$ (Fig. 7 top, $R^2 = 0.85$) and polarizability (Fig. 7 bottom, $R^2 = 0.65$). In both cases, expected trends were observed: the greater the polarizability and the lower the $\log S_w$, *i.e.* the more hydrophobic the IL pair, the more favourable were the LJ interactions between IL and glucose medium. These findings are consistent with our understanding of LJ interactions, which are dominated by weak induced-dipole London dispersion forces. The role of hydrophobicity in the dissolution of cellulose remains controversial;[8] however, the markedly amphiphilic nature of cellulose and its aqueous insolubility support our model's hypothesis. Furthermore, there are numerous reports suggesting that hydrophobic interactions (*versus* Hbonds) are the decisive factor for the behaviour of cellulose in the dissolution process.[58–61] For example, Lindman *et al.* observed in Cl NMR studies that quadrupole relaxation effects of chloride ions, which were attributed to interactions with hydroxyl groups of cellulose,[62] were much stronger for ILs with hydrophobic groups than for those that were hydrophilic.[63]

To briefly demonstrate how these properties can be applied to guide the design of ILs for high cellulose solubility, we revisited the [DEME]MEPA salt discussed previously. Predicted aqueous solubility, $\log S_w$ (−5.88) and polarizability (32.4 $Å^3$)

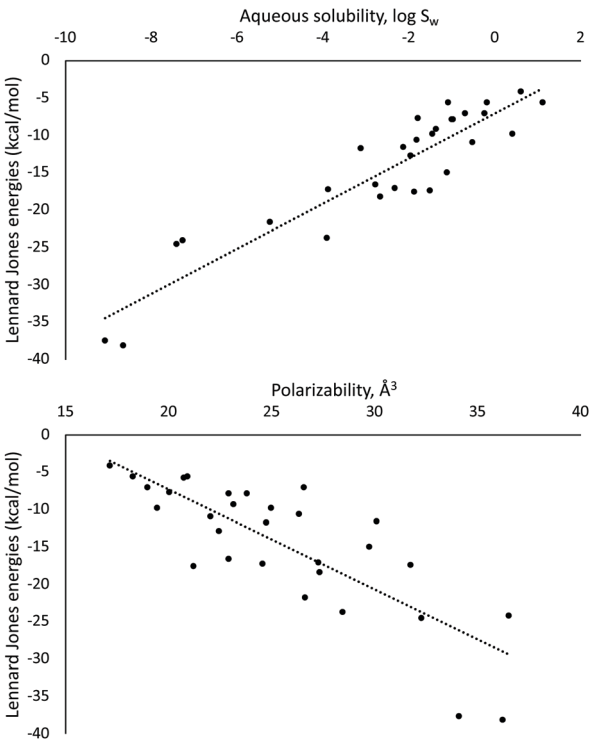


Fig. 7 Computed Lennard Jones interaction energies plotted as a function of IL aqueous solubility, $\log S_w$ (top) and IL polarizability (bottom). Top: $E_{LJ} = 3.0186 \times \log S_w - 7.0683$; $R^2 = 0.85$, $R^2_{adj} = 0.84$, standard error (SE) = 3.46; bottom: $E_{LJ} = -1.3414 \times$ Polrz + 19.581; $R^2 = 0.65$, $R^2_{adj} = 0.64$, SE = 5.22.

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

yielded E_{LJ} values of −24.8 and −23.9 kcal mol^{-1}, respectively, based on our linear models in Fig. 7. These values suggest molar-solubility range of *ca.* 61–63 g mol^{-1}, which is in qualitative agreement with the rest of Liu's dataset (Fig. 2).

While aqueous solubility and polarizability provide good structural interpretability as design guidelines, we further explored the role of individual structural components of ILs and the correlations between their properties and LJ interactions. In particular, we focused on the side chains (*R* groups) on the imidazolium and pyridinium cations in Liu's dataset, which are chemically, *i.e.* electronically, diverse (Fig. 2). While cations are thought to play a lesser role in cellulose dissolution, their impact is certainly not negligible.[8] A good example is a series of imidazolium chloride salts with varying lengths of the alkyl side chains (C2–C10), which was tested for cellulose dissolution by Mai *et al*. and yielded a wide range (0.8–20.5 g mol^{-1}) of solubility metrics.[12] It is important to note that while several meaningful correlations were noted in this study between R-group properties and E_{LJ}, the purpose of our analysis at this stage is to provide a qualitatively useful guide for structural modifications of IL cations that would increase cellulose dissolution. Evidently, in examining *R* groups across different ILs, we ignore much of the relevant chemistry, *e.g.* the anionic effect, or the important electronic interplay between the side chain and the cation core. The value proposition of this somewhat crude approach is that side-chain properties can be assessed on the fly using common chemistry toolkits, without the need for more complex molecular simulations. Here, we relied on QikProp 2.2, which calculates physicochemical properties using the PM3 semiempirical method, and estimates partition coefficients using structure–activity relationships.[43,45]

