



Machine Learning-based approach for Tailor-Made design of ionic Liquids: Application to CO₂ capture



Kexin Zhang ^{a,b,c,1}, Jiasheng Wu ^{a,1}, Hyeonsuk Yoo ^d, Yongjin Lee ^{a,d,*}

^a School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China

^b Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201203, China

^c University of Chinese Academy of Sciences, Beijing 100049, China

^d Department of Chemistry and Chemical Engineering, Education and Research Center for Smart Energy and Materials, Inha University, Incheon 22212, Republic of Korea

ARTICLE INFO

Keywords:

Ionic Liquid
Recurrent neural network
Monte Carlo tree search
Carbon capture

ABSTRACT

In this article, we present a machine learning-based approach for the tailor-made design of ionic liquids (ILs) promising toward the desired target applications. Our computational framework combines multi-player Monte Carlo tree search and recurrent neural network, within a parallel scheme of generating and testing multiple ILs simultaneously, to improve the efficiency of searching optimal structures. For two cases of CO₂ capture from 1) flue gas (CO₂/N₂) and 2) from syngas (CO₂/H₂), target-specific ILs were generated in our computational platform according to objective function values that combine three requirements of high CO₂ solubility, absorption selectivity of IL for CO₂, and easiness of subsequent desorption. Our results showed that high-performance ILs can be designed with great efficiency using our algorithm. Furthermore, topological data analysis on newly designed ILs demonstrated that our algorithm allows us to explore materials space widely to find high-performing ILs with good diversity.

1. Introduction

Ionic liquids (ILs) are defined as salts that exist in the liquid state at near-ambient temperature. Due to the extensively diverse combination of cation and anions, ILs have the intrinsic tunability of the physical and chemical properties, which has drawn continuously research interest to this class of compounds as a designer or task-specific solvents [1] for a wide range of chemical processes [2–7]. One prospective application of the ILs is as an alternative solvent to capture CO₂ [8–10] which is the most significant contributor to global warming, while current CO₂ capture processes based on using amine solutions can have high energy consumption for regenerating absorbent and corrosion issues. [11]

In principle, ILs provide an ideal platform to develop a tailor-made material for a target application because the properties of ILs can be tuned by modifying or replacing both the cation and anion. However, the vast chemical space of ILs can be regarded as a challenge that needs to be tackled, considering that in reality, the experimental trial-and-error is limited to test only a small fraction of all possible materials. Indeed, the number of ILs that are commercially available or could be constructed with ions published to date would be estimated as many as

1.5×10^5 or more that is significantly smaller than that of possible ILs (over 10^{18}) [12,13]. Therefore, it is increasingly desirable to develop efficient computational techniques that explore and customize the IL structures from a rich abundance of structural libraries to optimally match requirements for a given application.

As one promising approach to accomplish this demand, studies performing the -so-called- inverse design of materials [14] have been recently reported, by virtue of increasing advances in machine learning algorithms, among which deep learning [15,16] is mostly emerging. As demonstrated by the previous studies, the inverse design approach is capable to selectively generate target-specific structures for various types of materials including inorganic [17], organic molecules [18–20], and linker molecules of metal-organic frameworks [21–25]. However, to our knowledge, such effort has been lacking for designing ILs, while in the field of ILs the issue “vast chemical space needs to be explored” seems clear.

In this work, we aim to propose an inverse design approach-based computational framework for the tailor-made generation of target-specific ILs; to be specific, optimally designing side-branch structures for a given cation. The framework is constructed based on a multi-player

* Corresponding author at: School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China.

E-mail address: yongjin.lee@inha.ac.kr (Y. Lee).

¹ Contributed equally to this work.

Monte Carlo tree search (MP-MCTS) [26] combined with a recurrent neural network (RNN). To improve the efficiency of searching optimal structures, our computational tool also employs a parallel scheme [27] which generates and tests multiple ILs simultaneously. We test our algorithm by performing *in-silico* experiments to design high-performing ILs for two cases of CO₂ capture applications; 1) CO₂ separation from flue gas (CO₂/N₂) and 2) the separation of CO₂ from syngas (CO₂/H₂). The design of new ILs is directed to meet three requirements; high solubility of CO₂, absorption selectivity of IL for the target gas component, and easiness of subsequent desorption. Furthermore, we inquire into the capability of our algorithm in respect of discovering new diversity of top-performing ILs using topological data analysis [28].

2. Computational details

2.1. Representation of ionic liquid

A language model of describing chemical systems (i.e. ionic liquid (IL) in this work) was established by utilizing the Simplified Molecular Input Line Entry System (SMILES) [29] which represents molecules as sequenced strings of symbols based on principles of molecular graph theory. In the SMILES model, a string (S) is in a vector form of $S = \{s_1, s_2, \dots, s_T\}$, where symbol s_i represents chemical information such as atoms, ions, bonds, the number of rings, branch, or terminal symbol. Because this work is focused on functionally designing organic side branches in a cation of the pre-existing IL, we constructed a SMILES only for a cation, while an anion is represented as characters according to chemical formula. In our definition of SMILES strings, the following 22 symbols were used: $s_i \in \{\text{""}, \$, \#, (\text{, }), \text{--}, \text{1}, \text{=}, \text{C}, \text{F}, \text{H}, \text{N}, \text{O}, \text{S}, [\text{, }], \text{c}, \text{n}, \text{o}, \text{s}, \text{Cl}, \text{Br}\}$. (for the meaning of each symbol, see Table S1)

