



Cite this: *Green Chem.*, 2021, **23**, 7269

## A predictive toolset for the identification of effective lignocellulosic pretreatment solvents: a case study of solvents tailored for lignin extraction†

Ezinne C. Achinivu,<sup>‡,a,b</sup> Mood Mohan,<sup>‡,a,b</sup> Hemant Choudhary,<sup>‡,a,b</sup> Lalitendu Das,<sup>a,b</sup> Kaixuan Huang,<sup>a,c,d</sup> Harsha D. Magurudeniya,<sup>a,b</sup> Venkataramana R. Pidatala,<sup>a,c</sup> Anthe George,<sup>‡,a,b</sup> Blake A. Simmons,<sup>‡,a,c</sup> and John M. Gladden <sup>‡,a,b</sup>

Pretreatment of lignocellulosic biomass is essential for efficient conversion into biofuels and bioproducts. The present study develops a predictive toolset to computationally identify solvents that can efficiently dissolve lignin and therefore can be used to extract it from lignocellulose during pretreatment, a process known to reduce recalcitrance to enzymatic deconstruction and increase conversion efficiency. Two approaches were taken to examine the potential of eleven organic solvents to solubilize lignin, Hansen solubility parameters (HSP) and activity coefficients and excess enthalpies of solvent/lignin mixtures predicted by COSMO-RS (COnductor like Screening MOdel for Real Solvents). The screening revealed that diethylenetriamine was the most effective solvent, promoting the highest lignin removal (79.2%) and fermentable sugar yields (>72%). Therefore, a COSMO-RS-based predictive model for the lignin removal as a function of number and type of amines was developed. Among the fitted models, the non-linear regression model predicts the lignin solubility more accurately than the linear model. Experimental results demonstrated a >65% lignin removal and >70% of sugar yield from several amine-based solvents tested, which aligned very well with the model's prediction. Finally, to help understand the dissolution mechanism of lignin by these solvents, quantum theory of atoms in molecules (QTAIM) and quantum chemical calculations (interaction energies and natural bond orbital (NBO) analysis) was performed and suggest that amines exhibit strong electrostatic interactions and hydrogen bonding strengths with lignin leading to higher lignin removal. Together, these computational tools provide an effective approach for rapidly identifying solvents that are tailored for effective biomass pretreatment.

Received 7th April 2021,  
Accepted 12th July 2021

DOI: 10.1039/d1gc01186c  
[rsc.li/greenchem](http://rsc.li/greenchem)



<sup>a</sup>Deconstruction Division, Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, USA. E-mail: [jmgladden@lbl.gov](mailto:jmgladden@lbl.gov), [jmgladd@sandia.gov](mailto:jmgladd@sandia.gov)

<sup>b</sup>Department of Biomass Science and Conversion Technology, Sandia National Laboratories, 7011 East Avenue, Livermore, California 94551, USA

<sup>c</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA

<sup>d</sup>Key Laboratory of Forestry Genetics & Biotechnology (Nanjing Forestry University), Ministry of Education, Nanjing 210037, People's Republic of China

†Electronic supplementary information (ESI) available: COSMO-RS calculated chemical potentials of investigated molecules, viscosity validation, lignin developed predictive model plots, and quantum-based calculated intermolecular geometries, QTAIM properties, biomass compositional analyses, and the coordinates (XYZ) of most stable conformation structural information. See DOI: [10.1039/d1gc01186c](https://doi.org/10.1039/d1gc01186c)

‡These authors contributed equally to this work.

## 1. Introduction

Lignocellulosic biomass is one of the primary natural resources that can be utilized as a renewable source of key intermediates to produce bioenergy, chemicals, and biomaterials.<sup>1,2</sup> However, the distinctive physical and chemical diversity of its major constituents (cellulose, hemicellulose, and lignin), along with its complex microstructure and chemical linkages make it difficult to be effectively processed or fractionated into its various constituents. This key hurdle must be overcome for the development of sustainable biorefineries and a robust bio-based economy.<sup>1,3</sup> To effectively accomplish this, the cross-linked matrix of lignin and hemicelluloses that surrounds the cellulose fibers must be disrupted.<sup>4</sup> In particular, the extraction of lignin, the component that predominantly makes the biomass recalcitrant to deconstruction, is necessary

to enable the efficient conversion of lignocellulosic intermediates into valuable products (fuels and/or chemicals).<sup>1,4–7</sup>

There are several different pretreatment strategies that have been investigated for the separation of pure lignin and amongst them four chemical industrial processes are noteworthy: sulfite, kraft, soda and organosolv pretreatments.<sup>6,8</sup> Out of these methods, the organosolv fractionation process has been widely accepted as one of the most promising techniques for biomass fractionation due to its comparatively low environmental impact, high delignification efficiency, and the diversity of products that are released.<sup>1,9,10</sup> Organosolv pulping or fractionation is one of the methods of biomass fractionation that can produce high-quality cellulosic biofuels (*via* monomeric sugar fermentation), along with a high purity lignin. Unlike other pretreatment methods, the organosolv process is sulfur free, thereby producing lignin streams with a high level of purity for subsequent valorization.<sup>6,9,11</sup> Additionally, this approach is particularly appealing because of the possibility of recovery and recycling of the organic solvent.<sup>12–14</sup> In a typical organosolv process, an organic solvent is used to pretreat lignocellulosic biomass with or without the addition of external catalysts.<sup>13,14</sup> Organic solvents such as short alkyl chain aliphatic alcohols (*e.g.*, methanol, ethanol), polyols (*e.g.*, glycerol, ethylene glycol, triethylene glycol), amines, alkanolamines, organic acids, acetone, dioxanes, and phenols have been widely used for the organosolv process.<sup>12,15</sup> In most cases, the biomass pretreated by organic solvents is very susceptible to hydrolysis (*via* enzymes) and can be readily deconstructed to yield monomeric sugars.<sup>12,14,16</sup> Cheng *et al.* studied the ability of 12 organic solvents including alcohols, alcohol ethers, lactones, and alkanolamines, to fractionate poplar and rice straw and reported at least 70% delignification.<sup>17</sup> Zhao *et al.* also reported >90% conversion of the polysaccharides (cellulose/hemicelluloses) for alcohol-pretreated biomasses,<sup>18</sup> and, Qin *et al.* reported that ethylenediamine can be applied to corn stover, resulting in glucose and xylose yields of 92% and 70% respectively after enzymatic digestion.<sup>19</sup>

Despite the promise for the organosolv processes, the near limitless possibilities for solvent selection have not been fully explored within the context of a robust multi-product biorefinery. Solvents like alcohols and diols have dominated the organosolv literature,<sup>9,15,20</sup> yet many other possible solvents may exist with better performance and/or recyclability. The identification of these solvents would be greatly accelerated by the development of a computational toolset that could predict lignin solubilization and be systematic and efficient. Nevertheless, researchers still require guidelines to be established for the choice of successful solvent systems to become methodical. These guidelines or design rules should offer insights into the key chemical functionalities within a solvent that promote lignin dissolution, as well as the structural and conformational variations within a solvent group that can affect it. Lastly, it would be ideal if this toolset could aid in revealing the mechanistic factors that control lignin dissolution, which would help further refine the design/development of new and effective solvent systems for lignin.

Alongside experimental studies, molecular simulations have also been employed to understand the dissolution mechanism of biomass and its components. Researchers have adopted quantum chemical (QC) and molecular dynamics (MD) simulations, which provide fundamental insights of the molecular systems (*e.g.*, lignin and ionic liquids). Solubility parameters such as Hildebrand<sup>21–23</sup> and Hansen solubility parameters (HSP)<sup>17,24–27</sup> and the COSMO-RS (Conductor like Screening MOdel for Real Solvents) model have been widely used to design and develop effective solvents for biomass delignification.<sup>28–31</sup> Balaji *et al.*<sup>28</sup> and Casas *et al.*<sup>29,30</sup> screened various ionic liquids (ILs) to understand the lignin dissolution ability by predicting Hildebrand solubility parameters and thermodynamic parameters namely excess enthalpy and activity coefficient using COSMO-RS. Casas reported that the strong exothermic behavior of excess enthalpy and lower activity coefficients are beneficial for higher lignin dissolution.<sup>29,30</sup> However, in both studies, only lignin's monomeric structures were employed as a model component. Later, Zhang *et al.*<sup>32</sup> and Ji *et al.*<sup>33</sup> performed quantum chemical (density functional theory) simulations to reveal the mechanism of lignin dissolution in imidazolium-based ionic liquids. It has been reported that the stronger H-bonding interaction between lignin and IL is responsible for the greater ability to dissolve lignin. These molecular simulation techniques can help in identifying new potential effective solvents for biomass pretreatment. However, the dissolution mechanism of lignin from lignocellulosic biomass using molecular solvents and the development of a predictive model for lignin removal has not yet been fully addressed. There is still a need to develop a predictive model for lignin removal, which can describe the solubility, while exploring the relationship between lignin dissolution and pretreatment effectiveness.

The present study attempts to develop both predictive models to identify the best solvents for lignin dissolution and multiscale simulation approaches tailored to provide mechanistic insights into how these solvents interact with lignin. First, HSP were used to screen a wide range of molecular solvents and identify ones that may be effective at lignin extraction from lignocellulose, which were then tested experimentally to determine the accuracy of the HSP predictions. Next, COSMO-RS calculations were performed to examine the same solvents and study the solvent/lignin mixture's thermodynamic properties such as excess enthalpy, activity coefficient, and sigma potentials. The excess enthalpy and activity coefficients were then used as a method to rank these solvents' ability to dissolve lignin, and this approach was compared to using HSP and found to have better predictability. The initial screening revealed that solvents containing amine were effective at dissolving lignin, so a broader class of amine-based solvents were screened using excess enthalpy and activity coefficients of solvent/lignin mixtures, and several amines were tested experimentally for their ability to extract lignin from biomass and promote efficient enzymatic saccharification of the lignocellulosic polysaccharides. This data was used to validate a COSMO-RS-based predictive model that was developed for the



lignin removal as a function of number and type of amines, which can be used for future screening efforts of amine-based solvents. Finally, to gain a deeper mechanistic understanding of how these solvents act to dissolve lignin, quantum chemical simulations are performed to study the solvent's interactions with lignin. Quantum theory of atom in the molecule (QTAIM), reduced density gradient (RDG), and natural bonding orbital (NBO) analysis were also carried out to investigate the strength and nature of H-bonding present in the lignin/molecular solvents. This analysis provided key insights into lignin dissolution and revealed that H-bonding between solvent and lignin is a major driver of lignin dissolution. The predictive toolset developed in this study combined with the mechanistic insights into lignin dissolution lay a strong foundation for rapidly identifying effective solvents for biomass pretreatment and developing cost-effective lignocellulosic conversion technologies.

## 2. Results and discussion

### 2.1. Screening of molecular solvents for lignin dissolution

Lignin is a complex and random polymer that is held together with strong bonds, such as ether linkages (carbon–oxygen) and carbon–carbon bonds, as well as, weak inter- and intra-molecular forces, such as hydrogen bonds.<sup>34</sup> The complete (or near-complete) dissolution of lignin and its removal from lignocellulosic biomass is challenging owing to the scarcity of effective, efficient, economic, and environmentally benign solvents. Therefore, methods that enable rapid identification of effective lignin solvents could be instrumental in expanding the list of available solvents and identifying those that can be incorporated into cost-effective lignocellulose conversion technologies.

Hansen solubility parameters (HSP) could possibly be used to provide an expedient route to identifying a short list of good potential solvents for lignin dissolution. HSP values for many molecular solvents have been determined and these values are readily available. In addition, Hansen and Björkman report relative energy difference (RED) values for many solvents compared to lignin.<sup>27</sup> RED values can be used to estimate a solvent's ability to dissolve a solute. If the RED value is less than 1, then the affinity between the solute and the solvent is said to be higher and will result in a higher dissolution capacity. If the RED is greater than 1, the affinity between the solvent and solute is lower, resulting in poor dissolution. These RED values were used as an initial screen to identify a short list of solvents with RED values less than 1 that could be tested experimentally for lignin dissolution. The solvents were intentionally selected to have different molecular functionalities to maximize the chemical space covered. Functional groups included amines (diethylenetriamine), lactams (2-pyrrolidone), alcohols (dipropylene glycol, benzyl alcohol, furfuryl alcohol, guaiacol, and 2-ethoxyethanol), ethers (2-ethoxyethanol and di-propylene glycol), esters (isobutyl acetate and trimethyl phosphate), and aromatics (benzyl alcohol, furfuryl alcohol, guaiacol, and furfural) (Fig. 1A).

While the Hansen and Björkman reported HSP and RED values for extracted woody lignin (14.9 for polar ( $\delta_p$ ), 16.9 for hydrogen-bonded ( $\delta_h$ ), and 21.9 for dispersion ( $\delta_d$ ) contributor) are readily available, they are also based on kraft lignin extracted from pine trees during paper pulping, which is unlikely to have the same properties as intact lignin within plant biomass.<sup>27,35</sup> In addition, it should be noted that the reported HSP values cannot be assumed to be universal for all lignin samples as there is an extensive chemical diversity that exists between the lignins from different biomass sources. Considering these issues, we sought to identify an alternate set of lignin HSP that could be used to calculate lignin RED values to accurately rank the selected solvent's ability to dissolve lignin. Thielemans and Wool have reported the HSP values for lignin as  $\delta_p = 13.7$ ,  $\delta_h = 11.7$ , and  $\delta_d = 16.7$ .<sup>36</sup> In their model, the solubility behavior of the modified lignin was described using the Flory–Huggins solubility theory, combined with the group contribution model developed by Hoy.<sup>36–38</sup> This is one of the more practical lignin models available because it has contributions for a large number of functional groups, and accounts for a variety of structural features, which is important for a complex polymer like lignin.

To develop a more accurate set of RED values for lignin solvents, the Thielemans and Wool reported HSPs were used to calculate a new set of RED values for the same solvents identified by Hansen and Björkman using COSMOquick. In the new set of RED values, diethylenetriamine and trimethylphosphate have the lowest RED values and are expected to be the most suitable solvents for delignification, while the other solvents are expected to extract little to no lignin (Table 1). To validate these predictions, the grassy crop sorghum was pretreated with the solvents listed in Table 1 at 140 °C for 3 h at 20 wt% solids loading. Pretreatment with diethylenetriamine resulted in the highest lignin extraction (79.2%) as predicted, but pretreatment with trimethylphosphate resulted in an unexpectedly low-level lignin extraction (28.5%; Table 1). None of the other solvents were able to extract high levels of lignin from sorghum. Therefore, the calculated RED values do not appear to be very predictive for lignin extraction from lignocellulose, and screening solvents with these values will likely result in many false positives.