From Table S3,† one notable correlation with E_{LJ} emerged that can be mechanistically justified: the ionization energy, IE (R^2 = 0.81). Ionization energy, which corresponds to the negative energy of the highest-occupied molecular orbital (HOMO) according to Koopman's theorem, is the energy required to remove the most-loosely bound electron from a molecule. We noted that though the range of IE values was relatively narrow in our set of *R* groups, *ca.* 10–12 eV, higher IE values more frequently corresponded to greater magnitudes of E_{LJ} (Table S3†). This result is consistent with the notion that alkyl chains have lower-lying HOMOs and donate electron density to the cation ring, increasing cation–cellulose interactions.[53] In contrast, alkyl chains substituted with electronegative heteroatoms have higher-lying HOMOs that draw electron density from the cation ring, diminishing cation–cellulose interactions.[53]

Informing design of ILs for minimal ecotoxicity

Physicochemical properties of ILs. Though initially heralded as 'green' solvents due to their favourable physicochemical properties when compared to traditional systems, many ILs have since shown to be toxic to enzymes, microorganisms and cells as well as whole animals and plants.[64,65] This is disconcerting given that most ILs are not readily biodegradable.[66] Quantitative structure–activity and structure–property relationships to assess toxicity of ILs toward different species and cell lines have been developed, offering guidance in designing new solvent systems.[67] For example, Couling *et al*. developed a predictive model to determine the chemical and structural factors that govern toxicity to aquatic organisms.[68] However, among the many computational approaches proposed for cellulose dissolution, none explicitly consider toxicity as a fully-integrated design criterion. To develop an effective design strategy for safer ILs, it is essential that these models incorporate available mechanistic information. While toxicity mechanisms of ILs are complex and not entirely understood, most authors point to the disruption of cell membranes.[69] Thus, ILs with lipophilic substituents, which are more likely to interact with cell membranes, are more toxic as a result.[70] Some ILs are also known to inhibit specific enzymes, such as acetylcholinesterases.[64] To that end, reactivity of ILs is also an important factor that should be incorporated into toxicity modelling.

To deliver an easy-to-use tool for designing safer ILs, we considered our previously-developed safer chemical design guidelines[24] to assess toxicity of ILs to aquatic species. Our guidelines define a 'safer' chemical space based on cutoff values in two key properties applicable to ILs: the HOMO–LUMO band gap, ΔE, which must be greater than 6 eV, and the octanol–water partition coefficient, $\log P_{o/w}$, which must be less than 2. These rules are mechanistically justified in the context of ILs: a chemical with a high band gap will have a relatively low chemical reactivity, and one that is not sufficiently lipophilic will not be readily bioavailable (and in the specific case of ILs, less likely to trigger narcosis).[71] In evaluating these parameters for ILs, it is important to consider the interacting ion pair. While toxicity is generally assumed to relate to the nature of the cation,[69] membrane permeability of ILs depends on the ion-pair interactions, as does our interpretation of reactivity *via* orbital-energy calculations. The latter can be shown in Fig. 8, which illustrates the effect of pairing the cation and anion on the HOMO and LUMO orbitals of [Apy]Cl (Fig. 2). Expectedly, if both ions are present then HOMO is largely localized to the anion while LUMO is mainly composed of atomic orbitals of the cation. IL pairs were noted to have smaller band gaps than cations, which results from lowering of the LUMO energy due to delocalization of the orbital in the ion pair and from raising the energy of the HOMO due to presence of the (nucleophilic) anion.

As detailed in the Methods section, our calculations of $\log P_{o/w}$ rely on rigorous QM/MM/MC simulations, which have been extensively validated and found more robust than structure-based estimation methods.[43–45] To that end, and because experimental partition coefficients are highly concentration-dependent,[41] we deemed relative trends across different ILs more important than absolute values, as long as our modelling approach adhered to our previously-developed protocol, allowing the use of an existing cutoff for minimal bioavailability ($\log P_{o/w}$ = 2).[24]

We emphasize that broader design rules, such as those reported here, are particularly relevant due to the established variability in experimental conditions in testing for acute