2.2. Multi-Player Monte Carlo tree search

Fig. 1 shows a schematic of our algorithm to generate ionic liquids (ILs). It begins with providing input information about a pristine cation (i.e. in this work, imidazolium ring without branches), anion, and target application, and then side branches are generated using the multi-player Monte Carlo Tree Search (MP-MCTS) [26]. Like a standard MCTS, MP-MCTS starts with the root node (symbolled as "") and grows a search tree node by node with repetition of the four operations that are composed of selection, expansion, simulation, and backpropagation; a node corresponds to one symbol of SMILES. The special feature of MP-MCTS lies in the operation of the selection step. In the selection step, a path from the root to an edge node at the current level is built by selecting the children nodes which maximize the upper confidence bound (UCB) [30] (see SI for details). In MP-MCTS, the selection of

symbols is performed in rotation among multiple players, as depicted in **Fig. 1**. In our study, three players were employed in MP-MCTS, and each player attempts to design the designated branch in an imidazolium cation.

Next, at the expansion step, a new child node is added under the edge node corresponding to the currently working player. Then the remaining path from the added node to the terminal node is obtained via rollout processes [30] (i.e., simulation step). During a rollout process, a new symbol is added one by one until the terminal symbol '\$' occurs or the string length reaches the limit which is set as 43 in this work; note that a rollout process is operated by all players, and continued until all players obtain terminal symbols (i.e. complete molecule). Given a partial string of symbols, the conditional probability distribution for the next symbol is determined by the recurrent neural network (RNN) trained on 414,972 SMILES selected from the MOSES database [31] with the criteria of zero formal charges on all constituent atoms.

A complete SMILES string obtained by the rollout is checked for its validity using RDKit [32] and converted to an organic molecule in 3D coordinates. After an organic molecule for the side branch is constructed, we identified atoms to be connected to the cation. The first (in an order of strings in SMILES) nonhydrogen atom was chosen as a connection site except the case that the atom is a part of an aromatic ring. Then, one hydrogen atom attached to the atom (connection site) was removed and a single bond was formed between the atom and the connection point of a cation. The constructed cation was geometrically optimized using RDKit [32]. For ILs successfully built, the reward function values were estimated through atomistic simulations. The details about the reward function and its estimation in our algorithm are discussed in "Results and Discussion".

As the last stage of an iteration in MP-MCTS, the backpropagation step propagates the reward function value backward and updates nodes' information including the UCB scores along the path back to the root node. Then, a new iteration is initiated again.

2.3. Parallelization of MP-MCTS

In the MP-MCTS, the most time-consuming part is the simulation step where the reward function value is estimated at every MP-MCTS iteration. Because halt of MCTS operation is required before backpropagation step to avoid data corruption between MCTS iterations, searching for the optimum structure would take a long time as the more computationally expensive method is employed to evaluate the reward function value. Therefore, instead of pausing the MCTS and waiting for the completion of the simulation step, operating multiple MCTS would be beneficial because multiple structures can be tested simultaneously. This parallelization of MCTS is feasible because procedures of evaluating the reward function value do not require any information about the MCTS tree and

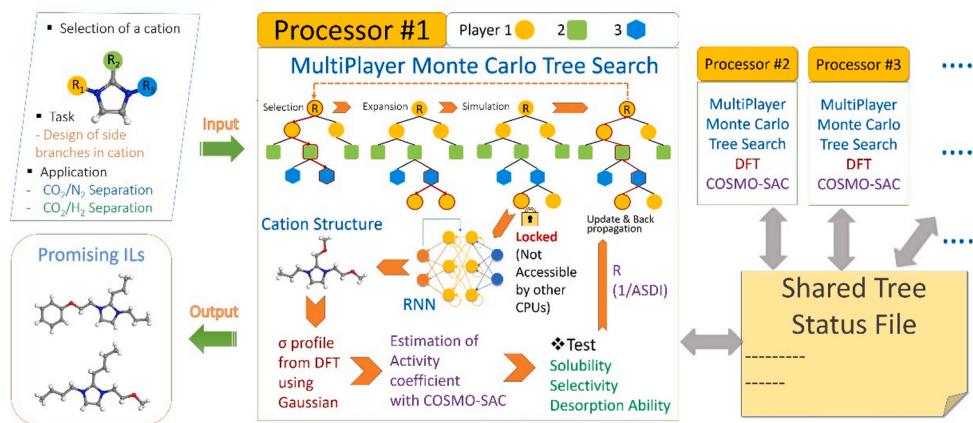


Fig. 1. A schematic of our algorithm to design an application-specific ionic liquid. With given inputs about target application and cation, side branches of the cation core are generated by our algorithm combining the multi-player Monte Carlo tree search (MP-MCTS) and recurrent neural network (RNN). A new IL (a preset anion + cation with newly designed branches) undergoes a performance check for the target application and then internal parameters in MP-MCTS are updated. With structured parallel strategy, multiple attempts of designing ILs are executed simultaneously, and the corresponding processors share the same tree state file. When a certain node is under atomistic simulation to evaluate the reward (R) value, the node is locked and not accessible by other processors.

thus are independent of each other. Among viable parallel MCTS algorithms [33], we employed a structured parallel programming MCTS [27] approach. This approach assigns each operation of the MCTS to a separate processing element according to its computation patterns regardless of iteration, while conventional parallel methods are based on iteration-level parallelism assigning each iteration of MCTS to a separate processor. In our parallel approach, a status file is generated to store one MCTS tree which is visited and updated by multiple processors in common (see Fig. 1). To prevent any data corruption caused by simultaneous accesses from multiple processors, a node lock mechanism is employed in our algorithm. When evaluation of reward function value is in progress, the corresponding node of MCTS is marked as “activated state”, which makes the node not to be selected by other processors. After the evaluation is done, the node is turned back to “deactivated state” at the backpropagation step. With this strategy, multiple processors can work together within one MCTS tree without any wait or disturbance.