It is unclear why the HSP values are not very predictive for lignin dissolution, but one explanation is that HSP values are used to measure the intermolecular affinity between solvent and solute but do not account for their intramolecular affinities, which can affect their behavior. Therefore, to better understand both the inter- and intramolecular interactions in a lignin/solvent mixture, COSMO-RS calculations were performed to study the mixture's thermodynamic properties such as excess enthalpy, activity coefficient, and sigma potentials. Typically, monomeric and dimeric structures of lignin have been used as lignin models to perform these molecular simulations.<sup>28–30</sup> However, the monomeric and dimer structures of lignin do not directly represent the lignin molecule due to the absence of many different linkages present in lignin. Therefore, to obtain more realistic results, a lignin



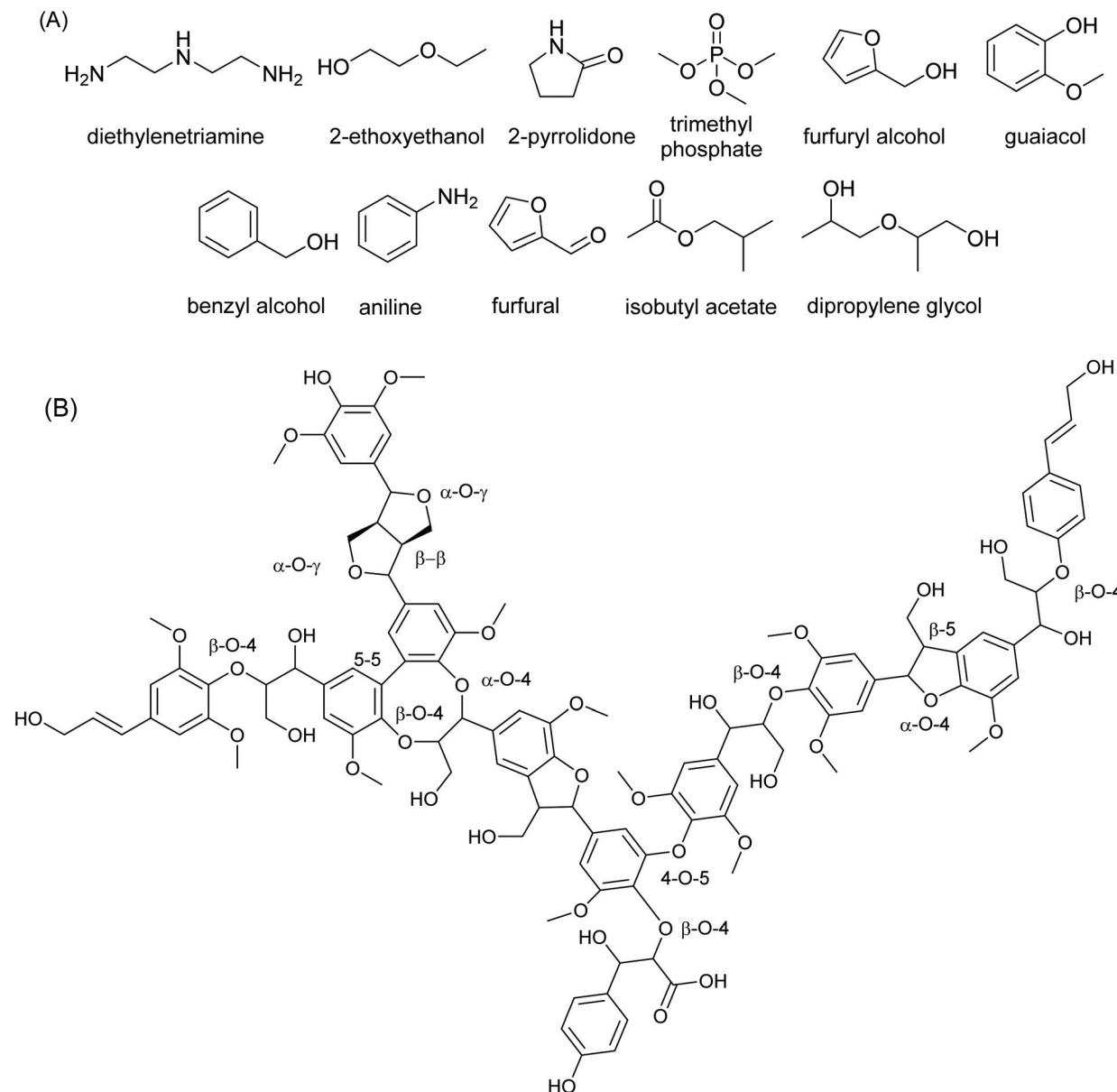


Fig. 1 Chemical structures of (A) the organic solvents screened (B) the lignin model used in this study for COSMO-RS calculations.

structure was generated based on the G/S ratio of grassy biomass and built by joining all the major lignin linkages ( $\beta$ -O-4,  $\beta$ - $\beta$ , 4-O-5,  $\alpha$ -O-4, and  $\beta$ -5) present in the native lignin (Fig. 1B). As mentioned earlier, lignin is a heterogeneous macromolecule, therefore, it is not possible to create a single lignin structure that can fully capture that heterogeneity or represent all lignins. However, many insights can be gained by simply ensuring coverage of the typical linkages found in lignin for the biomass used for pretreatment, which is this study is the grass sorghum.

Two thermodynamic properties may be useful in predicting lignin dissolution in a solvent, excess enthalpy ( $H^E$ ) and logarithmic activity coefficients ( $\ln(\gamma)$ ). The  $H^E$  is a useful thermodynamic property for measuring the difference in the strength

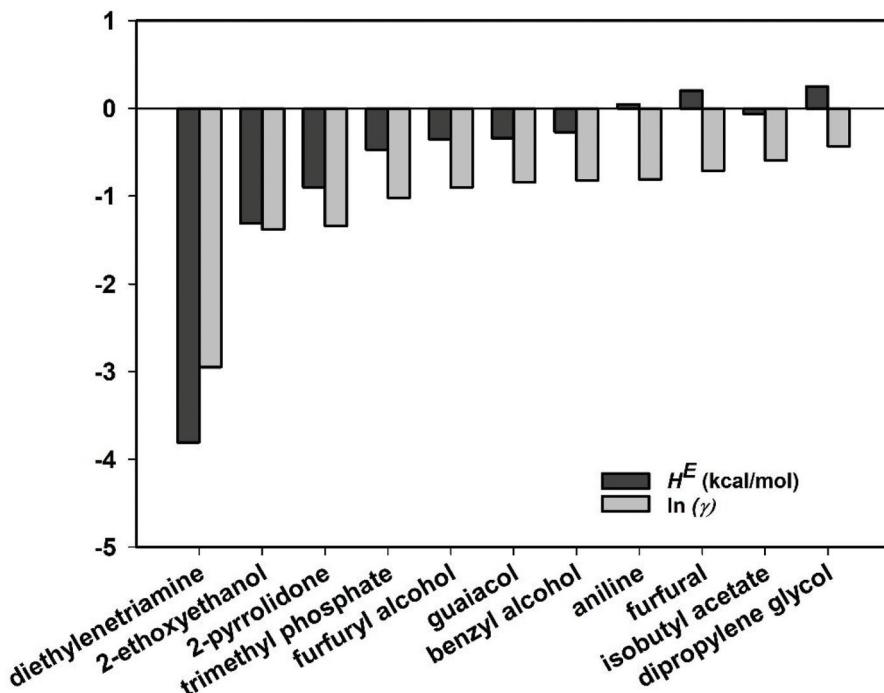
of interactions between dissimilar species (*i.e.*, lignin–solvents) in the mixture. While the  $\ln(\gamma)$  values are often used as a quantitative descriptor for the dissolution power of a solvent. In the literature,  $\ln(\gamma)$  has been reported as the dominating parameter in deciding the capability of a solvent and has also been successfully employed in previous studies to predict the solubility of cellulose in ILs.<sup>29,39,40</sup> Studies have reported that both  $H^E$  and  $\ln(\gamma)$  parameters are good indicators of cellulose and lignin solubility in a solvent.<sup>29,30,41</sup> Therefore, both  $H^E$  and  $\ln(\gamma)$  parameters were calculated for the model grass lignin in the same set of solvents screened by HSP to determine if they can be used to accurately predict lignin dissolution (Fig. 2). The solvent diethylenetriamine was determined to possess significantly lower  $H^E$  and  $\ln(\gamma)$  values (*i.e.*, more negative) than



**Table 1** Hansen solubility parameters and RED values for lignin and the investigated molecular solvents calculated based on the COSMOquick correlated with the experimental lignin removal

Lignin/solvents	Lignin removal (%)	Hansen solubility parameters				RED <sup>a</sup>	RED <sup>b</sup>
		$\delta_D$	$\delta_P$	$\delta_H$	$\delta_T$		
Lignin <sup>a</sup>	—	16.7	13.7	11.7	24.57	—	—
Diethylenetriamine	79.20	16.7	13.3	14.3	25.70	0.192	0.785
Trimethylphosphate	30.32	16.7	15.9	10.2	25.21	0.194	0.913
2-Pyrrolidone	36.28	18.2	12.0	9.0	23.58	0.320	0.599
2-Ethoxyethanol	28.51	16.2	9.2	14.3	23.49	0.386	0.925
Dipropylene glycol	24.07	16.5	10.6	17.7	26.42	0.494	0.831
Furfuryl alcohol	22.37	17.4	7.6	15.1	24.26	0.520	0.821
Guaiacol	26.68	18.0	7.0	12.0	22.74	0.525	0.761
Furfural	12.72	18.6	14.9	5.1	24.37	0.563	0.989
Benzyl alcohol	19.82	18.4	6.3	13.7	23.79	0.612	0.800
Aniline	16.75	20.1	5.8	11.2	23.73	0.762	0.897
Isobutyl acetate	18.29	15.1	3.7	6.3	16.77	0.862	1.470

<sup>a</sup> Taken from Thielemans and Wool.<sup>36</sup> <sup>b</sup> Hansen and Björkman<sup>26,34</sup> reported RED values for lignin.



**Fig. 2** COSMO-RS predicted excess enthalpy and logarithmic activity coefficients of lignin in molecular solvents.

the other solvents, including trimethylphosphate, which had a similar RED value as diethylenetriamine and was therefore predicted to be a good lignin solvent. The COSMO-RS predicted results are much more consistent with the experimental lignin removal than the HSP RED values (Table 1), suggesting the use of COSMO-RS to predict  $H^E$  and  $\ln(\gamma)$  parameters of lignin in solvents is a more realistic method to determine a solvents' ability to extract lignin from lignocellulose.

The goal of lignin extraction from lignocellulose is to increase the efficiency of enzymatic digestion of the plant polysaccharides. Therefore, enzymatic hydrolysis of sorghum pretreated with these solvents was performed using commercial

enzyme cocktails (Fig. 3). The pretreated sorghum was first washed to remove the solvent to prevent interference with enzymatic digestion and the sugar yields were calculated (eqn (5)) based on the recovered solids (Fig. S1 and S2†). These results indicate that there is a direct correlation between low  $H^E$  and  $\ln(\gamma)$  parameters of lignin in molecular solvents and saccharification efficiency. Diethylenetriamine had the lowest  $H^E$  and  $\ln(\gamma)$  parameters and promoted the highest lignin removal and highest glucose and xylose yields of 72.6% and 78.6%, respectively (Fig. 3). All other solvents investigated were unable to extract significant quantities of lignin ( $\leq 36\%$ ) and had low sugar yields ( $\leq 18\%$  glucose and  $\leq 13\%$  xylose), indicat-



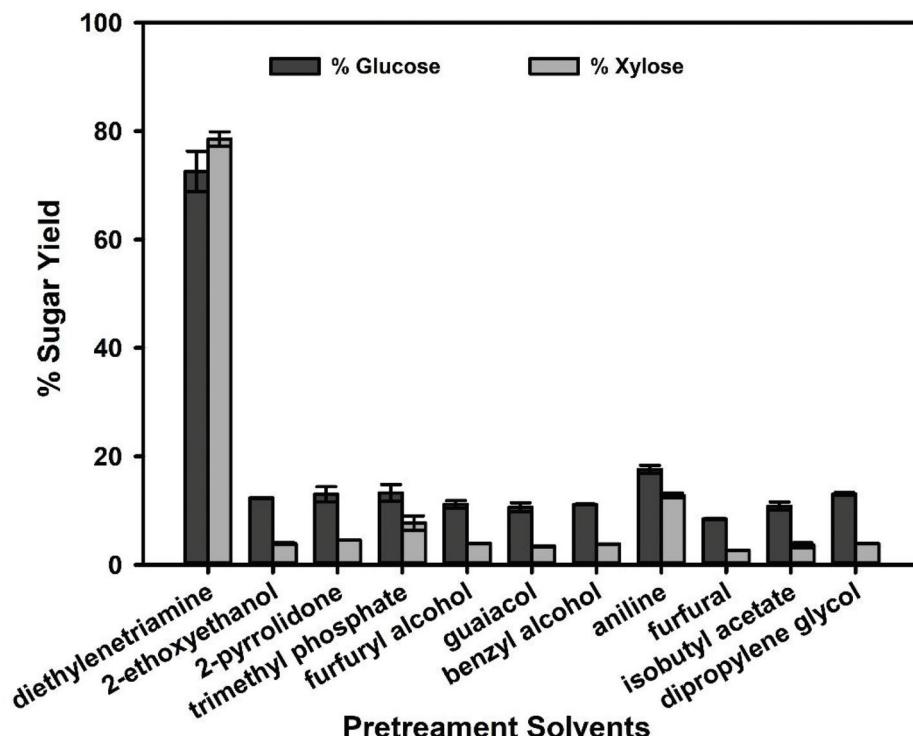


Fig. 3 Glucose and xylose yields of pretreated sorghum after enzymatic hydrolysis.

ing that amine solvents pretreat biomass more effectively than the other functional group categories investigated. The direct correlation between lignin removal efficacy during pretreatment and saccharification efficiency has been observed in many other studies in the literature.<sup>17,18,42</sup> Overall, these data suggest that a general rule can be postulated that solvents that enable high lignin solubility and subsequent plant polysaccharide digestibility will have a  $H^E$  value for lignin of  $\leq -1.5$ .