 This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

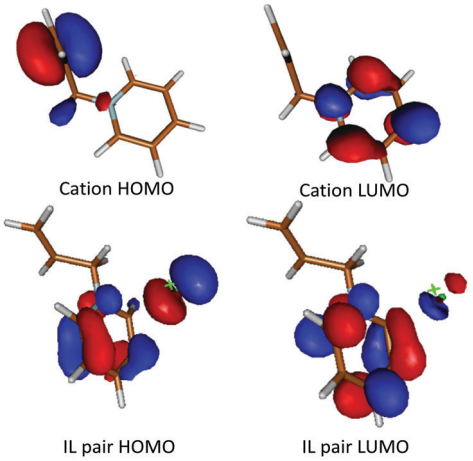


Fig. 8 HOMO and LUMO orbitals of [Apy]Cl calculated at the mPW1PW91/MIDIX+/SMD level of theory, showing difference between the cation and the IL pair in orbital positions and polarization. The Cl^- anion is marked in green colour.

aquatic toxicity, differences in responses across aquatic species and the many (poorly understood) mechanisms of aquatic toxicity.[69,72] Computed ΔE and $\log P_{o/w}$ values, along with experimental hazard labels for acute aquatic toxicity, are summarized in Table 1. Since only [Emim]Ac from the set by Liu *et al.*[21] had a measured lethal concentration at 50% (LC_{50}) value for aquatic species, we included key ILs compiled from several other studies for reference. In our selection, we focused on accepted studies from common aquatic model species, such as *Daphnia magna* or zebrafish, that evaluated both toxic and safe ILs, the latter being far less common.[73–75] Several cholinium-amino acid ([Ch]AA) ILs, which are proposed to be safer and more biodegradable than imidazolium- and pyridinium-based ILs, were included in Table 1 as well.[64] Because for many compounds we identified multiple toxicity thresholds and for some, such as [Ch]AAs, only toxicity potentials toward enzymes and bacteria were tested, we reported toxicity categories rather than numerical acute toxicity thresholds in Table 1. These categories were assigned based on qualitative report summaries in the absence of whole-animal tests or, where available, using measured LC_{50} values and following acute toxicity rating scale set by the US Fish and Wildlife Service (FWS) and the US EPA Safer Choice Program: highly toxic (<1 mg L^{-1}), moderately toxic (1–10 mg L^{-1}), slightly toxic (10–100 mg L^{-1}) and relatively nontoxic (>100 mg L^{-1}).

From Table 1, we find good agreement between our design guidelines for safer chemicals (high ΔE and low $\log P_{o/w}$) and toxicity-concern categories assigned to the 22 ILs with experimental toxicity thresholds. The trend observed between computed properties and aquatic toxicity can be readily extended to other substituents on imidazolium and pyridinium halides,[76] which were omitted from Table 1 for brevity. The model yields good accuracy in category prediction (81%), which is on par with our previous studies,[77] and a very low false-negative rate (1 compound in Table 1). The model thus achieves very good sensitivity (89%). In Table 1, we note that the performance of our design guidelines for safer ILs could be potentially boosted by increasing the $\log P_{o/w}$ cutoff value to 3, resulting in 90% accuracy. Such an adjustment to the cutoff in $\log P_{o/w}$, a property which represents chemical bioavailability in the model, would be perfectly reasonable given the difference in bioavailability mechanisms between ILs and neutral organics, on which the original design guidelines were trained.[70,78]

From the dataset by Liu *et al.*,[21] [Emim]Ac and [Emim]DEP emerged as safe toward aquatic species based on both high band gap and low $\log P_{o/w}$ values. In accordance with experimental observations, imidazolium and pyridinium ILs show higher acute toxicity than "natural" [Ch]AA ILs in Table 1.[69] However, the anion impacts the toxicity category, as shown by

Table 1 Computed HOMO–LUMO gap (ΔE) and predicted $\log P_{o/w}$ values for a series of ILs tested by Liu *et al.*[21] and key ILs tested for ecotoxicity and reported in the literature.[64,68,74,75,79,80] IL structures corresponding to name acronyms in column 1 are provided in Fig. 2 and S1.† Shaded rows mark mispredictions, *i.e.* false positives or negatives, based on safer design guidelines. Computed performance metrics are provided at the end: Accuracy = (TP + TN)/(TP + TN + FP + FN), Sensitivity = TP/(TP + FN), Specificity = TN/(TN + FP), Precision = TP/(TP + FP) where TP = true positives, TN = true negatives, FP = false positives and FN = false negatives