Considering the possibility that the same cation is designed and tested repeatedly by multiple processors, during a run for designing IL, our algorithm stores MCTS history including cations generated and the corresponding reward function values, and this database is shared among different processors.

3. Results and discussion

We firstly checked the capability of our algorithm based on multi-player Monte Carlo Tree Search (MP-MCTS) in designing target-specific ILs as the promising absorbents for two cases of CO₂ capture applications; 1) CO₂ separation from flue gas (CO₂/N₂) and 2) the separation of CO₂ from syngas (CO₂/H₂). To address this, we chose an IL (hereinafter referred to as an original IL) consisted of the 1-ethyl-3-methylimidazolium cation ([Emim]⁺) and bis(trifluoromethylsilyl) anion (TFSI⁻), and made attempts to re-functionalize three branches R₁, R₂, and R₃ (see Fig. 1) of imidazolium ring core in the cation to achieve the desired properties, which has been demonstrated as an effective strategy to adjust chemical/physical properties of imidazolium cation in the previous studies [13,34,35].

The existence of three branches provides a huge degree of freedom for designing optimal structures; for instance, even for the case of designing only aliphatic hydrocarbon branches with the maximum number of carbons as 10 at each branching site, the number of possible branches (including hydrogen atom) is 873 for each chain, which results in more than 600 million cations (=873³). Here, we, therefore, imposed the limitation on the number of substituent groups as 10 in each branch.

For each case, 10 independent runs were made. For a single run, a procedure of generating ILs continues until the maximum value of the reward function is not updated during additional 1000 cations are generated.

3.1. Reward function

For efficient CO₂ capture and separation process, an optimal absorbent is required to have a high CO₂ solubility which is judged according to Henry's law constant in dilute solutions. In addition to the high solubility, the absorption selectivity of IL for the target gas component is essential, because multicomponent gas mixtures are faced in practical gas absorption processes rather than single-component gases. In the gas separation process, the easiness of subsequent desorption is another important concern in the aspects of solvent reuse and energy saving. To take account of these factors in evaluating the performance of ILs candidates for carbon capture and separation, we utilized the Absorption-Selectivity-Desorption index [10], ASDI, integrating the main thermodynamic properties of gas solubility, selectivity, and desorption capacity as follows.

$$ASDI = H_{i/IL} \times \frac{1}{S_{i/j}} \times D \quad (1)$$

Here, $H_{i/IL}$ is Henry's law constant of gas i in IL and the solvent selectivity toward target gas component i over other undesirable components j ($S_{i/j}$) is defined as $H_{j/IL}/H_{i/IL}$, with $i = \text{CO}_2$ and $j = \text{N}_2$ or H_2 . The desorption capacity (D) of ILs is estimated as $H^{\text{abs}}/H^{\text{des}}$ where H^{abs} and H^{des} denote Henry's law constants of CO₂ at the absorption (298.15 K) and desorption (323.15 K) temperature, respectively.

The multiplication form in Eq. (1) is intended to balance different thermodynamic performance indices, considering a possible trade-off between them [36,37]. It is worth noting that we adopted $H_{i/IL}$, $S_{i/j}$, and D in mass-based units because ILs with higher mass-based absorption performance are more practically desirable than those with mole-based absorption performance [10].

$H_{i/IL}$ is estimated as $H_{i/IL} = \gamma_i^\infty P_i^s$. P_i^s is the saturated vapor pressure of the gas (see SI for calculation details) and γ_i^∞ is the infinite dilution activity coefficient of the gas in the solution. In our study, γ_i^∞ was calculated by COSMO-SAC model [38–40] with the σ profiles obtained from the density functional theory calculation within Gaussian09 at B3LYP 6–31 g(d,p) level [41].

Because higher solubility, selectivity, desorption capacity are desired, and hence a lower ASDI value is favorable for the gas separation. To take account of this, the reward function (R) in MP-MCTS was defined as

$$R = \begin{cases} 1/ASDI \\ 0 \text{ when IL is not constructed} \end{cases} \quad (2)$$

3.2. Tailor-Made design of ionic liquids

Tables 1 and 2 show examples of newly designed ILs promising for CO₂/N₂ and CO₂/H₂ separations, respectively. The complete lists for both cases are provided as Supporting Information. For both applications, our algorithm was successful in generating novel ILs better performing than the original IL consisted of the 1-ethyl-3-methylimidazolium cation ([Emim]⁺) and bis(trifluoromethylsilyl) anion (TFSI⁻). For instance, with side groups of -(C₃H₇), -(C₃H₇), and -(C₂H₄O)(C₆H₆), our approach designed a promising IL for CO₂/N₂ separation having a significantly low ASDI value (=0.002045) with $H_{i/IL} = 329.39$, $\frac{1}{S_{i/j}} = 9.71 \times 10^{-6}$ and $D = 0.6393$, compared to the original IL showing ASDI = 0.003575 with $H_{i/IL} = 569.36$, $\frac{1}{S_{i/j}} = 9.68 \times 10^{-6}$ and $D = 0.6488$. The same IL was also obtained as the top candidate for CO₂/H₂ separation and predicted to have the ASDI of 0.6438 ($H_{i/IL} = 329.39$, $\frac{1}{S_{i/j}} = 0.00305$ and $D = 0.6438$) which is 2 times lower than ASDI = 1.387 of the original IL with $H_{i/IL} = 542.16$, $\frac{1}{S_{i/j}} = 0.003921$ and $D = 0.6524$. It would be instructive to discuss how our algorithm takes the design strategies to improve the absorption performance of IL. From structural formulas of cations for top-performing ILs in Tables 1 and 2, it is identified that side branches in promising cations have long alkyl chains or ether groups which has been proposed as an effective strategy to enhance the solubility of CO₂ in the IL from the previous studies [34,42–44]. The results obtained for the fourth-ranked cation in CO₂/N₂ study suggest that there would be gain with the introduction of amine functional group into cation due to the strong interaction between CO₂ and negatively charged nitrogen atoms.