To better understand the experimental observations, sigma ( $\sigma$ )-potentials of the isolated molecules (solvents and lignin) were predicted using COSMO-RS. The  $\sigma$ -potential is a measure of the affinity of the system to a surface of polarity  $\sigma$ , which provides insights into a solvent's interactions with itself and with lignin. The  $\sigma$ -potential is divided into three regions: H-bond acceptor ( $\sigma > +0.01 \text{ e } \text{\AA}^{-2}$ ), H-bond donor ( $\sigma < -0.01 \text{ e } \text{\AA}^{-2}$ ), and non-polar ( $-0.01 \text{ e } \text{\AA}^{-2} < \sigma < +0.01 \text{ e } \text{\AA}^{-2}$ ) regions. Fig. S3a† depicts the  $\sigma$ -potentials of lignin and molecular solvents. On the negative side of screening charge density (SCD:  $\sigma > -0.01 \text{ e } \text{\AA}^{-2}$ ), the  $\sigma$ -potential ( $\mu(\sigma)$ ) value of diethylenetriamine is more negative than the other solvents, which implies that diethylenetriamine has more affinity to interact with the H-bond donor surfaces (blue color in Fig. S3b†) and has higher H-bond basicity, both of which would promote greater lignin solubility. In contrast, the  $\mu(\sigma)$  value is positive in the region of large positive screening charge density values ( $\sigma > +0.01 \text{ e } \text{\AA}^{-2}$ ), which reflects diethylenetriamine's lack of H-bond donor surfaces (Fig. S3b†). Thus, the intramolecular interaction in diethylenetriamine is very weak, which enables the high interacting strength with the lignin. These results

indicate that diethylenetriamine and potentially other amine-based solvents have an excellent ability to dissolve lignin from lignocellulosic biomass.

## 2.2. Development of COSMO-RS-based predictive model for lignin solubility in amines

The initial solvent screen identified diethylenetriamine as an effective pretreatment solvent. This prompted a more in-depth analysis of amine-based solvents, including a broad class of amines with one or more amine functional groups. A diverse set of amines was selected to help develop a deeper understanding of the key structural features in the amines that contribute to lignin extraction and to enable the development of an effective predictive model for amine-based dissolution of lignin. The new set of solvents was selected that vary in the number, type, and position of amine groups, as well as the carbon chain length of the molecule. The amines selected include 1,3-diaminopropane, 1,4-diaminobutane, 1,5-diaminopentane, 1,2-diaminopropane, spermidine, spermine, 2,2-dimethyl-1,3-propanediamine, ethylenediamine, and pentylamine. They were compared to diethylenetriamine as the baseline (see Fig. 4).

Fig. 6 shows the COSMO-RS predicted  $H^E$  and  $\ln(\gamma)$  of lignin in these amine containing solvents. In the previous section, we established a general rule for lignin solubility and biomass digestibility as  $H^E$  value  $\leq -1.5$ . Since all the selected amines have values below this cutoff, there is a strong indication that they will all be effective solvents for lignin extraction, except possibly 2,2-dimethyl-1,3-propanediamine with a borderline



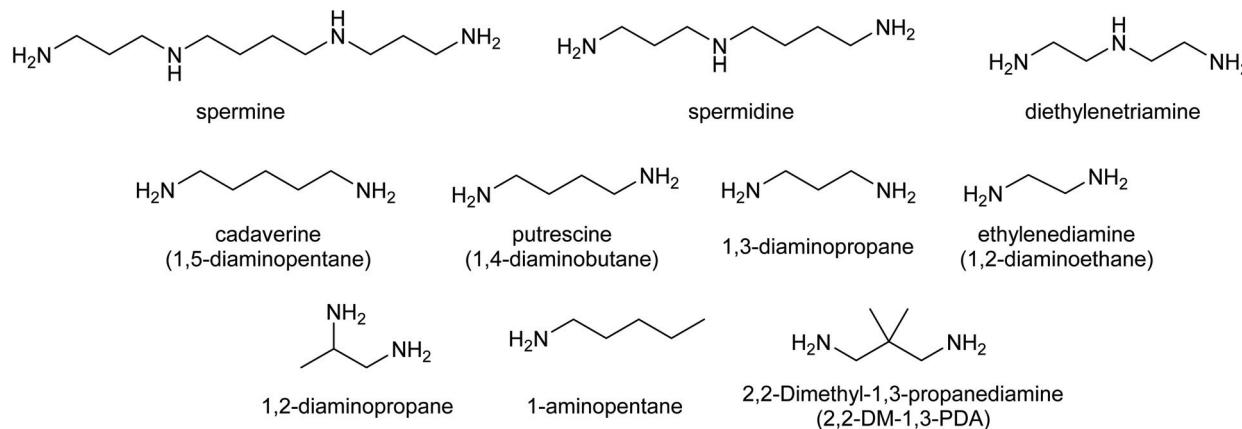


Fig. 4 Chemical structures of the amines evaluated in this study.

$H^E$  value of  $-1.57$  (Fig. 5). Spermine and spermidine have the lowest  $H^E$  and  $\ln(\gamma)$ , indicating that they may be able to extract and solubilize the greatest amount of lignin. As the number of carbon and amine groups increases, the  $H^E$  and  $\ln(\gamma)$  are predicted to be more negative. The polyamines (compounds with  $>2$  amine groups) have  $H^E \leq -3.8$  and are predicted to have the highest lignin solubility capacity, while the diamines, are expected to have an intermediate lignin dissolution ability  $-3.8 \leq H^E \leq -2.5$ . Finally, the branched diamine (2,2-dimethyl-1,3-propanediamine) and monoamine (1-aminopentane) with  $-2.5 < H^E < -1.5$  are expected to have the lowest lignin extraction capacity.

To confirm these predictions, biomass pretreatment experiments using these solvents were performed in a similar manner as the prior round of screening. Overall, the results indicate that all the poly and diamines were effective solvents. However, an examination of the experimental averages of lignin removal suggest that the polyamines do not actually have the highest lignin extraction capacity (66.1%–79.2%), but rather the diamines (74.1%–85.8%) (Fig. 6). This is not com-

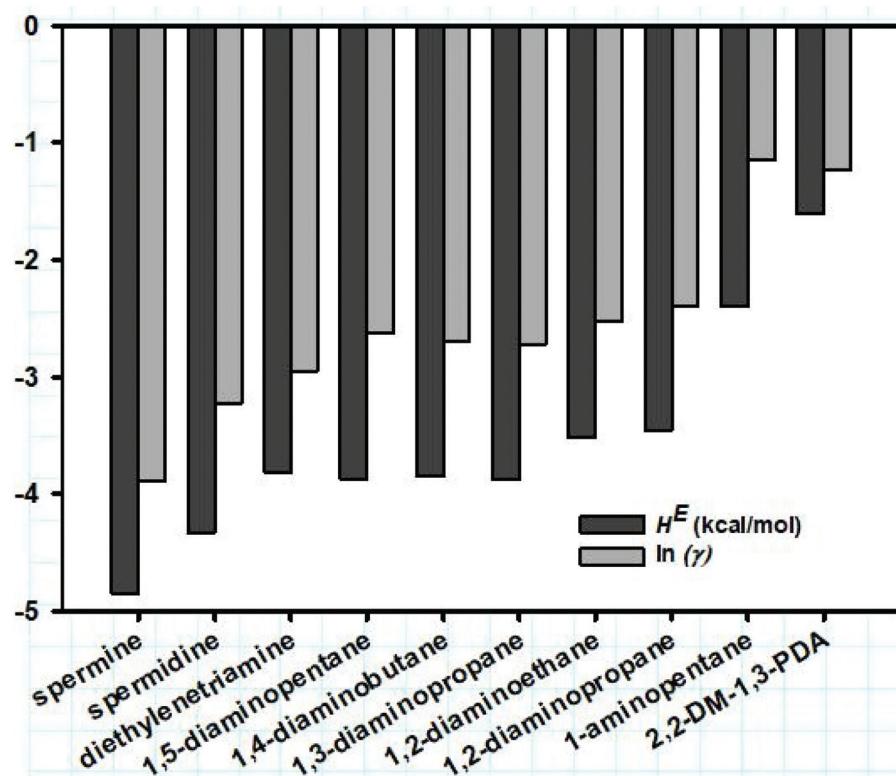
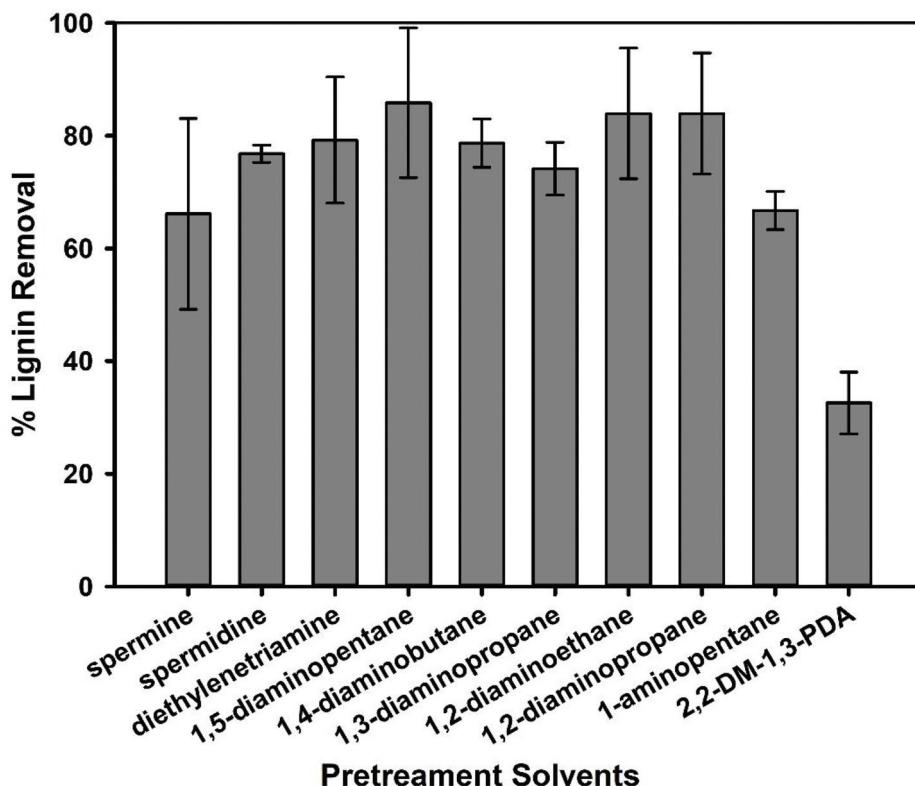


Fig. 5 COSMO-RS predicted excess enthalpy and logarithmic activity coefficients of lignin.





**Fig. 6** Solubility of lignin measured after biomass pretreatment in amines. Pretreatment conditions: 20% solids loading, 140 °C, and 3 h of reaction time.

pletely unexpected as both spermine and spermidine have a higher viscosity ( $\eta$ ) than the diamines (Table S1†). Increases in a solvent's viscosity can limit effective mass transfer, which is known to have a negative effect on a pretreatment solvent's dissolution power.<sup>43–46</sup> Therefore, solvents with low  $H^E$  values should also be cross checked for high viscosity when screening. In addition to viscosity, the basicity and polarity of the solvent is another indicator of their ability to dissolve lignin. For example, 2,2-dimethyl-1,3-propanediamine has a lower viscosity than spermine and spermidine but only extracted a relatively low amount of lignin (32.6%). This is likely due to the lower polarity and basicity of this solvent (see Fig. S4†). Therefore, viscosity, polarity and basicity are important parameters to consider when attempting to predict dissolution. Since viscosities do have a notable impact, we demonstrated that we could predict the amine solvent viscosities using the COSMO-RS model and validated the predictions using the available experimental viscosity data for diethylenetriamine (Fig. S5†).<sup>47</sup> This means COSMO-RS can be used to predict  $H^E$ ,  $\ln(\gamma)$ , and  $\eta$  to facilitate identification of good lignin solvents.

Using these COSMO-RS-predicted quantities, three different lignin solubility prediction models (linear and non-linear; eqn (1)–(3)) were developed for the amines and then validated. To develop a predictive model, the parameters excess enthalpy (correlated to the interactions), activity coefficient (related to the dissolution capability), viscosity (associated with the mass transfer rate), and the dissociation constant ( $pK_a$ ) related to the

strength of acid/base were considered. The following equations were developed to predict the dissolution of lignin, but excluded pentylamine, which was used to validate the models.

Non-linear model 1:

$$\text{Lignin sol.}(\%) = b_0 + (b_1 \times \exp(H^E)) + \left( \frac{b_2}{\ln(\gamma)} \right) + (b_3 \times \eta) + (b_4 \times pK_a) \quad (1)$$

$b_0 = 210.65$ ;  $b_1 = -34.41$ ;  $b_2 = 92.46$ ;  
 $b_3 = -3.72$ ; and  $b_4 = -9$

Non-linear model 2:

$$\text{Lignin sol.}(\%) = b_0 + (b_1 \times \exp(H^E)) + \left( \frac{b_2}{\ln(\gamma)} \right) + (b_3 \times \eta) \quad (2)$$

$b_0 = 130.33$ ;  $b_1 = 41.31$ ;  
 $b_2 = 124.76$ ; and  $b_3 = -4.96$

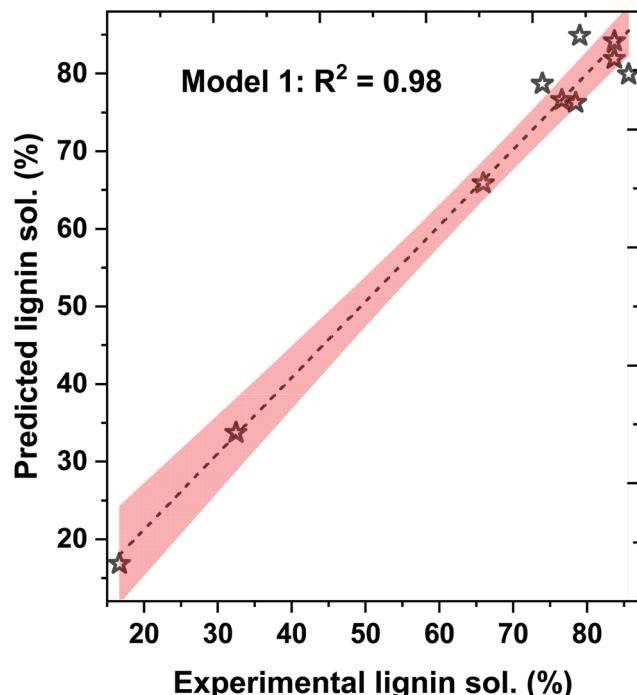
Linear model 3:

$$\text{Lignin sol.}(\%) = b_0 + (b_1 \times H^E) + (b_2 \times \ln(\gamma)) + (b_3 \times \eta) + (b_4 \times pK_a) \quad (3)$$

$b_0 = 89.24$ ;  $b_1 = -45.63$ ;  $b_2 = 34.23$ ;  
 $b_3 = -2.55$ ; and  $b_4 = -8.46$

Here, the  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_4$  are the fit coefficients (*i.e.*, constants). Experimental and predicted lignin solubility for





**Fig. 7** Experimental data and COSMO-RS-based model 1 predicted lignin solubility for amines with 95% confidence error band.

amines based on these models are shown in Fig. 7 (model 1) and Fig. S6<sup>†</sup> (models 2 and 3). To assess the potential performance of the developed model equations, they were used to predict the solubility of lignin for pentyamine, which was excluded from the original training set. The predicted lignin solubility for models 1, 2, and 3 is 60.2%, 58.9%, and 53.1%, respectively while the experimental lignin solubility in pentyamine is  $66.7 \pm 3.4\%$ . Further, the predictive models were also evaluated with lignin solubility data from the literature where *Miscanthus* biomass was pretreated with ethylenediamine at higher temperature (180 °C).<sup>48</sup> The experimental lignin solubility was reported as  $\sim 71 \pm 4\%$ , while the predicted solubility for models 1, 2, and 3 is 65.1%, 53.5%, and 96.1%, respectively. These data indicate that the non-linear model (1) predicts the lignin solubility more accurately than the linear models. This reveals that the relationship between the solvent type and lignin dissolution capacity (within the realm of biomass pretreatment) is not a simple linear relationship. When factors such as mass transfer and chemical reactivity are coupled, non-linear relationships have been more suitable at describing the experimental results. However, these developed lignin solubilities non-linear model could be applicable for amines only when  $H^E \leq 0.2$  and  $\ln(\gamma) \leq -0.75$ .