IL	ΔE (eV)	$\log P_{o/w}$	Design guidelines	Exptl ecotox category	Ref.
[Amim]Cl	5.93	4.6	Not safe	N/A	N/A
[Apy]Cl	4.89	4.22	Not safe	N/A	N/A
[Emim]Ac	7.02	-1.31	Safe	Nontoxic	EChA[a]
[Emim]DEP	7.32	0.76	Safe	N/A	N/A
[EtOMmim]Cl	6.18	3.43	Not safe	N/A	N/A
[HOEtmim]Br	5.92	3.29	Not safe	N/A	N/A
[HOEtpy]Br	4.70	3.76	Not safe	N/A	N/A
[C4mim]Cl	6.18	5.31	Not safe	Slightly toxic	73
[C4mim]Br	5.41	5.27	Not safe	Moderately toxic	73
[C4mim]NO3	6.41	-1.38	Safe	Nontoxic	74
[C2mim]NO3	6.41	-2.27	Safe	Nontoxic	74
[C4py]Cl	4.50	5.00	Not safe	Slightly toxic	73
[3MPyC6]Br	3.72	6.96	Not safe	Highly toxic	67
[3MpyC4]Br	3.88	5.4	Not safe	Slightly toxic	67
[C6mim]Br	5.05	6.96	Not safe	Moderately toxic	67
[35MpyC4]Br	3.98	5.86	Not safe	Slightly toxic	67
[4C4Am]Br	6.28	7.73	Not safe	Moderately toxic	67
[4C4Phos]Br	5.82	0.02	Not safe	Moderately toxic	67
[2CAm]2CNCO2	6.84	-2.04	Safe	Slightly toxic	78
[C6mim]NO3	6.47	-0.21	Safe	Nontoxic	74
[C6mim]Cl	5.93	6.20	Not safe	Nontoxic	74
[C4pyrr]Tf2N	8.12	2.41	Not safe	Nontoxic	79
[C2mim]OTs	6.53	0.06	Safe	Nontoxic	79
[C4py]Tf2N	6.29	2.86	Not safe	Nontoxic	79
[Ch]Gly	7.84	-3.08	Safe	Nontoxic	64
[Ch]Ala	7.30	-3.70	Safe	Nontoxic	64
[Ch]Val	7.02	-2.65	Safe	Nontoxic	64
[Ch]Leu	6.91	-2.37	Safe	Nontoxic	64
Accuracy (%)	81	90[b]			
Sensitivity (%)	89	89[b]			
Specificity (%)	75	92[b]			
Precision (%)	73	89[b]			

[a] Hazard category assigned based on LC_{50} > 120 mg L^{-1} obtained from EChA database (EC# 604-344-8, CAS# 143314-17-4). [b] Performance metrics calculated using $\log P_{o/w}$ > 3 cutoff for safer ILs.

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

the case of [C4mim]NO3 (nontoxic) *vs.* [C4mim]Br (moderately toxic) *vs.* [C4mim]Cl (slightly toxic). This difference is reflected in both ΔE and low $\log P_{o/w}$ values. Because quaternary ammonium salts, particularly those with shorter alkyl side chains, are recognized as having lower acute aquatic toxicity than corresponding aromatic ILs;[72] we revisited the [DEME]MEPA compound. This IL was used above to demonstrate the utility of our computational design framework for cellulose dissolution, indicating high performance. Based on computed ΔE and low $\log P_{o/w}$ values, 7.71 eV and −1.63, respectively, [DEME]MEPA is predicted as being acutely nontoxic toward aquatic species.

Design strategy

As illustrated in Fig. 9, our computational model suggests that design opportunities for new ILs exist in the chemical space defined by several computed physicochemical properties, namely by aqueous solubility, $\log S_w$, polarizability, the octanol–water partition coefficient, $\log P_{o/w}$, and the HOMO–LUMO band gap, ΔE.

These properties have foundation in current mechanistic knowledge of the cellulose-dissolution process,[8] and the modes of action for acute aquatic toxicity.[24] The fact that [Emim]Ac, [Emim]DEP and [DEME]MEPA emerged as having high solubility of cellulose while being nontoxic to aquatic species based on our models suggests that an intersection in the IL chemical space exists for these properties to both optimize function and minimize ecotoxicity. It is interesting to note that the natural cholinium-amino acid ILs investigated here, which were all identified as safe toward aquatic species, also displayed properties indicative of good solubility. Cholinium Leucine, [Ch]Leu, fared the best with predicted Lennard Jones energy, E_{LJ} = −15.1 kcal mol^{-1} (based on aqueous solubility, $\log S_w$ = −2.66), which resulted in solubility comparable to [Emim]DEP, *ca.* 40 g mol^{-1}. This is not surprising given structural similarities between [Ch]Leu and [DEME]MEPA, which showed remarkable dissolving ability.