In designing cations, we adopted the inverse of ASDI as the reward function R to evaluate the carbon capture performance property of ILs. Because the ASDI is defined in one functional form integrating the three thermodynamic properties of gas solubility, selectivity, and desorption capacity, in general, it would be expected that optimizing three properties together is challenging because optimization of one property can be achieved at significant expenses of others. Therefore, it is worth discussing whether the ASDI (i.e. reward function R) is properly chosen

Table 1

Examples of top-promising cations designed using our algorithm for CO₂ separation from flue gas (CO₂/N₂). Along with each cation structure, SMILES is shown. ‘No.’ is the rank in terms of the reward function value R(=1/ASDI). H_{i/IL}, 1/S_{i/IL}, and D is Henry’s law constant of CO₂ in ionic liquid, the inverse of selectivity of CO₂ over N₂, desorption capacity obtained from COSMO-SAC, respectively. For COSMO-SAC simulations, ionic liquids consist of the newly obtained cation and bis triflimide anion (TFSI⁻). The complete list of top structures is given as the Supporting Information.

No.	Cation Structure	SMILES	R (bar ⁻¹)	ASDI(10 ⁻³ bar)	H _{i/IL} (bar)	1/S _{i,j} (10 ⁻⁶)	D
1		CCCc1n(CCC)cc[n+]1CCOc1cccc1	488.9	2.045	329.4	9.71	0.639
2		CCCc1n(CCOC2cccc2)cc[n+]1CCOC	488.5	2.047	339.4	9.46	0.637
3		Cc1n(C)cc[n+]1C	485.7	2.059	367.3	8.85	0.633
4		CCCc1n(OCCNCC2cccc2)cc[n+]1OCC1cccc1	476.7	2.098	331.5	9.91	0.638
5		CCCC[n+]1ccn(CCOC2cccc2)c1CCOc1cccc1	476.3	2.099	345.2	9.48	0.642
6		CCCC[n+]1ccn(CCC)c1CCOc1cccc1	475.1	2.105	334.8	9.81	0.641
7		CCCCn1cc[n+](CCOC2cccc2)c1CCC	474.2	2.109	331.7	9.93	0.640
8		CCCc1n(CCOC)cc[n+]1CCC	473.7	2.111	330.4	10.10	0.635
9		CCCc1n(CCOC2cccc2)cc[n+]1CCOc1cccc1	473.6	2.112	351.6	9.36	0.641
10		CCC[n+]1ccn(CCOC)c1COC	472.5	2.116	339.3	9.82	0.635
Ref.		CCN1C = C[N+](=C1)C	279.7	3.575	569.4	9.68	0.649

and how our algorithm designs novel cations by balancing three properties (H_{i/IL}, $\frac{1}{S_{i,j}}$, D), especially the balance between H_{i/IL} and $\frac{1}{S_{i,j}}$, or H_{i/IL} and D, considering a possible trade-off between them. Under the situation that smaller H_{i/IL} (greater solvent ability to CO₂), $\frac{1}{S_{i,j}}$ (higher selectivity to CO₂), and D (better recyclability of ILs) are desired to achieve lower ASDI (i.e. higher R value), Figs. 2 and 3 show that our algorithm successfully designed highly performing ILs (high R value) by satisfying together good solubility, selectivity, and recyclability, for both CO₂/N₂ and CO₂/H₂ separation applications; reduction of H_{i/IL} is achieved in conjunction with decreases in $\frac{1}{S_{i,j}}$ and D. Compared to the CO₂/H₂ separation, the results for CO₂/N₂ separation seem less satisfactory; $\frac{1}{S_{i,j}}$ seems to be varied regardless of H_{i/IL}, as shown in Fig. 2(a). Although it might seem to indicate incapability of our algorithm, it would be explained rather by the fact that the original IL is already quite optimized for the application (as shown by low $\frac{1}{S_{i,j}} = 9.68 \times 10^{-6}$) with leaving a little room for further upgrade. Indeed, [Emim][TFSI] has been proposed as a promising absorbent for CO₂/N₂ separation, due to the strong interaction between the positive charge on the carbon of CO₂ with the negative on the fluorine atoms of the anion [45].

3.3. The efficiency of designing novel ILs

It would be interesting to see how efficiently our algorithm based on MP-MCTS generates promising ILs for a target application. For comparison, we performed a comparative study of designing three branches of cations separately one by one via consecutive three standard MCTS. The effectiveness of each approach can be judged according to the success rate of obtaining promising IL per cycle and cycles required to find an optimum IL (note: each cycle produces one IL). As shown in Fig. 4, during the same cycle span, MP-MCTS designed the promising ILs with higher success rate than standard MCTS; 9.83 or 13.1% with a criterion of R greater than 450 or 1.4 for CO₂/N₂ or CO₂/H₂, respectively, which is remarkably efficient compared to the success rate of 0.25 or 1.67% in the case of using standard MCTS. Also, the promising IL was firstly found after 54 or 37 MP-MCTS cycles that are significantly earlier than 1097 or 897 cycles for the cases using standard MCTS, for CO₂/N₂ or CO₂/H₂, respectively. These results indicate that our algorithm based on MP-MCTS is successful in designing promising ILs more efficiently than standard MCTS does.