### 2.3. Pretreatment with amines and enzymatic saccharification of residual biomass

As mentioned earlier, high lignin extraction has been directly correlated with an increase in enzymatic saccharification efficiency of pretreated lignocellulose. Therefore, we tested whether this correlation also holds for biomass pretreated

with the new set of polyamines. Typically, efforts to extract lignin from biomass are not completely selective and some portions of polysaccharides are also extracted, so it is important to measure the solid recovery and composition post-pretreatment to fully understand the extent of this non-selective extraction and its impact on biomass deconstruction. For the set of amines studied, the solid recovery after pretreatment ranged from 61.5%–68.1% (Fig. S7<sup>†</sup>). The composition of the residual biomass was relatively similar across the different pretreatments, with an average loss of  $27.1 \pm 2.3\%$  cellulose and  $30.2 \pm 1.9\%$  hemicellulose (xylan and arabinan) during pretreatment (Fig. S7 and 8<sup>†</sup>). In many cases, lignin extraction is accompanied by some extent of hemicellulose/cellulose removal, so these results are not surprising. In an actual biorefinery, these extracted polysaccharides would be returned to the saccharification reaction after solvent recovery. Nevertheless, all amines investigated in this study afforded  $>70\%$  glucose yields, with the highest glucose yield of  $\sim 90\%$  (Fig. 8). The glucose yields with these amines could be organized in the following order: 1,2-diaminopropane > 1,5-diaminopentane > 2,2-dimethyl-1,3-propanediamine > ethylenediamine > spermidine > 1,4-diaminobutane > pentyamine > 1,3-diaminopropane > diethylenetriamine > spermine. Similarly,  $>65\%$  xylose yields were attained with these amines, with the highest xylose yield of  $\sim 92\%$  obtained also for 1,2-diaminopropane. The xylose yields followed the similar order as of glucose (Fig. 8).

The efficiency of sugar release for most amines was as expected based on the extent of lignin removal. Interestingly, 2,2-dimethyl-1,3-propanediamine extracted relatively low amounts of lignin (32.6%) but still permitted high sugar yields. This was very unexpected and does not fit well with previously reported studies of pretreatment solvents that selectively extract lignin. However, while lignin extraction can drive increases in saccharification efficiency, it is not the only factor that influences the release of sugars from biomass. Lignocellulose is complicated, and there are many possible outcomes of solvent-based pretreatment that can impact enzymatic sugar release, such polysaccharide extraction, or modification of macrostructure of biomass to increase the accessible surface area for enzymatic hydrolysis, *etc*. This is an interesting observation and suggests that amines are potentially acting to reduce the recalcitrance to enzymatic digestion by other mechanisms than lignin dissolution. Since this study is focused on lignin extraction, those other possibilities will be a focus area of future studies.

### 2.4. Cellulose structure and allomorphs

To study the structural changes in the polymers that occur during pretreatment, PXRD was utilized to determine the polymorph of cellulose and the proportions of crystalline and non-crystalline components in the biomass sample. The untreated sorghum displays diffraction patterns characteristic to the cellulose I polymorph, which is typical for native cellulose that is found in untreated lignocellulosic biomass<sup>49–51</sup> (Fig. 9 and S9<sup>†</sup>) with a crystallinity content of 83.3%. Note: This is higher



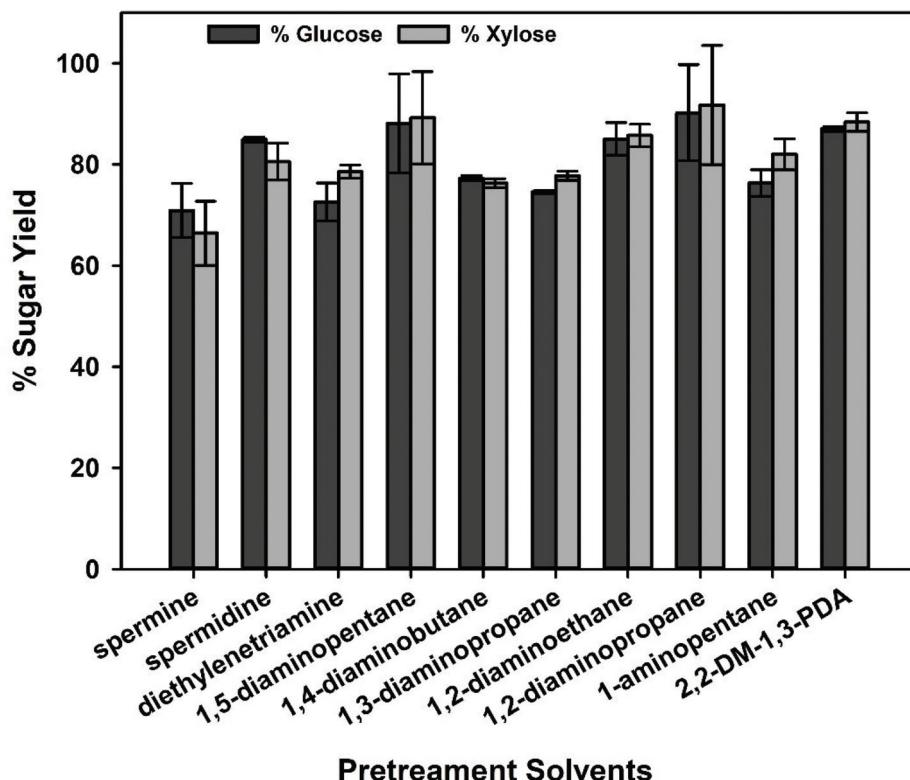


Fig. 8 Glucose and xylose yields after enzymatic hydrolysis of pretreated sorghum with amines.

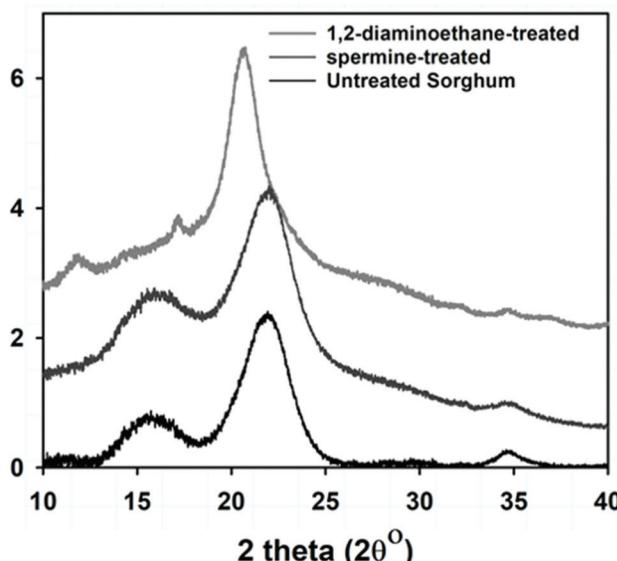
than that previously reported for sorghum (~50–55%)<sup>52</sup> due to the different radiation source that is used in these measurements (cobalt *vs.* copper). Nevertheless, we can compare the results qualitatively to determine the impact of pretreatment on the biomass. When 1,2-diaminoethane was utilized to pretreat the biomass, a change in the polymorph was observed. The identified peaks in the diffractogram match what has been previously reported as cellulose III<sup>49,50,53</sup> and depicted a lower crystallinity index (70.3%) (Fig. 9 and S10†). This observation is in line with previous reports associated with amine pretreatment. Qin *et al.* showed that using 1,2-diaminoethane (also known as ethylenediamine) resulted in biomass modification to cellulose III especially when high temperatures are used followed by soaking and washing in ethanol (unlike water).<sup>51</sup> On the other hand, the biomass recovered after pretreatment with the polyamine-spermine displayed mixed polymorphism. The diffractogram shows peaks analogous to both cellulose I and III, although most of the crystalline peaks are from cellulose I polymorph (Fig. 9 and S11†). Nevertheless, the recovered biomass has the highest amorphous character with a crystallinity index of 65.7%.

The recovered biomass (after pretreatment) was subjected to saccharification and yielded an average sugar (glucose/xylose) of 68.6% and 85.4% for spermine and ethylenediamine, respectively. Although ethylenediamine pretreated biomass had a slightly higher crystallinity index, it permitted higher sugar yields. This indicates that while the level of cry-

tallinity is known to impact biomass digestibility, sugar yields are impacted by a combination of factors, such as delignification. However, in conjunction with polymorph transformation, amine pretreatment can significantly reduce the crystallinity index, which can positively impact the cellulose digestibility. In their previous work, researchers reveal that ethylenediamine molecules penetrate the hydrophilic edges of the stacked sheets and enlarge cellulose III volume in the (010) direction.<sup>51</sup> These changes have been reported to increase the enzymatic saccharification rate by 5 times,<sup>54</sup> while other studies found that initial rates of digestion were strongly correlated with amorphous content, not the allomorph type.<sup>55,56</sup>

While celluloses having a higher amorphous content are typically easier to enzymatically digest, the accessibility of the plant cell-wall to the various glycoside hydrolases is also a very important factor in determining hydrolysis rate. While enzyme accessibility could be affected by crystallinity, it is also known to be affected by the lignin and hemicellulose contents/distribution, the particle size, and the porosity of the biomass. Since the lignin and hemicellulose removal for ethylenediamine were higher (83.9% and 32.5%) than that of spermine (66.1% and 29.7%), the greater sugar yields from the ethylenediamine pretreated biomass appear to also be influenced by their enhanced removal, potentially providing the enzymes greater access to the polysaccharides. It is well known that lignin plays a more important role than cellulose crystallinity on the digestibility of lignocellulose,<sup>57</sup> and both the chemistry





**Fig. 9** X-ray diffraction profiles for untreated and treated sorghum including the relative percentage of each polymorph and crystallinity index (\*measured by method of Segal *et al.*).<sup>59</sup>

	I* (%)	III* (%)	CI** (%)
<b>Untreated</b>	61.8	0	83.3
<b>1,2-diaminoethane-treated</b>	7.3	67.7	70.3
<b>spermine-treated</b>	47.6	19.5	65.7

and physical barrier lignin provides can reduce the level of enzymatic hydrolysis.<sup>58</sup> Therefore, lignin solubilization and removal remains the main focus area for optimization in this study.

## 2.5. Understanding the mechanistic behavior of lignin dissolution

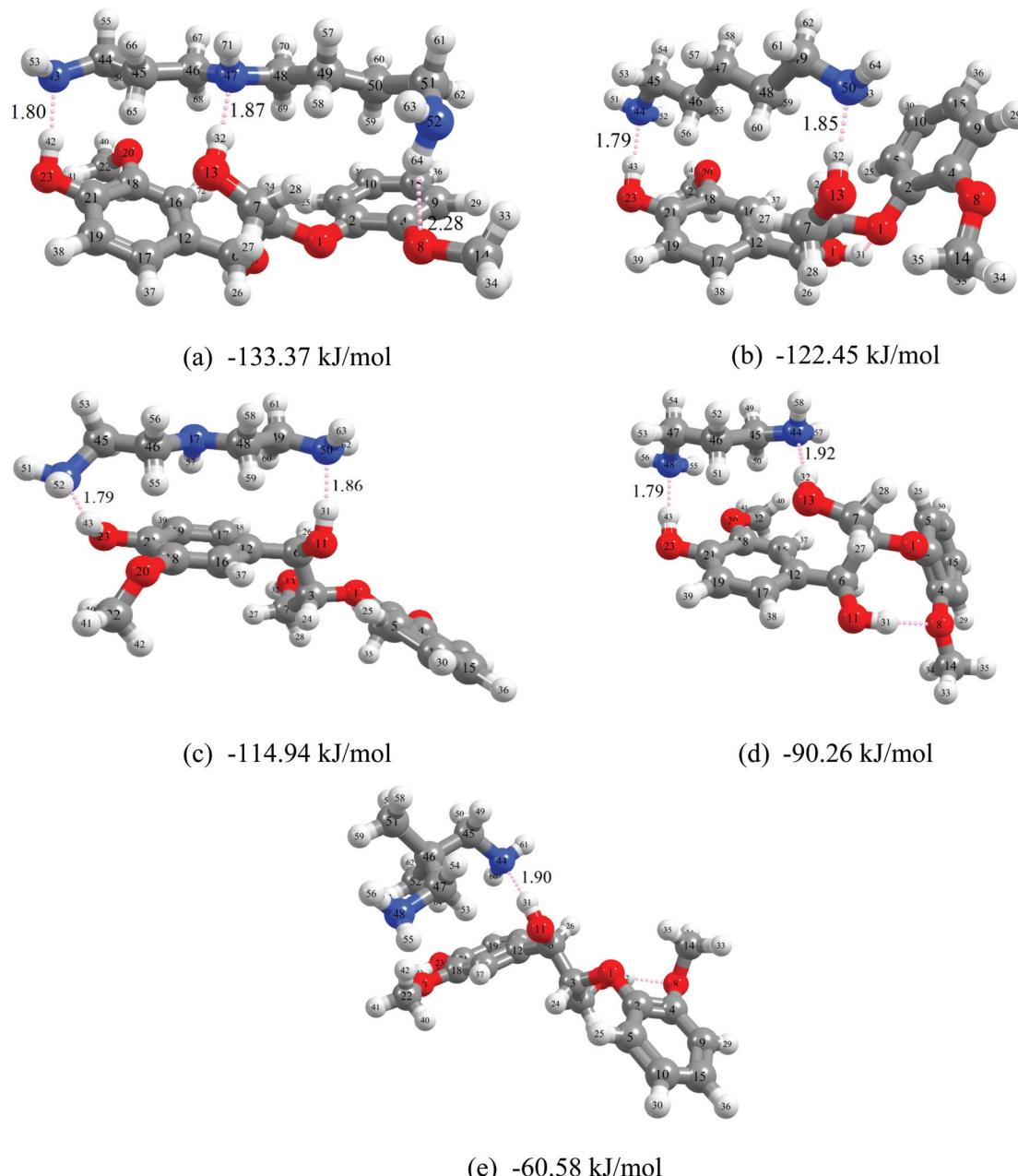
To better understand the dissolution mechanism of lignin in the molecular solvents, quantum chemical (QC) simulations were performed. QC calculation is a prevailing computational technique for investigating the underlying molecular interactions of solvent/solute systems. These calculations can be used to investigate different types of interactions between solvents and solutes and help researchers to understand which interactions drive a solvent's ability to dissolve a solute. It is computationally difficult to conduct QC calculations with large macromolecules such as polymeric type lignin structure, so to simplify our analysis, we chose a representative lignin dimer, guaiacyl glycerol- $\beta$ -guaiacyl ether (GGE) as a model molecule since it has the most common  $\beta$ -O-4 monomer-monomer linkage found in lignin. These complex solvent/solute systems can adopt many different conformations, which complicates analysis, so the most stable energy conformer structures of lignin GGE-molecular solvents were obtained and used to help understand several aspects of their interactions in more detail. According to Boltzmann distribution, the conformers with the lowest energy represent the maximum proportion of distribution. The lowest energy conformers are used to construct the initial structures for the geometry optimization between lignin-amines and lignin-organic solvents. All the lignin-amine/organic solvent conformers with relative energies are provided in the ESI (Fig. S12–S20 and Tables S2–S10†) and the most stable conformers are depicted in Fig. 10 and Fig. S21.†