From a structural-modification standpoint, it is imperative to comment on the relationship between hydrophobicity of a molecule (low aqueous solubility) and its lipophilicity as measured by $\log P_{o/w}$:

$$\log S_w = \log S_o - \log P_{o/w} \quad (2)$$

Eqn (2) shows these two properties are not interchangeable. A compound can be hydrophobic but have relatively low $\log P_{o/w}$ if it also has low solubility in octanol, $\log S_o$. While studying drug solubility in water-based systems, Bergström and Larsson showed that the correlation between $\log S_w$ and $\log P_{o/w}$ on 292 drugs was only modest, with R^2 of *ca.* 0.56.[81] In our dataset of 31 ILs, no significant correlation between $\log S_w$ and $\log P_{o/w}$ of the interacting IL pair was observed ($R^2 \sim 0.1$, Table S2†). Using eqn (2), we evaluated octanol solubility, $\log S_o$ in order to rank the ILs in the dataset from the least to the most soluble in octanol (Table S4†). The top 8 ILs identified in this manner had both $\log P_{o/w} < 3$ and showed aqueous solubility, $\log S_w$ comparable to or lower than [Emim]Ac, suggesting relatively high dissolving ability of cellulose based on our model. From these 8 ILs, 5 were quaternary ammonium salts and 5 had a carboxylate anion (Table S2 and Fig. S1†), which is considered the most effective anion in dissolving cellulose.[82] Thus, provided generous cutoff in $\log P_{o/w}$ (value of 2–3, as suggested here), new ILs can be identified *via* structural modifications of existing molecular scaffolds that satisfy minimal ecotoxicity criteria while having sufficiently low aqueous solubility to yield strong LJ energetics in our model, and by extension, high solubility of cellulose.

By relying on mechanistic yet reduced models of the dissolution process, we sought balance between reliability and efficiency. Basic physicochemical properties, such as ionization energy (IE) of IL side chains, can be estimated instantaneously by most available software. Even more rigorous QM/MM/MC simulations used to calculate interaction energetics require only *ca.* 1–2 h of real computer time per compound on high-performance CPUs, and can be parallelized on multi-CPU workstations of computing clusters to screen large datasets. The setup of these calculations is straightforward using our computational protocol and the glucose solvent box developed in this study. Likewise, density functional calculations needed to evaluate ΔE values are relatively fast owing to the small basis set used; *ca.* 12–24 h of CPU time are needed for most ILs. When considered all together, these computing requirements are certainly greater than those of structure-based statistical models used to predict safety[67] and cellulose dissolution metrics;[12] however, they are comparable to the COSMO-RS model,[21] which they outperform, and are far-less costly than highly accurate but very demanding large-scale molecular simulations.[8]

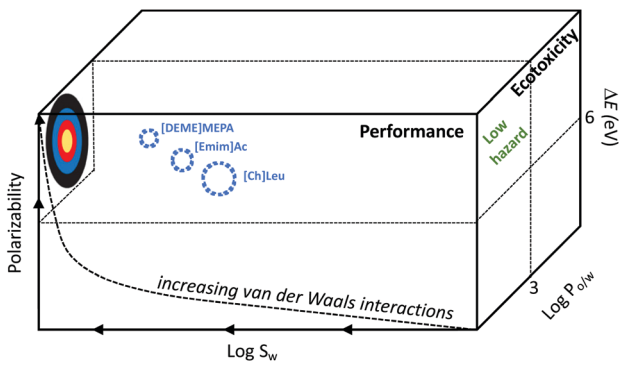


Fig. 9 A 3-D flowchart for designing novel ILs for cellulose dissolution that pose minimal ecotoxicity based on key physicochemical properties. Green marks chemical space with low hazard based on cutoff values in ΔE and $\log P_{o/w}$; arrows along the front-plane *x*- and *y*-axes denote directions in $\log S_w$ and polarizability that optimize van der Waals interactions between ILs and cellulose and increase performance. Exemplary ILs, [DEME]MEPA, [Emim]Ac and [Ch]Leu, which were predicted to be both efficient and safe, are included for reference in their relative positions in the diagram.

 This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

Conclusions

Rational design of functional chemicals with minimal toxicity remains a grand challenge in green chemistry that is yet to be addressed systematically. Here, we demonstrate the development of a computationally inexpensive design strategy for ionic liquids (ILs) as promising designer solvents for cellulose dissolution, a key step in biomass processing. Many ILs have been recognized for their ability to dissolve cellulose by disrupting the vast network of noncovalent interactions between glucose chains. However, much of the vast chemical space encompassed by possible IL-pair combinations and substitutions remains largely unexplored for their potential utility in this application while also considering their environmental safety. To that end, our rational design strategy for developing safer ILs for cellulose dissolution, as described in this report, is based on a two-tier computational framework. The first tier relies on strong quantitative agreement between cellulose solubility and energetics calculated in explicit QM/MM/MC simulations, while the second tier exploits physicochemical properties of ILs to optimize said computed energetics and to inform design of environmentally nontoxic IL analogues.