3.4. On the diversity of newly designed ILs

A set of ILs generated in this work was enriched with structures promising for a target application. While the result shows the ability of

Table 2

Examples of top-promising cations designed using our algorithm for CO₂ separation from syngas gas (CO₂/H₂). Along with each cation structure, SMILES is shown. ‘No.’ is the rank in terms of the reward function value R(=1/ASDI). H_{i/IL}, 1/S_{i/IL}, and D is Henry’s law constant of CO₂ in ionic liquid, the inverse of selectivity of CO₂ over H₂, desorption capacity obtained from COSMO-SAC, respectively. For COSMO-SAC simulations, ionic liquids consist of the newly obtained cation and bis triflimide anion (TFSI⁻). The complete list of top structures is given as the Supporting Information.

No.	Cation Structure	SMILES	R (bar ⁻¹)	ASDI (bar)	H _{i/IL} (bar)	1/S _{i,j} (10 ⁻³)	D
1		CCCc1n(CCC)cc[n+]1CCOc1cccc1	1.553	0.644	329.4	3.057	0.639
2		CCCCc1n(CCOC)cc[n+]1CCCC	1.545	0.647	320.1	3.167	0.638
3		CCCn1cc[n+]1CCC)c1CC	1.542	0.649	327.2	3.113	0.637
4		CCCCn1cc[n+]1(CCOC)c1CCC	1.536	0.651	325.3	3.133	0.637
5		CCCC[n+]1ccn(CCC)c1CCC	1.531	0.653	320.7	3.128	0.640
6		CCCc1n(CCOC)cc[n+]1CCC	1.527	0.655	330.4	3.119	0.635
7		CCCc1n(CCC)cc[n+]1CCC	1.519	0.658	325.0	3.166	0.640
8		CCCCc1n(CCOC)cc[n+]1CCOC	1.515	0.660	339.4	3.052	0.637
9		CCCCc1n(CCC)cc[n+]1CCCC	1.515	0.660	317.1	3.242	0.642
10		CCCC[n+]1ccn(CCOC)c1Cc1cccc1	1.511	0.662	333.4	3.111	0.638
Ref.		CCN1C = C[N+]1(=C1)C	0.733	1.365	569.4	3.680	0.649

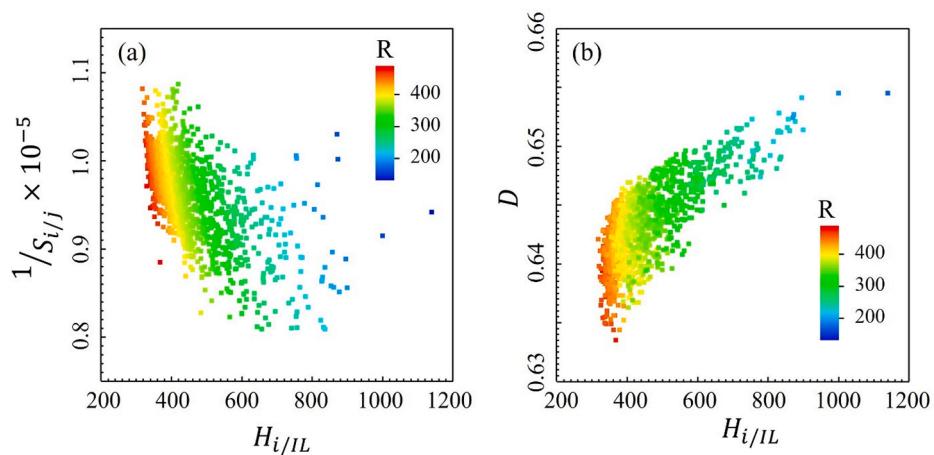


Fig. 2. (a) The inverse of the solvent selectivity toward target gas component *i* (=CO₂) over other undesirable components *j* (=N₂) (1/S_{i,j}) and (b) desorption capacity (D) as a function of Henry’s law constant of CO₂ in IL (H_{i/IL}). Color coding of the squares corresponds to the reward function value R (=1/ASDI) in MP-MCTS. These graphs are for a total of 1454 ionic liquids newly designed using our algorithm for CO₂ separation from flue gas (CO₂/N₂).

our algorithm to achieve a goal set out in terms of performance property, yet it is also important that an algorithm explores materials space broadly and designs top ILs eventually with good diversity rather than

produce only similar ILs repeatedly. To check this, we assessed the diversity of the generated ILs using a topological data analysis (TDA) [28,46]. For TDA, a similarity between ionic liquids was determined by