All the optimized conformers show that the ideal interaction site for the lignin-amine interaction is the  $\alpha$ -OH,  $\gamma$ -OH, and phenol-OH group of lignin. In addition, interaction energies, non-covalent interactions, critical bonding topological characteristics, and an assessment of the strength of the electron donor-acceptor orbital interactions have also been performed for only the most stable lowest energy conformers.

**2.5.1. Optimized geometries and interaction energies of lignin-amines/organic solvents.** We first set out to determine the optimized geometries of lignin GGE dissolved in various molecular solvents and assess their interaction energies, which will bring insights into how strongly the solvents interact with the lignin through H-bonding and how that might impact lignin dissolution (Fig. 10 and Fig. S21†). From these figures, it appears that the amines (except 2,2-dimethyl-1,3-propanediamine) form stronger H-bonds O-H $\cdots$ N (1.79 Å–1.92 Å) with lignin than the non-amine organic solvents (2-ethoxy ethanol, furfuryl alcohol, isobutyl acetate, and benzyl alcohol) O-H $\cdots$ O (1.88 Å–2.02 Å), thereby resulting in higher interaction energies between amines and lignin. Moreover, amines form multiple hydrogen bonds with the lignin molecule. Therefore, it appears that the intermolecular H-bonds between the amines and lignin may be a vital factor in promoting the higher solubility of lignin in these solvents. A comparison of the interaction energies of amines with lignin indicates that spermidine has the highest interaction energies, followed by 1,5-diaminopentane, diethylenetriamine, and 1,3-diaminopropane.

On a fundamental level, the total interaction energy of a solvent with lignin is decomposed into four chemically meaningful contributors: electrostatic, exchange-repulsion, induction, and dispersion (see Fig. 11). The electrostatic energy corresponds to the classic electrostatic interaction between the promoted fragments as they are brought into their positions in



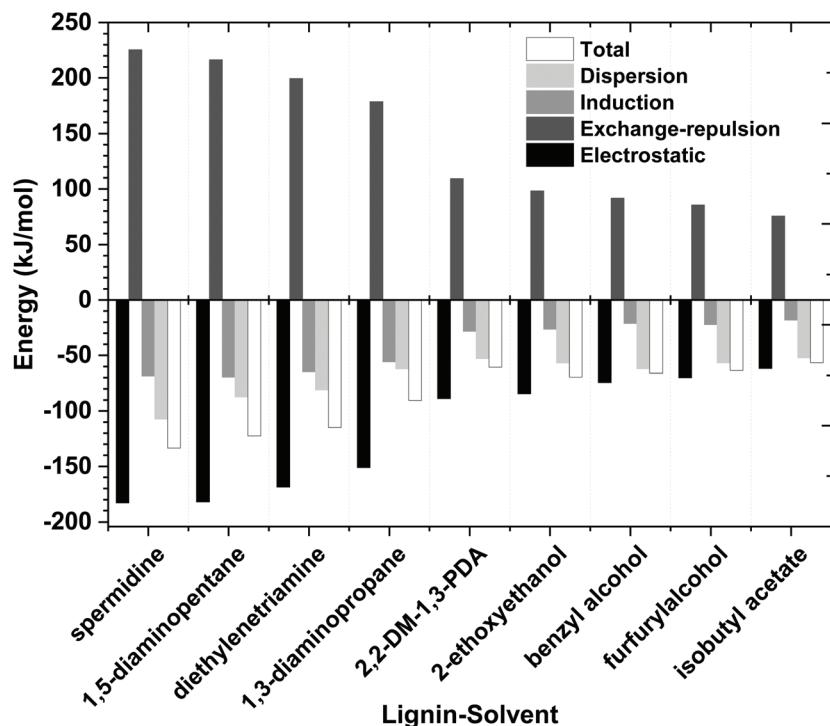


**Fig. 10** Optimized geometries for lignin GGE–amines (a) spermidine, (b) 1,5-diaminopentane, (c) diethylenetriamine, (d) 1,3-diaminopropane, and (e) 2,2-dimethyl-1,3-propanediamine. The H-bonds are indicated by dotted lines, the bond lengths are in Angstrom ( $\text{\AA}$ ) and given with corresponding atom numbers. The color scheme used for different atoms is C (gray), O (red), N (blue), and H (white), respectively.

the final complexes, the term exchange repulsion accounts for Pauli repulsion between closed-shell fragments and is perpetually positive. The induction term, sometimes referred to as the orbital interaction or the polarization energy, arises from the orbital relaxation and the orbital mixing between the fragments (charge transfer). Dispersion energy which represents the amount of energy required to promote the fragments from their equilibrium geometry to the structure they will take up in the combined molecule. The stronger the interaction energy (more negative magnitude) between lignin and molecular sol-

vents, the higher the anticipated lignin dissolution capacity. Analysis of these four contributors illustrates that the electrostatic interaction is the dominating attractive component between lignin and the amine, while dispersion and induction energies play a minor role in stabilizing the lignin–amine complexes. In the case of non-amine organic solvent systems, dispersion interactions are almost equal to the electrostatic interactions. Also, the induction interactions are significant in lignin–amine complexes relative to those calculated for the non-amine organic solvent complexes. This higher induction





**Fig. 11** Functional group intramolecular symmetry adapted perturbation theory (FI-SAPT) decomposition of the non-bonded interaction energies between lignin and molecular solvents.

energy in lignin–amine complexes indicates that a substantial charge transfer occurred between lignin and amines. In terms of magnitude, all the energy components are greater in the lignin–amine complexes than the lignin–organic solvent complexes. The order of attractive interactions in lignin–amine complexes is electrostatic > dispersion > induction. Overall, the QC calculated interaction energies are in good agreement with the COSMO-RS predicted interactions and lignin solubility. The QC and COSMO-RS results can be used to suggest the solubility of lignin is lower in these types of organic solvents due to the weaker interactions.

**2.5.2. Reduced density gradient (RDG) analysis of lignin–amines/organic solvents.** RDG analysis was carried out to examine the strength of the non-covalent interactions (NCIs) between the lignin and amines/organic solvents. This method can be used to visualize the different interaction energy contributions, such as hydrogen bond, van der Waals, and steric repulsion. NCIs are assessed using the normalized and dimensionless reduced density gradient (eqn (4))

$$\text{RDG} = \frac{1}{\left[2(3\pi^2)^{\frac{1}{3}}\right]} \frac{|\nabla\rho(r)|}{\rho(r)^{\frac{4}{3}}} \quad (4)$$

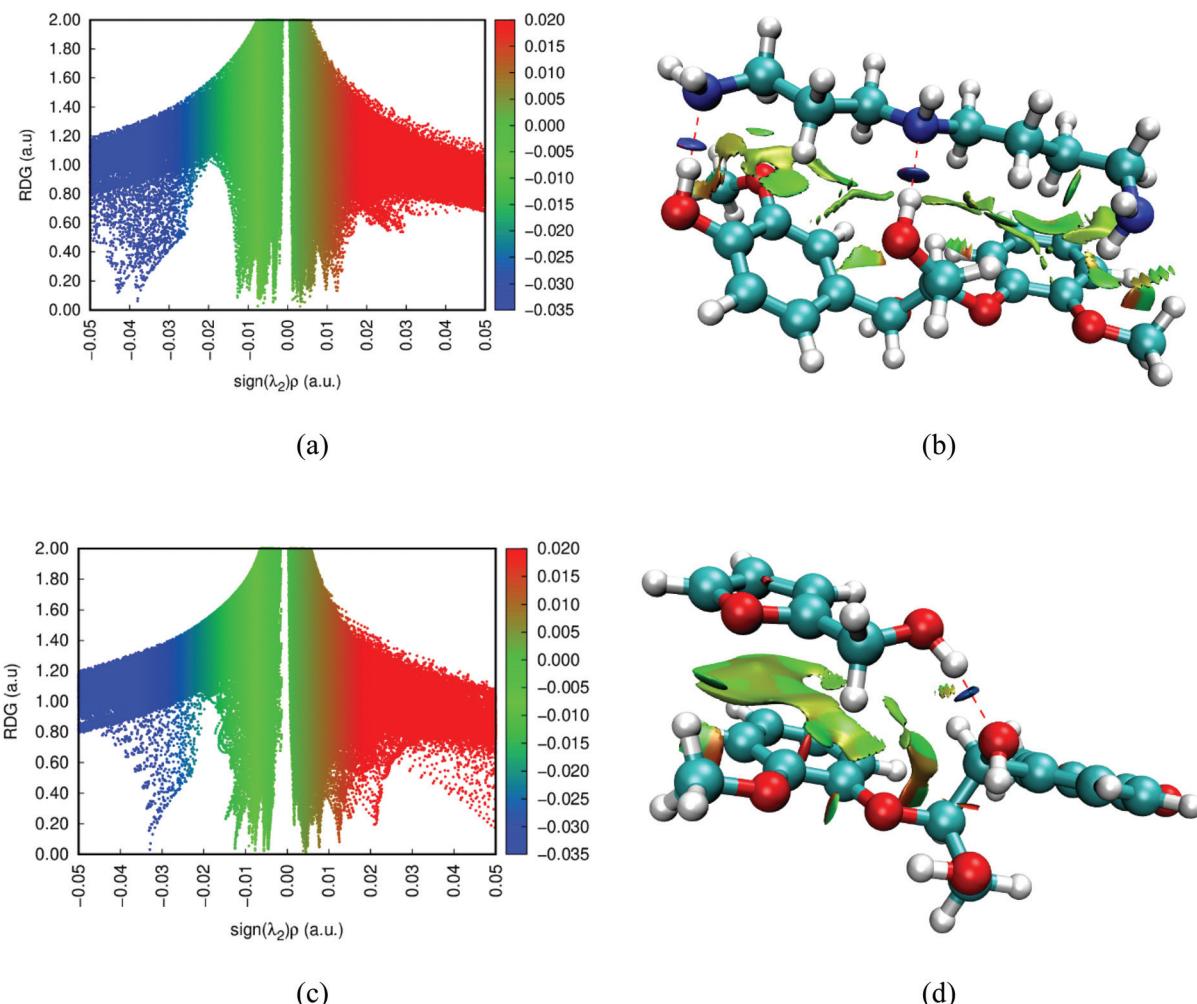
Regions where the RDG and electron densities are low representing the non-covalent interactions. Therefore, the iso-surface of RDG at lower electron densities was used to visualize the position and nature of NCIs in 3D space. This is done by plotting the RDG *vs.* sign of second Hessian eigenvalue ( $\lambda_2$ )

multiplied with the electron density ( $\rho(r)$ ) ( $\text{sign}(\lambda_2) \cdot \rho(r)$ ) in a scatter plot.

Two representative solvents, spermidine (amine) and furfuryl alcohol (non-amine), were selected to explore this analysis in detail (Fig. 12a and c) and plots of the remaining solvents are provided in Fig. S22.† Scanning across  $\text{sign}(\lambda_2) \cdot \rho(r)$  from positive to negative values, there are several spikes in RDG scatter plot that correspond to the steric repulsion (red color), van der Waal (green color) interaction, and hydrogen bonding (blue color). In Fig. 12b and d, the interactions are visualized, and colored surfaces correspond to the respective colors in the respective NCI scatter plots. Examination of the NCI plots show that the amine-based solvents have spikes in the negative region of  $\text{sign}(\lambda_2) \cdot \rho(r)$  that are more negative (O–H…N:  $-0.032 < \text{sign}(\lambda_2) \cdot \rho(r) < -0.044$ ) than the non-amine organic solvents (O–H…O:  $\text{sign}(\lambda_2) \cdot \rho(r) > -0.03$ ), which indicates that the strength of the H-bond interactions (blue region) is much stronger between lignin and amines. On the other hand, in the attractive region, multiple spikes are observed for lignin–amine interactions which are consistent with the geometrical analysis. Also, in the lignin–amine systems, the major steric repulsions (red color) occurred within the lignin molecule while it occurred between the lignin and solvent in the lignin–organic solvents, weakening their interaction and potentially explaining the lower lignin solubility observed in non-amine solvents (Fig. 12b and d).