The goal of our combined approach was to deliver practical guidelines that are mechanistically-sound, transparent and useful in informing structural modifications likely to yield safer and effective solvent systems for cellulose dissolution. The key structure–property relationships identified, which are consistent with existing mechanistic knowledge, are formalized here through both qualitative and quantitative models that show good correlations and consistency with experimental studies owing to a robust theoretical approach. Our work aims to instigate further efforts toward a holistic computational framework for designing safer and efficacious ILs. To that end, while a mechanistically-diverse ecotoxicity model is a good indicator of a broader toxicity potential,[83] additional criteria of 'greenness' and sustainability ought to be incorporated in future studies. In particular, subsequent model iterations should consider physicochemical hazards of ILs as well as comprehensive environmental fate and persistence metrics based on the entire life-cycle of ILs.[84–86]

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 and to Prof. Orlando Acevedo (University of Miami) for sharing his OPLS force-field parameters for ionic liquids. This work was supported by the National Science Foundation (NSF1805080).

References

1 A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer and T. Tschaplinski, *Science*, 2006, **311**, 484–489.
2 D. Tilman, R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville and R. Williams, *Science*, 2009, **325**, 270–271.
3 H. Tadesse and R. Luque, *Energy Environ. Sci.*, 2011, **4**, 3913–3929.
4 M. E. Himmel, S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, 2007, **315**, 804–807.
5 S. Y. Ding and M. E. Himmel, *J. Agric. Food Chem.*, 2006, **54**, 597–606.
6 K. Tashiro and M. Kobayashi, *Polymer*, 1991, **32**, 1516–1526.
7 C. A. Bishop, *Vacuum Deposition onto Webs, Films and Foils*, 2nd edn, 2011, pp. 1–530.
8 Y. Li, J. J. Wang, X. M. Liu and S. J. Zhang, *Chem. Sci.*, 2018, **9**, 4027–4043.
9 R. D. Rogers and K. R. Seddon, *Science*, 2003, **302**, 792–793.
10 N. V. Plechkova and K. R. Seddon, *Chem. Soc. Rev.*, 2008, **37**, 123–150.
11 R. P. Swatloski, S. K. Spear, J. D. Holbrey and R. D. Rogers, *J. Am. Chem. Soc.*, 2002, **124**, 4974–4975.
12 N. L. Mai and Y. M. Koo, *ACS Sustainable Chem. Eng.*, 2016, **4**, 541–547.
13 K. C. Badgujar and B. M. Bhanage, *Bioresour. Technol.*, 2015, **178**, 2–18.
14 S. A. Forsyth, J. M. Pringle and D. R. MacFarlane, *Aust. J. Chem.*, 2004, **57**, 113–119.
15 M. Barycki, A. Sosnowska, M. Piotrowska, P. Urbaszek, A. Rybinska, M. Grzonkowska and T. Puzyn, *J. Cheminf.*, 2016, **8**, 40.
16 B. G. Janesko, *Phys. Chem. Chem. Phys.*, 2011, **13**, 11393–11401.
17 J. Kahlen, K. Masuch and K. Leonhard, *Green Chem.*, 2010, **12**, 2172–2181.
18 B. Mostofian, J. C. Smith and X. L. Cheng, *Interdiscip. Sci.: Comput. Life Sci.*, 2011, **3**, 308–320.
19 H. M. Cho, A. S. Gross and J. W. Chu, *J. Am. Chem. Soc.*, 2011, **133**, 14033–14041.
20 J. Zhang, D. Peng, Z. Song, T. Zhou, H. Cheng, L. Chen and Z. Qi, *Chem. Eng. Sci.*, 2017, **162**, 355–363.
21 Y. R. Liu, K. Thomsen, Y. Nie, S. J. Zhang and A. S. Meyer, *Green Chem.*, 2016, **18**, 6246–6254.
22 A. Klamt, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1338.
23 W. L. Chen and S. T. Lin, *Phys. Chem. Chem. Phys.*, 2017, **19**, 20367–20376.
24 J. Kostal, A. Voutchkova-Kostal, P. T. Anastas and J. B. Zimmerman, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 6289–6294.