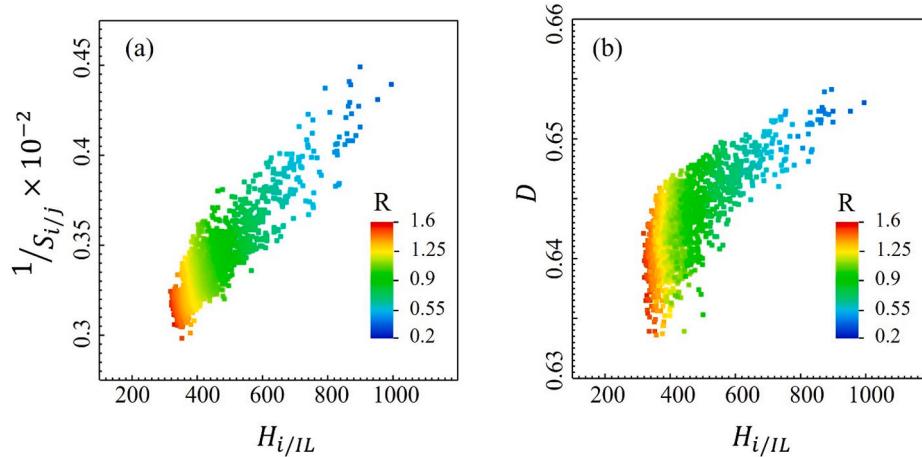


Fig. 3. (a) The inverse of the solvent selectivity toward target gas component i ($=\text{CO}_2$) over other undesirable components j ($=\text{H}_2$) ($1/S_{i/j}$) and (b) desorption capacity (D) as a function of Henry's law constant of CO_2 in IL ($H_{i/IL}$). Color coding of the squares corresponds to the reward function value R ($=1/\text{ASDI}$) in MP-MCTS. These graphs are for a total of 1779 ionic liquids newly designed using our algorithm for the separation of CO_2 from syngas (CO_2/H_2).

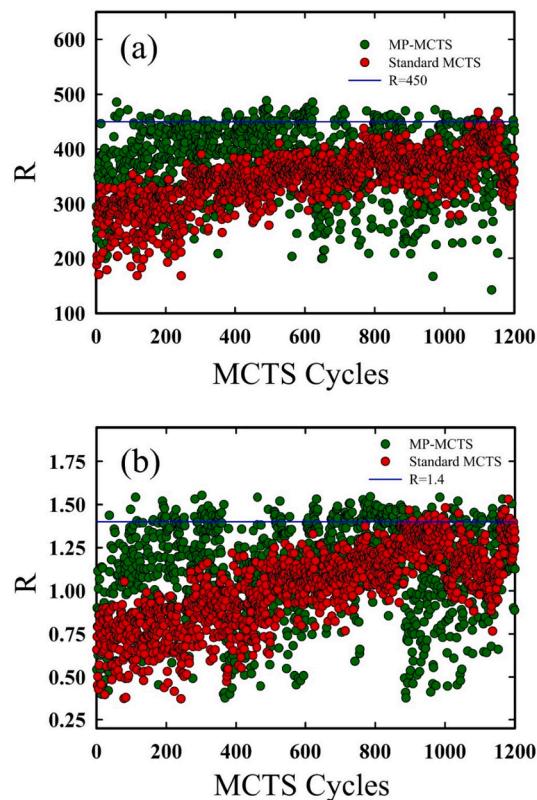


Fig. 4. The evolution of the reward function value R with MCTS cycles in designing ionic liquids (ILs) for CO_2 separation from (a) flue gas (CO_2/N_2) and (b) syngas (CO_2/H_2). The green or red colored circles correspond to MP-MCTS (i.e., design of three branches together) or standard MCTS (i.e., design of side branch one by one), respectively. The blue-colored horizontal solid line corresponds to $R = 450$ or 1.4 that is the criterion for promising ILs in this analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

comparing branches in cations with noting that the imidazolium part of cation and anion are the same for each case. To measure similarity, a side branch was represented using a topological fingerprint [32] which is a bit vector generated based on the molecular fragments. A distance (L) between different fingerprints can be estimated by utilizing the Tanimoto coefficient [47] as follows

$$L = 1 - \text{Tanimoto coefficient} = 1 - \frac{c}{a + b - c} \quad (3)$$

where a (or b) represents bits set for each fingerprint, and c is bits set in common in the two fingerprints. Considering that each cation has three branches and reflection symmetry between R_1 and R_2 , a pair-wise distance between two cations A and B was determined by averaging L for 6 pairs of branches (before reflection : $R_{1,A} - R_{1,B}, R_{2,A} - R_{2,B}, R_{3,A} - R_{3,B}$, and after reflection: $R_{1,A} - R_{3,B}, R_{2,A} - R_{2,B}, R_{3,A} - R_{1,B}$) as $\bar{L}_{AB} = (L_{R_{1,A}-R_{1,B}} + L_{R_{3,A}-R_{3,B}} + 2 \cdot L_{R_{2,A}-R_{2,B}} + L_{R_{1,A}-R_{3,B}} + L_{R_{3,A}-R_{1,B}})/6$, where $R_{i,j}$ denotes i^{th} branch in cation j . Then, a distance matrix was given as an input for generating mapper plots using TDA. We performed this analysis for the case which generated 2179 (or 2460) cations obtained for CO_2/N_2 (or CO_2/H_2) separation application, respectively. In Fig. 5, the mapper plots for two cases are shown. In a plot each node contains a set of similar structures and nodes connected by lines share some structural aspects in common. Two nodes apart without line connection indicate that structures in those nodes are dissimilar. In Fig. 5, nodes are colored according to the normalized order of the IL generation cycle for each run. The blue (or red) color represents an early (or final) cycle for the IL generation. In both mapper plots, the distribution of the same colored nodes is broad rather than adjacent, which would indicate that our algorithm in each run designs ILs with a different strategy of optimizing reward function. We observed sudden and discontinuous changes in coloring coding among neighboring nodes. This result would support that at some attempts (i.e. cycles) our algorithm made complex replacements or even generation of completely different branches, along with small modifications in the underlying structures as typically expected. Furthermore, from focused and yet dispersed locations of top-performing ILs on the mapper plots in the insets of Fig. 5, it is recognized that our algorithm can find structural diversity of top ILs rather than design analogous ones repeatedly.