**2.5.3. QTAIM analysis of lignin–amines/organic solvents.** To gain deeper insights into the intermolecular interactions





**Fig. 12** RDG scatter and NCI plots (isovalue 0.5 a.u.) of (a) lignin-spermidine (RDG), (b) lignin-spermidine (NCI), (c) lignin-furfuryl alcohol (RDG), and (d) lignin-furfuryl alcohol (NCI). The RDG/NCI plots are colored on a blue-green-red scale according to values of  $\text{sign}(\lambda_2) \rho(r)$ , ranging from  $-0.045$  to  $0.025$ . Blue indicates strong attractive interactions, green indicates the vdW interaction, and red indicated steric repulsions. The color scheme used for different atoms is C (cyan), O (red), N (blue), and H (white), respectively.

between lignin and amine or organic solvents, we conducted quantum theory of atom in molecule (QTAIM) analysis. QTAIM analysis is used to examine critical bonding topological characteristics, such as electron density  $\rho(r)$ , Laplacian energy density  $\nabla^2\rho(r)$ , energy density  $H_{\text{BCP}}(r)$ , and H-bonding energies ( $E_{\text{HB}}$ ). The interatomic interactions are quantitatively represented by using signs of  $\rho(r) > 0$ ,  $\nabla^2\rho(r) > 0$ , and  $H_{\text{BCP}}(r) < 0$  at bond critical points (BCPs) and classified as closed-shell interactions, which include the H-bonding and van der Waals interactions. The positive sign of  $\rho(r)$  corresponds to the strength of the hydrogen bond, a positive sign of  $\nabla^2\rho(r)$  represents the characteristic of hydrogen bonds (non-covalent type), and  $H_{\text{BCP}}(r)$  corresponds to the nature of the hydrogen bonds (a positive sign corresponds to non-covalent, and a negative sign corresponds to covalent). According to Koch and Popelier, the  $\rho(r)$  and  $\nabla^2\rho(r)$  at the BCP for hydrogen bond interactions must be in the range of  $0.002$ – $0.035$  a.u. and  $0.014$ – $0.139$  a.u., respectively.<sup>60,61</sup> As mentioned above, the  $\rho(r)$

for the vdW region tends to be smaller ( $0.002$ – $0.009$  a.u.) than hydrogen bonding and steric repulsion.

Table 2 reports the  $\rho(r)$ ,  $\nabla^2\rho(r)$ , and  $H_{\text{BCP}}(r)$  of hydrogen bond critical points for the lignin-amine ( $\text{O-H}\cdots\text{N}$ ) and lignin-organic solvent ( $\text{O-H}\cdots\text{O}$ ) systems. For the lignin-amine system, the values of  $\rho(r)$  are in the range of  $0.036$ – $0.044$  a.u., which is higher than the Koch and Popelier proposed range for electron densities for the hydrogen bond. Whereas, in the case of lignin-organic solvents, the values of  $\rho(r)$  and  $\nabla^2\rho(r)$  lies within the Koch and Popelier proposed ranges ( $0.002$ – $0.035$  a.u. and  $0.014$ – $0.139$  a.u.). From these electron densities, the  $\text{O-H}\cdots\text{N}$  bond between lignin and the amine solvent is predicted to be stronger than the  $\text{O-H}\cdots\text{O}$  bond between lignin and non-amine solvents. The Laplacian electron densities at the BCP show positive values for both lignin-amines/organic solvent systems, implying that the characteristics of H-bonding interactions are non-covalent. Further examining energy densities ( $H_{\text{BCP}}(r)$ ), they are negative for O-

**Table 2** Topological electron density (a.u.) properties of H-bonded (H···N & H···O) complexes at bond critical points (BCPs) for the interaction of lignin–amine/organic solvent systems and H-bond energies ( $E_{HB}$ ) calculated at the B3LYP-GD3/6-311+G (d, p) level of theory. Atom notations are given in Fig. 10 and Fig. S21†

Amine/organic solvent	H-Bond	Distance (Å)	$\rho(r)$	$\nabla^2\rho(r)$	$H_{BCP}(r)$	$E_{HB}$ (kJ mol <sup>-1</sup> )
Spermidine	O23–H42···N43	1.80	0.0436	0.0957	-0.0059	-46.84
	O13–H32···N47	1.87	0.0382	0.093	-0.0032	-38.84
1,5-Diaminopentane	N52–H64···O8	2.28	0.0128	0.0435	0.0013	-10.76
	O23–H43···N44	1.79	0.045	0.0957	-0.0067	-48.94
Diethylenetriamine	O13–H32···N50	1.85	0.0393	0.0947	-0.0037	-40.68
	N44–H52···O20	2.38	0.0124	0.0424	0.0011	-10.89
1,3-Diaminopropane	O11–H31···N50	1.86	0.038	0.095	-0.0032	-39.49
	O23–H43···N44	1.79	0.045	0.099	-0.0062	-48.94
2,2-Dimethyl-1,3-propanediamine	O23–H43···N48	1.79	0.0449	0.0994	-0.0065	-49.73
	O13–H32···N44	1.92	0.0334	0.0898	-0.0011	-32.28
2-Ethoxy ethanol	N44–H55···O20	2.37	0.0124	0.0441	0.0014	-10.89
	O11–H31···N44	1.90	0.0347	0.0895	-0.0018	-34.12
Benzyl alcohol	O11–H31···O46	1.88	0.030	0.101	0.0008	-31.10
	O57–H59···O20	2.01	0.020	0.078	0.0023	-19.42
Furfuryl alcohol	O50–H56···O11	1.82	0.033	0.1184	0.0006	-37.13
Isobutyl acetate	O23–H43···O46	1.92	0.0247	0.0953	0.0023	-25.19

$H_{BCP}(r) = G_{BCP}(r) + V_{BCP}(r)$ ;  $E_{HB}$  (kJ mol<sup>-1</sup>) =  $\frac{1}{2}V(r) \times 2624.25$ ;  $G_{BCP}(r)$  is the Lagrangian kinetic energy density;  $V_{BCP}(r)$  is the potential energy density.

H···N and positive for O–H···O at the BCP, indicating the amine–lignin interactions are more covalent in nature while the non-amine solvents have more non-covalent or weak interactions. These calculations ( $\rho(r)$ ,  $\nabla^2\rho(r)$ , and  $H_{BCP}(r)$ ) help in explain why the organic solvents exhibit lower interactions with lignin.

Additionally, hydrogen bonding energies ( $E_{HB}$ ) are also calculated using the potential energy densities  $V_{BCP}(r)$ . The hydrogen bonding energy of O–H···N (lignin–amine) is much stronger than the O–H···O bond energy. It is worthwhile to mention that 2,2-dimethyl-1,3-propanediamine showed similar electronic properties ( $\rho(r)$ ,  $\nabla^2\rho(r)$ , and  $E_{HB}$ ) as the organic solvents, which is consistent with the low solubility of lignin observed with this solvent. In addition to the discussions, there is a strong correlation between Hessian second eigenvalue ( $\lambda_2$ ) and  $E_{HB}$  values (Fig. S23†). The lower the  $\lambda_2$  value, the stronger the hydrogen bonding energy. Overall, the electronic properties clearly indicate that the amine solvents are highly effective solvents for the lignin removal and have much stronger hydrogen bonding energies than the organic solvents.

**2.5.4. Natural bonding orbital (NBO) analysis.** NBO analysis was performed to understand the strength of the electron donor–acceptor orbital interactions between lignin and the solvents. The electron donor–acceptor bond energies are reported in Table 3. The larger the electron donor–acceptor bond energy ( $E^{(2)*}$ ), the stronger the interaction. In the lignin–solvent systems, the lone pair (LP) electrons of the amine (nitrogen atom) or organic solvent (oxygen atom) interact with the anti-bonding orbital (BD\* or  $\sigma^*$ ) of the lignin dimer (hydroxyl proton). These electron donor–acceptor orbitals are responsible for the formation of strong H-bonds between them. In the complex systems, the amine/organic solvent moieties act as an electron donor and lignin moiety is the electron acceptor. Examining the lignin–organic solvents, 2-ethoxyetha-

nol and benzyl alcohol show the strongest electron donor–acceptor ( $LP(O) \rightarrow \sigma^*(O11–H31)$ ) orbital energies. In the case of lignin–amines complexes, the electron donor–acceptor ( $LP(N) \rightarrow \sigma^*(O–H)$ ) orbital energies are two to three-time stronger than the organic solvent systems. These results are consistent with the QTAIM and COSMO-RS predictions.

In addition, the  $\pi$ – $\pi$  and CH– $\pi$  stacking interactions in the lignin–solvent complexes were also examined, where applicable.  $\pi$ – $\pi$  interactions are observed in the benzyl alcohol and furfuryl alcohol lignin complexes whereas CH– $\pi$  stacking interactions were observed in the isobutyl acetate, 2-ethoxyethanol, and amine–lignin complexes. Overall, the strength of CH– $\pi$  interactions was predicted to be relatively stronger than  $\pi$ – $\pi$ . However, compared to  $LP(O) \rightarrow \sigma^*$  orbital energies, the CH– $\pi$  stacking interactions are not significant and therefore less relevant to lignin dissolution. Overall, the QC calculations indicate that hydrogen bonding interactions are playing a vital role in the dissolution of lignin. Organic solvents also exhibit significant hydrogen bonding energies, but due to the strong steric repulsions and weaker polarity, the net result is lower lignin solubility.

### 3. Materials and methods

#### 3.1. Materials

In this study, Sorghum (*Sorghum bicolor*) biomass was obtained from Idaho National Labs (Idaho Falls, Idaho, USA) and used as a raw material. The acquired biomass samples were milled (Thomas-Wiley Model 4, Swedesboro, NJ) and sieved to attain a homogeneous particle size of 2 mm, and finally oven-dried (40 °C, 24 h). The resulting biomass was then placed in an airtight plastic bag and stored in a cool dry place until further use. The following solvents were purchased



**Table 3** The electron donor and acceptor orbitals with their corresponding second-order interaction energies  $E^{(2)*}$  (NBO analysis) of lignin–amine/organic solvent systems. Atom notations are given in Fig. 10 and Fig. S21†

Amine/organic solvent	H-Bond interaction			$\pi$ -Stacking interaction		
	Donor	Acceptor	$E^{(2)*}$ (kJ mol <sup>-1</sup> )	Donor	Acceptor	$E^{(2)*}$ (kJ mol <sup>-1</sup> )
Spermidine	LP (1) N43	$\sigma^*$ O23–H42	101.87	$\pi$ C18–C21	$\sigma^*$ C45–H65	1.46
	LP (1) N47	$\sigma^*$ O13–H32	71.39	$\pi$ C2–C5	$\sigma^*$ C50–H59	1.42
	LP (2) O8	$\sigma^*$ N52–H64	5.31			
1,5-Diaminopentane	LP (1) N44	$\sigma^*$ O23–H43	107.73	$\pi$ C18–C21	$\sigma^*$ C46–H56	0.71
	LP (1) N50	$\sigma^*$ O13–H32	86.32			
Diethylenetriamine	LP (1) N44	$\sigma^*$ O23–H43	101.96	$\pi$ C12–C17	$\sigma^*$ C49–H60	1.30
	LP (1) N50	$\sigma^*$ O11–H31	80.34	$\pi$ C19–C21	$\sigma^*$ C47–H57	2.38
1,3-Diaminopropane	LP (1) N48	$\sigma^*$ O23–H43	102.92	$\pi$ C19–C21	$\sigma^*$ C46–H51	1.09
	LP (1) N44	$\sigma^*$ O13–H32	66.92	$\pi$ C16–C18	$\sigma^*$ C45–H50	0.59
2,2-Dimethyl-1,3-propaneamine	LP (1) N44	$\sigma^*$ O11–H31	69.92	$\pi$ C16–C18	$\sigma^*$ C52–H64	2.51
	LP (1) N48	$\sigma^*$ C22–H42	8.53			
2-Ethoxy ethanol	LP (1) O46	$\sigma^*$ O11–H31	13.68	$\pi$ C12–C17	$\sigma^*$ C45–H49	0.88
	LP (2) O46	$\sigma^*$ O11–H31	30.28			
	LP (1) O20	$\sigma^*$ O57–H59	8.11			
	LP (1) O20	$\sigma^*$ O57–H59	10.66			
Benzyl alcohol	LP (1) O13	$\sigma^*$ O51–H59	8.28	$\pi$ C4–C9	$\pi^*$ C47–C48	1.17
	LP (2) O13	$\sigma^*$ O51–H59	41.23	$\pi$ C47–C48	$\sigma^*$ C14–H35	1.21
Furfuryl alcohol	LP (1) O11	$\sigma^*$ O50–H56	13.01	$\pi$ C4–C9	$\pi^*$ C45–C46	1.13
	LP (2) O11	$\sigma^*$ O50–H56	39.02	$\pi$ C47–C48	$\pi^*$ C4–C9	1.17
Isobutyl acetate	LP (1) O46	$\sigma^*$ O23–H43	17.90	$\pi$ C16–C18	$\sigma^*$ C50–H60	1.42
	LP (2) O46	$\sigma^*$ O23–H43	11.33			

from Sigma Aldrich (St Louis, MO) and used as received: 2,2-dimethyl-1,3-propanediamine (purity: 99%), ethylenediamine (≥99), spermidine (≥99%), spermine (≥99%), 1,2-diaminopropane (99%), 1,3-diaminopropane (≥99%), 1,4-diaminobutane (99%), 1,5-diaminopentane (95%), diethylenetriamine (99%), pentylamine (99%). 2-Ethoxyethanol (≥99%), 2-pyrrolidone (≥99%), trimethyl phosphate (97%), furfuryl alcohol (98%), guaiacol (≥98%), benzyl alcohol (≥99%), isobutyl acetate (99%), aniline, (≥99.5%), furfural (99%), dipropylene glycol (99%), citric acid (ACS reagent ≥99.5%), sodium citrate tribasic dihydrate (ACS reagent, ≥99.0%) and sodium azide. Sulfuric acid (72% and 95–98%) was purchased from VWR, and sugar standards glucose (≥99.5%), xylose (≥99%), and arabinose (≥98%) were procured from Sigma-Aldrich for high-performance liquid chromatography (HPLC) analysis. Commercial cellulase (Cellic® CTec3) and hemicellulase (Cellic® HTec3) mixtures were provided by Novozymes, North America (Franklin, NC).