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

25 J. Kostal, A. Voutchkova-Kostal, B. Weeks, J. B. Zimmerman and P. T. Anastas, *Chem. Res. Toxicol.*, 2012, **25**, 2780–2787.
26 F. Melnikov, D. Botta, C. C. White, S. C. Schmuck, M. Winfough, C. M. Schaupp, E. P. Gallagher, B. W. Brooks, E. S. Williams, P. Coish, P. T. Anastas, A. Voutchkova-Kostal, J. Kostal and T. J. Kavanagh, *Chem. Res. Toxicol.*, 2019, **32**, 421–436.
27 W. B. Steele, L. A. Kristofco, J. Corrales, G. N. Saari, E. J. Corcoran, B. N. Hill, M. G. Mills, E. Gallagher, T. J. Kavanagh, F. Melnikov, J. B. Zimmerman, A. Voutchkova-Kostal, P. T. Anastas, J. Kostal and B. W. Brooks, *Chem. Res. Toxicol.*, 2020, **33**, 367–380.
28 T. Clymer, V. Vargas, E. Corcoran, R. Kleinberg and J. Kostal, *Green Chem.*, 2019, **21**, 1935–1946.
29 A. M. Voutchkova, L. A. Ferris, J. B. Zimmerman and P. T. Anastas, *Tetrahedron*, 2010, **66**, 1031–1039.
30 A. M. Voutchkova, J. Kostal, J. B. Steinfeld, J. W. Emerson, B. W. Brooks, P. Anastas and B. Zimmerman, *Green Chem.*, 2011, **13**, 2373–2379.
31 A. M. Voutchkova, T. G. Osimitz and P. T. Anastas, *Chem. Rev.*, 2010, **110**, 5845–5882.
32 A. M. Voutchkova-Kostal, J. Kostal, K. A. Connors, B. W. Brooks, P. T. Anastas and J. B. Zimmerman, *Green Chem.*, 2012, **14**, 1001–1008.
33 S. V. Sambasivarao and O. Acevedo, *J. Chem. Theory Comput.*, 2009, **5**, 1038–1050.
34 B. Doherty, X. Zhong, S. Gathiaka, B. Li and O. Acevedo, *J. Chem. Theory Comput.*, 2017, **13**, 6131–6145.
35 D. Bedrov, J. P. Piquemal, O. Borodin, A. D. MacKerell, B. Roux and C. Schroder, *Chem. Rev.*, 2019, **119**, 7940–7995.
36 V. Zeindlhofer and C. Schroder, *Biophys. Rev.*, 2018, **10**, 825–840.
37 M. Udier-Blagovic, P. M. De Tirado, S. A. Pearlman and W. L. Jorgensen, *J. Comput. Chem.*, 2004, **25**, 1322–1332.
38 W. L. Jorgensen and J. Tirado-Rives, *J. Comput. Chem.*, 2005, **26**, 1689–1700.
39 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Anonymous (Gaussian 16, Revision B.01)*, Gaussian, Inc., Wallingford CT, 2016.
40 V. S. Bernales, A. V. Marenich, R. Contreras, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2012, **116**, 9122–9129.
41 P. Jain and A. Kumar, *Phys. Chem. Chem. Phys.*, 2016, **18**, 1105–1113.
42 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
43 E. M. Duffy and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2000, **122**, 2878–2888.
44 W. L. Jorgensen and E. M. Duffy, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 1155–1158.
45 W. L. Jorgensen and E. M. Duffy, *Adv. Drug Delivery Rev.*, 2002, **54**, 355–366.
46 C. Adamo and V. Barone, *J. Chem. Phys.*, 1998, **108**, 664–675.
47 R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer and D. G. Truhlar, *Theor. Chim. Acta*, 1996, **93**, 281–301.
48 B. J. Lynch and D. G. Truhlar, *Theor. Chem. Acc.*, 2004, **111**, 335–344.
49 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
50 A. S. Gross, A. T. Bell and J. W. Chu, *J. Phys. Chem. B*, 2011, **115**, 13433–13440.
51 Y. Li, X. M. Liu, S. J. Zhang, Y. Y. Yao, X. Q. Yao, J. L. Xu and X. M. Lu, *Phys. Chem. Chem. Phys.*, 2015, **17**, 17894–17905.
52 T. Endo, S. Hosomi, S. Fujii, K. Ninomiya and K. Takahashi, *Molecules*, 2017, **22**, 178.
53 Y. L. Zhao, X. M. Liu, J. J. Wang and S. J. Zhang, *ChemPhysChem*, 2012, **13**, 3126–3133.
54 N. L. Mai, C. K. Kim, B. Park, H. J. Park, S. H. Lee and Y. M. Koo, *J. Mol. Liq.*, 2016, **215**, 541–548.
55 M. Zavrel, D. Bross, M. Funke, J. Buchs and A. C. Spiess, *Bioresour. Technol.*, 2009, **100**, 2580–2587.
56 A. M. Fernandes, M. A. A. Rocha, M. G. Freire, I. M. Marrucho, J. A. P. Coutinho and L. M. N. B. F. Santos, *J. Phys. Chem. B*, 2011, **115**, 4033–4041.
57 O. Acevedo and W. L. Jorgensen, *Acc. Chem. Res.*, 2010, **43**, 142–151.
58 B. Lindman, B. Medronho, L. Alves, C. Costa, H. Edlund and M. Norgren, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23704–23718.
59 M. Gubitosi, H. Duarte, L. Gentile, U. Olsson and B. Medronho, *Biomacromolecules*, 2016, **17**, 2873–2881.
60 B. Medronho, H. Duarte, L. Alves, F. Antunes, A. Romano and B. Lindman, *Nord. Pulp Pap. Res. J.*, 2015, **30**, 58–66.
61 C. H. Haigler, R. M. Brown Jr. and M. Benziman, *Science*, 1980, **210**, 903–906.
62 R. C. Remsing, R. P. Swatloski, R. D. Rogers and G. Moyna, *Chem. Commun.*, 2006, 1271–1273, DOI: 10.1039/b600586c.
63 B. Lindman, H. Wennerstrom and S. Forsen, *J. Phys. Chem.*, 1992, **96**, 5669–5670.
64 X. D. Hou, Q. P. Liu, T. J. Smith, N. Li and M. H. Zong, *PLoS One*, 2013, **8**(3), e59145.
65 P. T. P. Thi, C. W. Cho and Y. S. Yun, *Water Res.*, 2010, **44**, 352–372.