4. Conclusion

In this work, we have developed a computational approach to perform the tailor-made design of top-performing ionic liquid (IL) for target applications. For the case studies of CO_2 separation from flue gas (CO_2/N_2) and the separation of CO_2 from syngas (CO_2/H_2), it is demonstrated that high-performance ILs can be designed with great efficiency using our algorithm (named as MP-MCTS) utilizing the multiplayer Monte Carlo Tree Search combined with the recurrent neural network. Furthermore, we performed topological data analysis (TDA) on newly designed ILs in materials space. Mapper plots from TDA indicated

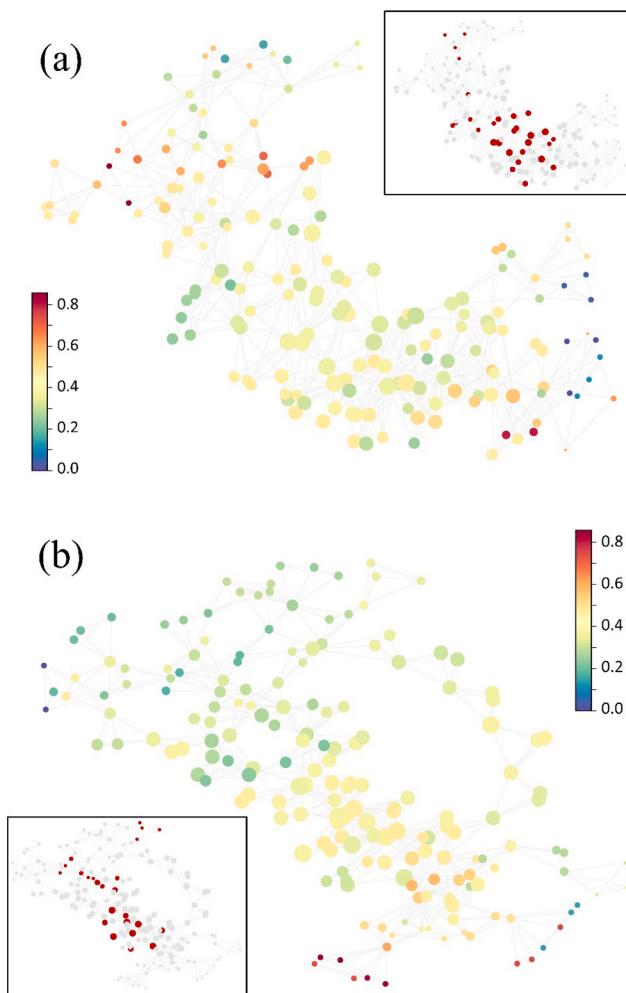


Fig. 5. Mapper plots of ionic liquids(ILs) generated for CO₂ separation from (a) flue gas (CO₂/N₂) and (b) syngas (CO₂/H₂). Nodes in the network represent clusters of similar structures, and gray edges connect nodes that contain structures in common. These figures were obtained with the Kepler Mapper (<https://kepler-mapper.scikit-tda.org/>). Nodes have different colors according to the normalized order of ILs generation during each run (Red: high value, Blue: low value). Normalization is based on the total number of ILs generated in each run. Insets show the distribution of top-performing ILs on the plots. (Red: presence, Gray: absence). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that our algorithm allows us not only to explore materials space widely rather than generate only similar structures repeatedly but also to successfully find high-performing ILs with a good diversity by often making a singular switch in branch design.

Although in this work we applied our algorithm for designing side branches of an imidazolium-based cation, it is expected that our approach based on multi-player strategies would be capable of discovering the promising cations with multiple functional groups for other cases. Furthermore, besides the two cases of CO₂ capture studied in this work, our approach can easily be extended to other applications of IL by properly choosing a performance criterion and a computer simulation tool to estimate it. However, it should be noted that our current algorithm does not take into account whether computationally designed structures are synthesizable and how to synthesize while those are important questions needed to be addressed. Although our work is currently focused on seeking optimum structures solely in terms of performance properties considering that directing a computational algorithm to design molecules for an increase in synthesizability would come at the cost of performance properties to the desired application

[48], this limitation of our algorithm is expected to be improved in our future works.

CRediT authorship contribution statement

Kexin Zhang: Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation. **Jiasheng Wu:** Methodology, Software, Validation, Investigation, Writing - original draft. **Hyeonsuk Yoo:** Visualization, Validation, Investigation. **Yongjin Lee:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgment

This research was supported by the National Natural Science Foundation of China (Grant No. 21950410525), the ShanghaiTech University Research Startup Fund, and Inha University Research Grant. We thank the high-performance computing platform of ShanghaiTech University and the Shanghai Supercomputer Center for the use of their computing resources. This work was also supported by the National Supercomputing Center with supercomputing resources including technical support (KSC-2020-CRE-0339).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.seppur.2021.119117>.