### 3.2. Pretreatment of sorghum biomass

The pretreatment of biomass sorghum was carried out using the conventional method that involves early separation (or washing) to remove the solvent after pretreatment. In a typical experiment, 1 g of the biomass was mixed with the 4 g of solvent and loaded into an ace pressure tube (50 mL, Ace Glass Inc., Vineland, NJ) and mixed well prior to the experiments. The pretreatment experiments were performed at 140 °C for 3 h of reaction time at a solid loading of 20 wt%. Post pretreatment, 25 mL of ethanol was added to the slurry before being transferred to a 50 mL Falcon tube and centrifuged at 4500 rpm to separate solids from liquid. The recovered solid was

further washed with a mixture of ethanol and water (1 : 1) to remove any residual organic solvents. Finally, the recovered solid fractions were lyophilized before enzymatic hydrolysis (EH) and compositional analysis (CA). All the experiments were performed in duplicate, and the average values are reported here. The solid recovery (%SR) after pretreatment was calculated based on the following equation.

$$\% \text{Solid recovery} (\% \text{SR}) = \frac{\text{weight of biomass recovered after pretreatment}}{\text{weight of biomass used for pretreatment}} \times 100 \quad (5)$$

### 3.3. Enzymatic hydrolysis

The enzymatic saccharification of pretreated and untreated biomass was carried out using commercially available enzymes, Cellic® Ctec3 and Htec3 (9 : 1 v/v) from Novozymes, at 50 °C in a rotary incubator (Enviro-Genie, Scientific Industries, Inc.). All reactions were performed at 5 wt% biomass loading in a 15 mL centrifuge tube (using 0.25 g of the pretreated or untreated biomass). The pH of the mixture was adjusted to 5 with 50 mM sodium citrate buffer supplemented with 0.02 wt% sodium azide to prevent microbial contamination. The total reaction volume included a total protein content of 10 mg per g biomass. The amount of sugars released was analyzed on an Agilent HPLC 1260 infinity system (Santa Clara, California, United States) equipped with a Bio-Rad Aminex HPX-87H column (300 × 7.8 mm<sup>2</sup>) and a refractive index detector. An aqueous solution of sulfuric acid (4 mM) was used as the eluent (0.6 mL min<sup>-1</sup>, column temperature 60 °C). All enzymatic saccharification was conducted in duplicate. The sugar yield was calculated as an overall process yield



using the formula below (eqn (6)), which accounts for sugars/oligosaccharides lost during pretreatment/washing.

$$\begin{aligned} \text{%Sugar yield(process)} \\ = \text{%SR} \times \frac{\text{weight sugars released after hydrolysis}}{\text{weight of sugars in the original biomass}} \end{aligned} \quad (6)$$

### 3.4. Compositional analysis

The compositional analysis of pretreated and untreated biomass sorghum was performed to determine the glucan, xylan, lignin and ash content following the two-step acid hydrolysis procedure previously described by NREL.<sup>62</sup> In summary, 300 mg of the dry extractive-free biomass was exposed to 3 mL of 72% w/w H<sub>2</sub>SO<sub>4</sub> and incubated at 30 °C for 1 h. Subsequently, the mixture was taken through secondary hydrolysis at 4% w/w H<sub>2</sub>SO<sub>4</sub> at 121 °C for 1 h. After the two-step acid hydrolysis, the hydrolysates were filtered using medium porosity filtering crucibles. The filtrates were spectrophotometrically analyzed for the acid-insoluble lignin (ASL) (NanoDrop 2000, Thermo Fisher Scientific, Waltham, MA) using the absorbance at 240 nm. Additionally, monomeric sugars (glucose and xylose) were determined by HPLC using an Agilent 1200 series instrument equipped with a refractive index detector and Bio-Rad Aminex HPX-87H column, coupled with a guard column assembly. Product separation was obtained at 60 °C with 4 mM H<sub>2</sub>SO<sub>4</sub> as a mobile phase at a flow rate of 0.6 mL min<sup>-1</sup>. Finally, the Klason lignin (acid-insoluble lignin – ASL) was determined gravimetrically by subtracting the weight of the oven-dried residual solids (105 °C) and the ash content (575 °C). All compositional analyses were conducted in duplicate. The amount of lignin removed can be calculated using the formula below (eqn (7)). Note: %lignin = %AIL + %ASL.

$$\begin{aligned} \text{%Lignin removal} = 100 - \text{%SR} \\ \times \frac{\text{%lignin}_{\text{after biomass pretreatment}}}{\text{%lignin}_{\text{original biomass}}} \end{aligned} \quad (7)$$

### 3.5. Structural characterization (P-XRD analyses)

The powder X-ray diffraction (P-XRD) analysis were performed to study the crystallinity behavior of cellulose before and after biomass pretreatment. The cellulose allomorph and crystallinity index was measured using in the 2θ range from 5 to 60° and an exposure time of 300 s with a voltage of 40 kV and current of 20 mA using Co-Kα radiation ( $\lambda = 1.7891$  nm). It is important to mention that the diffraction angle was converted to the analogous Cu-Kα ( $\lambda = 1.5418$  nm) for peak identification and analysis using previously assigned spectra. The peak deconvolution of the resulting diffractogram was performed using software PeakFit (SeaSolve Software Inc.). Gaussian/Lorentzian functions were applied in curve fitting analysis and iterations were repeated until the maximum *F* number was obtained. In all cases, the *F* number was >10 000, which corresponds to a *R*<sup>2</sup> value >0.99. Estimation of the content of cellulose I, cellulose III, and amorphous cellulose in the cellulosic

samples was established by using the peak areas of cellulose I, cellulose III, and amorphous cellulose, respectively. According to previously defined diffractogram, the Bragg angles of peak (110), (1̄10), (020), and (004) belonging to cellulose I are ~[14.8°, 16.3°, 22.3°, and 34.5°], respectively. The Bragg angles of peak (110) and (020) belonging to cellulose III are 11.3° and 20.0°, respectively. The Bragg angle of the amorphous peak is around 19.5–20.5°.<sup>49–51</sup> The crystallinity index was also calculated according to the method of Segal *et al.*, where the ratio of the height of the 002 peak (*I*<sub>002</sub>) and the height of the minimum (*I*<sub>AM</sub>) between the 002 and the 101 peaks.<sup>50,59</sup>

### 3.6. Computational details

**3.6.1 COSMO-RS calculations.** The COSMO-RS calculations were carried out to develop the lignin dissolution model in the molecular solvents. First, the structures of cellulose, lignin, and molecular solvents are drawn in the Avogadro freeware software.<sup>63</sup> Initially, the structures of all the investigated molecules were optimized by using Gaussian09 package at B3LYP (Becke 3-parameter hybrid functional combined with the Lee-Yang-Parr correlation) theory and 6-311+G(d,p) basis set To confirm the energy minima of the optimized structure and verify the presence of any imaginary frequency, frequency calculations have been performed at the same level of theory and no imaginary frequencies were present after optimization.

After a successful geometry optimization step, further, the COSMO file was generated using the BVP86/TZVP/DGA1 level of theory.<sup>41,64,65</sup> The ideal screening charges on the molecular surface were computed using the same level of theory *i.e.*, BVP86 through the “scrf = COSMORS” keyword.<sup>66,67</sup> The generated COSMO files were then used as an input in the COSMOtherm (version 19.0.1, COSMOlogic, Leverkusen, Germany) package.<sup>68,69</sup> BP\_TZVP\_19 parametrization was used to predict the sigma potentials, viscosity, excess enthalpy, and logarithmic activity coefficients of the isolated and mixture of molecular systems. In COSMO-RS calculations, the molar fraction of lignin was set as 0.2, whereas the molar fraction of solvents was set to 0.8 to mimic the experimental pretreatment setup.

The excess enthalpy of a binary mixture can be predicted by using the following expression (eqn (8)):<sup>65</sup>

$$H_M^E = \sum x_i H_i^E = \sum x_i [H_{(i, \text{mixture})} - H_{(i, \text{pure})}] \quad (8)$$

where,  $H_M^E$  is the excess enthalpy of solute in the mixture and defined as the enthalpy difference between component *i* in the mixture and in the pure state. On the other hand, excess enthalpy of a mixture is an algebraic sum of the three contributors (eqn (9)) such as electrostatic misfit, hydrogen bonding, and van der Waals interactions.

$$H_M^E = H_M^E(\text{misfit}) + H_M^E(\text{H-bond}) + H_M^E(\text{vdW}) \quad (9)$$

The activity coefficient of component *i* is associated with the chemical potential  $\mu_i$  and expressed as<sup>70</sup> (eqn (10)):

$$\ln(\gamma_i) = \left( \frac{\mu_i - \mu_i^0}{RT} \right) \quad (10)$$

where  $\mu_i^0$  is the chemical potential of the pure component  $i$ ,  $R$  is the real gas constant and  $T$  is the absolute temperature. The details of COSMO-RS calculation in predicting the sigma potential, excess enthalpies, and activity coefficients are provided in the COSMOtherm's user manual.<sup>69</sup>

**3.6.2. Quantum chemical calculations and quantum theory of atom in molecular (QTAIM) analysis.** In addition to COSMO-RS calculations, the quantum chemical (QC) calculations were performed to understand the mechanistic behavior of lignin dissolution in the investigated molecular solvents. For QC simulations, the complex structures of lignin and molecular solvent are combined and drawn using Avogadro freeware software.<sup>63</sup> The geometries of the complex molecular systems were fully optimized at hybrid B3LYP/6-311+G(d,p) function corrected for dispersion interaction using Grimme's dispersion damping (GD3) empirical term.<sup>71</sup> It is known that the combination of B3LYP level of theory and 6-311+G(d,p) basis set is an excellent compromise between the computational cost and accuracy of the computational results.<sup>32,72</sup> All the optimized geometries were obtained without the presence of any imaginary frequencies. The coordinates of the most stable conformers of lignin-solvents are presented in the ESI.†

From QC calculations, the interaction energy ( $\Delta E_{\text{total}}$ ) is calculated by following the eqn (11).<sup>39,73</sup>

$$\text{I.E. (kJ mol}^{-1}) = E_{\text{complex}} - \sum(E_{\text{isolated molecules}}) \quad (11)$$

here,  $E_{\text{complex}}$  is the total energy of the complex system (*i.e.*, lignin-solvent) in  $\text{kJ mol}^{-1}$ .  $E_{\text{isolated molecules}}$  are the individual energies of the lignin or molecular solvent in  $\text{kJ mol}^{-1}$ . Further, SAPT (symmetry-adapted perturbation theory) calculations were performed to study the decomposition of total interaction energies into more meaningful components using the PSI4 program.<sup>69,70</sup>

$$\Delta E_{\text{total}} \text{ (kJ mol}^{-1}) = \Delta E_{\text{elec}} + \Delta E_{\text{exch}} + \Delta E_{\text{ind}} + \Delta E_{\text{disp}} \quad (12)$$

$\Delta E_{\text{elec}}$ ,  $\Delta E_{\text{exch}}$ ,  $\Delta E_{\text{ind}}$ , and  $\Delta E_{\text{disp}}$  terms corresponding to the classic electrostatic, exchange-repulsion, induction, and dispersion interactions.

NBO analysis was employed to understand the strength of the electron donor-acceptor interactions involved in the system. The electron donor  $i-j$  acceptor delocalized stabilization energies ( $E^{(2)*}$ ) were estimated from the second-order perturbation approach and are expressed in the eqn (13) below.<sup>74,75</sup>

$$E^{(2)*} = \Delta E_{ij} = q_i \frac{F(i,j)^2}{\varepsilon_i - \varepsilon_j} \quad (13)$$

where  $\varepsilon_i$  and  $\varepsilon_j$  are the diagonal elements.  $q_i$  is the donor orbital occupancy and  $F(i, j)$  is the off-diagonal Fock matrix. The larger value of the  $E^{(2)*}$ , the more electron tends to transfer from donor to the acceptor.

In addition to the NBO analysis, QTAIM<sup>76</sup> analysis at the bond critical point (BCP) was performed to understand the strength (electron density,  $\rho(r)$ ), characterization (Laplacian energy density ( $\nabla^2 \rho(r)$ )), and nature (energy density  $H(r)$ ) of

the H-bond presented in lignin-molecular solvent systems using AllAIM (version 19.10.12) software.<sup>77</sup> The H-bond energy ( $E_{\text{HB}}$ ) was calculated using Espinosa's equation:  $E_{\text{HB}} = 1/2 \times V_{\text{BCP}}(r)$ , in which  $V_{\text{BCP}}(r)$  is the potential energy density at the BCP of the measured H-bond.<sup>78</sup> Further, to examine the nature of intermolecular interactions in the complex systems, reduced density gradient non-covalent interactions (RDG-NCI) were analyzed using Multiwfn<sup>79</sup> and VMD<sup>80</sup> packages.

**3.6.3. Calculation of solubility parameters.** The solubility parameter, defined as the square root of the cohesive energy density, is one of the key parameters that measure the polarity and quantify the 'like-dissolves-like' principle.<sup>27</sup> The solubility parameters of the molecular solvents under study were calculated by COSMOquick program.<sup>81</sup> The detailed calculation procedure of HSP using COSMOquick is reported elsewhere.<sup>82,83</sup> Based on the cohesive energy density assumptions, Hansen has broken the total solubility parameters ( $\delta_t$ ) into the three contributors such as polar ( $\delta_p$ ), hydrogen-bonded ( $\delta_h$ ), and dispersion ( $\delta_d$ ) forces.<sup>27</sup>

$$\delta_t = \sqrt{(\delta_p^2 + \delta_h^2 + \delta_d^2)} \quad (14)$$

In addition to  $\delta_t$ , Hansen also proposed a parameter called relative energy difference (RED) that correlates the interaction between a solute and a solvent. The RED is defined as the ratio between the radius of interaction ( $R_a$ ) to the 3D sphere radius of the solute ( $R_0$ ) as shown in the below eqn (15) and (16).<sup>27,35,84,85</sup>

$$R_a = \sqrt{4 \left( \delta_p^{\text{solute}} - \delta_p^{\text{solvent}} \right)^2 + \left( \delta_h^{\text{solute}} - \delta_h^{\text{solvent}} \right)^2 + \left( \delta_d^{\text{solute}} - \delta_d^{\text{solvent}} \right)^2} \quad (15)$$

$$\text{RED} = \frac{R_a}{R_0} \quad (16)$$

If the  $\text{RED} < 1$ , then the affinity of the solvent towards the solute is said to be higher. While If the  $\text{RED} > 1$ , the affinity between the solvent and solute is lower.