 This journal is © The Royal Society of Chemistry 2020

Published on 13 May 2020. Downloaded by George Washington University on 7/8/2020 4:02:14 PM.

66 D. Coleman and N. Gathergood, *Chem. Soc. Rev.*, 2010, **39**, 600–637.
67 N. Abramenko, L. Kustov, L. Metelytsia, V. Kovalishyn, I. Tetko and W. Peijnenburg, *J. Hazard. Mater.*, 2020, **384**, 121429.
68 D. J. Couling, R. J. Bernot, K. M. Docherty, J. K. Dixon and E. J. Maginn, *Green Chem.*, 2006, **8**, 82–90.
69 M. C. Bubalo, K. Radosevic, I. R. Redovnikovic, I. Slivac and V. G. Srcek, *Arh. Hig. Rada Toksikol.*, 2017, **68**, 171–179.
70 K. S. Egorova and V. P. Ananikov, *ChemSusChem*, 2014, **7**, 336–360.
71 J. Ranke, K. Molter, F. Stock, U. Bottin-Weber, J. Poczobutt, J. Hoffmann, B. Ondruschka, J. Filser and B. Jastorff, *Ecotoxicol. Environ. Saf.*, 2004, **58**, 396–404.
72 M. Petkovic, K. R. Seddon, L. P. N. Rebelo and C. S. Pereira, *Chem. Soc. Rev.*, 2011, **40**, 1383–1403.
73 A. S. Wells and V. T. Coombe, *Org. Process Res. Dev.*, 2006, **10**, 794–798.
74 D. B. Zhao, Y. C. Liao and Z. D. Zhang, *Clean: Soil, Air, Water*, 2007, **35**, 42–48.
75 C. Zhang, L. S. Zhu, J. H. Wang, J. Wang, T. T. Zhou, Y. Q. Xu and C. Cheng, *Ecotoxicol. Environ. Saf.*, 2017, **140**, 235–240.
76 K. M. Docherty and C. F. Kulpa, *Green Chem.*, 2005, **7**, 185–189.
77 F. Melnikov, J. Kostal, A. Voutchkova-Kostal, J. B. Zimmerman and P. T. Anastas, *Green Chem.*, 2016, **18**, 4432–4445.
78 K. S. Egorova, E. G. Gordeev and V. P. Ananikov, *Chem. Rev.*, 2017, **117**, 7132–7189.
79 S. P. F. Costa, P. C. A. G. Pinto, M. L. M. F. S. Saraiva, F. R. P. Rocha, J. R. P. Santos and R. T. R. Monteiro, *Chemosphere*, 2015, **139**, 288–294.
80 C. Pretti, C. Chiappe, D. Pieraccini, M. Gregori, F. Abramo, G. Monni and L. Intorre, *Green Chem.*, 2006, **8**, 238–240.
81 C. A. S. Bergstrom and P. Larsson, *Int. J. Pharm.*, 2018, **540**, 185–193.
82 H. Wang, G. Gurau and R. D. Rogers, *Chem. Soc. Rev.*, 2012, **41**, 1519–1537.
83 K. Howe, M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren and L. Matthews, *Nature*, 2013, **496**, 498–503.
84 S. Wellens, B. Thijs and K. Binnemans, *Green Chem.*, 2013, **15**, 3484–3485.
85 A. O. Diallo, C. Len, A. B. Morgan and G. Marlair, *Sep. Purif. Technol.*, 2012, **97**, 228–234.
86 M. C. Bubalo, K. Radosevic, I. R. Redovnikovic, J. Halambek and V. G. Srcek, *Ecotoxicol. Environ. Saf.ty*, 2014, **99**, 1–12.