## 4. Conclusions

The current work demonstrates an effective framework for (1) discovering and predicting high performing solvents for the dissolution of lignin and (2) understanding the mechanistic factors that control the lignin dissolution capacity of a solvent. This framework relies on multiscale simulation approaches to develop a predictive model to identify potential solvents for lignocellulosic biomass pretreatment, which are then experimentally verified. To initiate the development of this framework, several molecular solvents were screened using HSP and COSMO-RS models and amines were predicted to be effective solvents. Next, amines with different chemical functionalities were experimentally tested and shown to promote higher lignin removal and fermentable sugar yields than several non-amine solvents. The solvents 1,5-diaminopentane and 1,2-dia-

minopropane were found to be the most effective at biomass pretreatment, extracting >82% of lignin from biomass and enabling >88% yields of fermentable sugars. COSMO-RS was then used to develop a predictive model for lignin removal based on the several influential quantities:  $H^E$ ,  $\ln(\gamma)$ ,  $\eta$ , and  $pK_a$ . Comparison of the predictive model and experimental results shows that they are in excellent agreement, with a deviation of less than 10%.

Once effective lignin solvents were identified, QC calculations and QTAIM analysis were employed to understand the mechanism that drive the lignin solvent interactions and determine why the amines are more effective lignin extraction solvents than the other non-amine solvents examined in this study. QC and QTAIM analysis indicate that amines that form multiple strong H-bond interactions with lignin can extract high amounts of lignin from biomass. The use of computational platforms to both develop predictive models to identify effective pretreatment solvents and to then gain deeper insights into the mechanism of lignin dissolution by these solvents will lead to the rapid expansion of the list of solvents that can be used for efficient lignocellulose pretreatment and deconstruction. There are numerous considerations that must be made to effectively integrate a pretreatment technology into a biorefinery, including effectiveness on a broad range of feedstocks, fractionation of lignocellulose components, solvent cost, solvent recycling, generation of biomass-derived enzyme and microbe inhibitors, *etc.* Recommendations for subsequent research will study these additional components of process development and the expansive list of pretreatment solvents identified through the predictive framework established in this study will provide researchers and industry more options to consider in the development of highly efficient, low-cost lignocellulose conversion technologies.

## Conflicts of interest

The authors declare the following competing financial interest (s): EA, MM, HC, LD, KH, HDM, VRP, BAS, and JMG are named inventors on at least one related patent application. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

## Acknowledgements

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The United States

Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. XRD characterization was conducted with the help of Dr Tevy Kuykendall at the Molecular Foundry in Lawrence Berkeley National Laboratory.

## Notes and references

- 1 M. Galbe and O. Wallberg, *Biotechnol. Biofuels*, 2019, **12**, 294.
- 2 M. C. Y. Chang, *Curr. Opin. Chem. Biol.*, 2007, **11**, 677–684.
- 3 S. I. Mussatto and G. M. Dragone, in *Biomass fractionation technologies for a lignocellulosic feedstock based biorefinery*, Elsevier, 2016, pp. 1–22.
- 4 M. V. Rodionova, R. S. Poudyal, I. Tiwari, R. A. Voloshin, S. K. Zharmukhamedov, H. G. Nam, B. K. Zayadan, B. D. Bruce, H. J. M. Hou and S. I. Allakhverdiev, *Int. J. Hydrogen Energy*, 2017, **42**, 8450–8461.
- 5 E. C. Achinivu, R. M. Howard, G. Li, H. Gracz and W. A. Henderson, *Green Chem.*, 2014, **16**, 1114–1119.
- 6 Z. Strassberger, S. Tanase and G. Rothenberg, *RSC Adv.*, 2014, **4**, 25310–25318.
- 7 B. M. Upton and A. M. Kasko, *Chem. Rev.*, 2016, **116**, 2275–2306.
- 8 P. Kumar, D. M. Barrett, M. J. Delwiche and P. Stroeve, *Ind. Eng. Chem. Res.*, 2009, **48**, 3713–3729.
- 9 L. Yao, C. Chen, C. G. Yoo, X. Meng, M. Li, Y. Pu, A. J. Ragauskas, C. Dong and H. Yang, *ACS Sustainable Chem. Eng.*, 2018, **6**, 14767–14773.
- 10 J. Kautto, M. J. Realff, A. Ragauskas and T. Kässi, *BioResources*, 2014, **9**, 6041–6072.
- 11 C. Tang, J. Shan, Y. Chen, L. Zhong, T. Shen, C. Zhu and H. Ying, *Bioresour. Technol.*, 2017, **232**, 222–228.
- 12 Z. Zhang, M. D. Harrison, D. W. Rackemann, W. O. S. Doherty and I. M. O'Hara, *Green Chem.*, 2016, **18**, 360–381.
- 13 M. N. Boran and F. Karaosmanoglu, *J. Renewable Sustainable Energy*, 2018, **10**, 033104.
- 14 X. Zhao, S. Li, R. Wu and D. Liu, *Biofuels, Bioprod. Biorefin.*, 2017, **11**, 567–590.
- 15 A. Rodríguez, E. Espinosa, J. Domínguez-Robles, R. Sánchez, I. Bascón and A. Rosal, in *Pulp and paper processing*, ed. S. N. Kazi, InTech, 2018.
- 16 H. Chen, J. Zhao, T. Hu, X. Zhao and D. Liu, *Appl. Energy*, 2015, **150**, 224–232.
- 17 F. Cheng, T. Ouyang, J. Sun, T. Jiang and J. Luo, *BioResources*, 2019, **14**, 486–499.
- 18 X. Zhao, K. Cheng and D. Liu, *Appl. Microbiol. Biotechnol.*, 2009, **82**, 815–827.
- 19 L. Qin, X. Li, J.-Q. Zhu, W.-C. Li, H. Xu, Q.-M. Guan, M.-T. Zhang, B.-Z. Li and Y.-J. Yuan, *Ind. Crops Prod.*, 2017, **102**, 51–57.



20 F. Cheng, X. Zhao and Y. Hu, *Bioresour. Technol.*, 2018, **249**, 969–975.

21 D. T. Balogh, A. A. S. Curvelo and R. A. M. C. De Groot, *Holzforschung*, 1992, **46**, 343–348.

22 Y. Ye, Y. Liu and J. Chang, *BioResources*, 2014, **9**, 3417–3427.

23 J. Quesada-Medina, F. J. López-Cremades and P. Olivares-Carrillo, *Bioresour. Technol.*, 2010, **101**, 8252–8260.

24 L. Soh and M. J. Eckelman, *ACS Sustainable Chem. Eng.*, 2016, **4**, 5821–5837.

25 T. T. Kwok, K. R. Bright, M. J. Realff and A. S. Bommarius, *BioResources*, 2019, **14**, 5988–6003.

26 L. P. Novo and A. A. S. Curvelo, *Ind. Eng. Chem. Res.*, 2019, **58**, 14520–14527.

27 C. M. Hansen, *Hansen Solubility Parameters: A User's Handbook*, Crc Press, Boca Raton, 2nd edn, 2007.

28 C. Balaji, T. Banerjee and V. V. Goud, *J. Solution Chem.*, 2012, **41**, 1610–1630.

29 A. Casas, J. Palomar, M. V. Alonso, M. Oliet, S. Omar and F. Rodriguez, *Ind. Crops Prod.*, 2012, **37**, 155–163.

30 A. Casas, S. Omar, J. Palomar, M. Oliet, M. V. Alonso and F. Rodriguez, *RSC Adv.*, 2013, **3**, 3453–3460.

31 Q. Zhang, X. Tan, W. Wang, Q. Yu, Q. Wang, C. Miao, Y. Guo, X. Zhuang and Z. Yuan, *ACS Sustainable Chem. Eng.*, 2019, **7**, 8678–8686.

32 Y. Zhang, H. He, K. Dong, M. Fan and S. Zhang, *RSC Adv.*, 2017, **7**, 12670–12681.

33 W. Ji, Z. Ding, J. Liu, Q. Song, X. Xia, H. Gao, H. Wang and W. Gu, *Energy Fuels*, 2012, **26**, 6393–6403.

34 R. Vanholme, B. Demedts, K. Morreel, J. Ralph and W. Boerjan, *Plant Physiol.*, 2010, **153**, 895–905.

35 C. M. Hansen and A. Björkman, *Holzforschung*, 1998, **52**, 335–344.

36 W. Thielemans and R. P. Wool, *Biomacromolecules*, 2005, **6**, 1895–1905.

37 K. L. Hoy, *J. Coated Fabr.*, 1989, **19**, 53–67.

38 D. W. van Krevelen and K. te Nijenhuis, *Properties of Polymers*, 4th edn, 2009.

39 M. Mohan, P. Viswanath, T. Banerjee and V. V. Goud, *Mol. Phys.*, 2018, **116**, 1–21.

40 J. Kahlen, K. Masuch and K. Leonhard, *Green Chem.*, 2010, **12**, 2172–2181.

41 Y.-R. Liu, K. Thomsen, Y. Nie, S.-J. Zhang and A. S. Meyer, *Green Chem.*, 2016, **18**, 6246–6254.

42 Y. Liu, J. Zheng, J. Xiao, X. He, K. Zhang, S. Yuan, Z. Peng, Z. Chen and X. Lin, *ACS Omega*, 2019, **4**, 19829–19839.

43 D. Song, A. F. Seibert and G. T. Rochelle, *Energy Procedia*, 2014, **63**, 1268–1286.

44 D. Song, A. F. Seibert and G. T. Rochelle, *Energy Procedia*, 2017, **114**, 2713–2727.

45 X. Zhao, R. Wu and D. Liu, *Bioresour. Technol.*, 2018, **261**, 52–61.

46 J. P. F. Simão, M. G. V. S. Carvalho and C. M. S. G. Baptista, *Chem. Eng. J.*, 2011, **170**, 264–269.

47 A. Hartono and H. F. Svendsen, *J. Chem. Thermodyn.*, 2009, **41**, 973–979.

48 S. Padmanabhan, P. Schwyter, Z. Liu, G. Poon, A. T. Bell and J. M. Prausnitz, *3 Biotech.*, 2016, **6**, 23.

49 E. L. Hult, T. Iversen and J. Sugiyama, *Cellulose*, 2003, **10**, 103–110.

50 S. Park, J. O. Baker, M. E. Himmel, P. A. Parilla and D. K. Johnson, *Biotechnol. Biofuels*, 2010, **3**, 10.

51 L. Qin, W.-C. Li, J.-Q. Zhu, J.-N. Liang, B.-Z. Li and Y.-J. Yuan, *Biotechnol. Biofuels*, 2015, **8**, 174.

52 Ismojo, R. Hadiwibowo, E. Yuanita, E. Kustiyah and M. Chalid, *Macromol. Symp.*, 2020, **391**, 1900129.

53 J. P. Vandenbrink, R. N. Hilten, K. C. Das, A. H. Paterson and F. A. Feltus, *BioEnergy Res.*, 2012, **5**, 387–397.

54 S. P. S. Chundawat, G. Bellesia, N. Uppugundla, L. da Costa Sousa, D. Gao, A. M. Cheh, U. P. Agarwal, C. M. Bianchetti, G. N. Phillips, P. Langan, V. Balan, S. Gnanakaran and B. E. Dale, *J. Am. Chem. Soc.*, 2011, **133**, 11163–11174.

55 A. Mittal, R. Katahira, M. E. Himmel and D. K. Johnson, *Biotechnol. Biofuels*, 2011, **4**, 41.

56 T. Cui, J. Li, Z. Yan, M. Yu and S. Li, *Biotechnol. Biofuels*, 2014, **7**, 134.

57 Z.-H. Liu, L. Qin, B.-Z. Li and Y.-J. Yuan, *ACS Sustainable Chem. Eng.*, 2015, **3**, 140–146.

58 Z. Yu, K.-S. Gwak, T. Treasure, H. Jameel, H. Chang and S. Park, *ChemSusChem*, 2014, **7**, 1942–1950.

59 L. Segal, J. J. Creely, A. E. Martin and C. M. Conrad, *Text. Res. J.*, 1959, **29**, 786–794.

60 U. Koch and P. L. Popelier, *J. Phys. Chem.*, 1995, **99**, 9747–9754.

61 P. L. A. Popelier, *J. Phys. Chem. A*, 1998, **102**, 1873–1878.

62 A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter and D. Templeton, NREL, 2008, TP-510-42618.

63 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.

64 M. Mohan, T. Banerjee and V. V. Goud, *J. Chem. Eng. Data*, 2016, **61**, 2923–2932.

65 M. Gonzalez-Miquel, M. Massel, A. DeSilva, J. Palomar, F. Rodriguez and J. F. Brennecke, *J. Phys. Chem. B*, 2014, **118**, 11512–11522.

66 M. Mohan, V. V. Goud and T. Banerjee, *Fluid Phase Equilib.*, 2015, **395**, 33–43.

67 R. Anantharaj and T. Banerjee, *Ind. Eng. Chem. Res.*, 2010, **49**, 8705–8725.

68 Y. Li and Y. Jin, *Renewable Energy*, 2015, **77**, 550–557.

69 F. Eckert and A. Klamt, *AIChE J.*, 2002, **48**, 369–385.

70 K. A. Kurnia, S. P. Pinho and J. A. P. Coutinho, *Ind. Eng. Chem. Res.*, 2014, **53**, 12466–12475.

71 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

72 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.

73 Z.-D. Ding, Z. Chi, W.-X. Gu, S.-M. Gu, J.-H. Liu and H.-J. Wang, *Carbohydr. Polym.*, 2012, **89**, 7–16.

74 M. Mohan, P. K. Naik, T. Banerjee, V. V. Goud and S. Paul, *Fluid Phase Equilib.*, 2017, **448**, 168–177.

75 J. Guo, D. Zhang, C. Duan and C. Liu, *Carbohydr. Res.*, 2010, **345**, 2201–2205.



76 R. F. W. Bader, *A Quantum Theory*, 1990.

77 T. A. Keith, *TK Gristmill Software*, Overland Park, KS, USA, 2019.

78 E. Espinosa, E. Molins and C. Lecomte, *Chem. Phys. Lett.*, 1998, **285**, 170–173.

79 T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.

80 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38, 27.

81 C. Loschen and A. Klamt, *Ind. Eng. Chem. Res.*, 2012, **51**, 14303–14308.

82 A. Niederquell, N. Wytttenbach and M. Kuentz, *Int. J. Pharm.*, 2018, **546**, 137–144.

83 M. Diedenhofen, F. Eckert, A. Hellweg, H. C. Steffen, U. Huniar, A. Klamt, C. Loschen, L. Koch, J. Reinisch, J. Schwobel, M. Suray, S. Terzi, K. Wichmann and T. R. Rosenbaum, *COSMOquick, Version 1.7*, COSMOLogic GmbH & Co. KG, Leverkusen, Germany, 2018.

84 H. Yu, J. Hu and J. Chang, *Ind. Eng. Chem. Res.*, 2011, **50**, 7513–7519.

85 C. M. Hansen, *J. Paint Technol.*, 1967, **39**, 104–117.