References

- N.V. Plechkova, K.R. Seddon, Applications of Ionic Liquids in the Chemical Industry, *Chem. Soc. Rev.* 37 (1) (2008) 123–150, <https://doi.org/10.1039/b006677j>.
- J. Dupont, From Molten Salts to Ionic Liquids: A “Nano” Journey, *Acc. Chem. Res.* 44 (11) (2011) 1223–1231, <https://doi.org/10.1021/ar2000937>.
- J. Hulsbosch, D.E. De Vos, K. Binnemans, R. Ameloot, Biobased Ionic Liquids: Solvents for a Green Processing Industry? *ACS Sustain. Chem. Eng.* 4 (6) (2016) 2917–2931, <https://doi.org/10.1021/acssuschemeng.6b00553>.
- M. Armand, F. Endres, D.R. MacFarlane, H. Ohno, B. Scrosati, Ionic-Liquid Materials for the Electrochemical Challenges of the Future, *Nat. Mater.* 8 (8) (2009) 621–629, <https://doi.org/10.1038/nmat2448>.
- M.A.R. Martins, U. Domańska, B. Schröder, J.A.P. Coutinho, S.P. Pinho, Selection of Ionic Liquids to Be Used as Separation Agents for Terpenes and Terpenoids, *ACS Sustain. Chem. Eng.* 4 (2) (2016) 548–556, <https://doi.org/10.1021/acssuschemeng.5b01357>.
- L.F. Lepre, D. Andre, S. Denis-Quanquin, A. Gautier, A.A.H. Pádua, M. Costa Gomes, Ionic Liquids Can Enable the Recycling of Fluorinated Greenhouse Gases. *ACS Sustain. Chem. Eng.* 7 (19) (2019), <https://doi.org/10.1021/acssuschemeng.9b04214>.
- L.A. Blanchard, D. Hancu, E.J. Beckman, J.F. Brennecke, Green Processing Using Ionic Liquids and CO₂, *Nature* 398 (6731) (1999) 28–29, <https://doi.org/10.1038/19887>.
- M. Ramdin, T.W. de Loos, T.J.H. Vlugt, State-of-the-Art of CO₂ Capture with Ionic Liquids, *Ind. Eng. Chem. Res.* 51 (24) (2012) 8149–8177, <https://doi.org/10.1021/ie3003705>.
- S. Zeng, X. Zhang, L. Bai, X. Zhang, H. Wang, J. Wang, D. Bao, M. Li, X. Liu, S. Zhang, Ionic-Liquid-Based CO₂ Capture Systems: Structure, Interaction and Process. *Chem. Rev.* 117 (14) (2017) 9625–9673, <https://doi.org/10.1021/acscchemrev.7b00072>.
- J. Wang, Z. Song, H. Cheng, L. Chen, L. Deng, Z. Qi, Computer-Aided Design of Ionic Liquids as Absorbent for Gas Separation Exemplified by CO₂ Capture Cases, *ACS Sustain. Chem. Eng.* 6 (9) (2018) 12025–12035, <https://doi.org/10.1021/acssuschemeng.8b02321>.
- J.F. Brennecke, B.E. Gurkan, Ionic Liquids for CO₂ Capture and Emission Reduction, *J. Phys. Chem. Lett.* 1 (24) (2010) 3459–3464, <https://doi.org/10.1021/jz1014828>.
- J.D. Holbrey, K.R. Seddon, Ionic Liquids, *Clean Technol. Environ. Policy* 1 (4) (1999) 223–236, <https://doi.org/10.1007/s100980050036>.
- F. Yan, M. Larney, K. Jariwala, S. Bowser, K. Damodaran, E. Albenze, D.R. Luebke, H.B. Nulwala, B. Smit, M. Haranczyk, Toward a Materials Genome Approach for

- Ionic Liquids: Synthesis Guided by Ab Initio Property Maps, *J. Phys. Chem. B* 118 (47) (2014) 13609–13620, <https://doi.org/10.1021/jp506972w>.
- [14] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering, *Science* (80-) 361 (6400) (2018) 360–365, <https://doi.org/10.1126/science.aat2663>.
- [15] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the Game of Go with Deep Neural Networks and Tree Search, *Nature* 529 (7587) (2016) 484–489, <https://doi.org/10.1038/nature16961>.
- [16] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature* 521 (7553) (2015) 436–444, <https://doi.org/10.1038/nature14539>.
- [17] M. Dieb, S. Ju, K. Yoshizoe, Z. Hou, J. Shiomi, K. Tsuda, MDTs, Automatic Complex Materials Design Using Monte Carlo Tree Search, *Sci. Technol. Adv. Mater.* 18 (1) (2017) 498–503, <https://doi.org/10.1080/14686996.2017.1344083>.
- [18] M.H.S. Segler, T. Kogej, C. Tyrchan, M.P. Waller, Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks, *ACS Cent. Sci.* 4 (1) (2018) 120–131, <https://doi.org/10.1021/acscentsci.7b00512>.
- [19] X. Yang, J. Zhang, K. Yoshizoe, K. Terayama, K. Tsuda, ChemTS: An Efficient Python Library for de Novo Molecular Generation, *Sci. Technol. Adv. Mater.* 18 (1) (2017) 972–976, <https://doi.org/10.1080/14686996.2017.1401424>.
- [20] M. Sumita, X. Yang, S. Ishihara, R. Tamura, K. Tsuda, Hunting for Organic Molecules with Artificial Intelligence: Molecules Optimized for Desired Excitation Energies, *ACS Cent. Sci.* 4 (9) (2018) 1126–1133, <https://doi.org/10.1021/acscentsci.8b00213>.
- [21] Y. Bao, R.L. Martin, M. Haranczyk, M.W. Deem, In Silico Prediction of MOFs with High Deliverable Capacity or Internal Surface Area, *Phys. Chem. Chem. Phys.* 17 (18) (2015) 11962–11973, <https://doi.org/10.1039/c5cp00002e>.
- [22] Y. Bao, R.L. Martin, C.M. Simon, M. Haranczyk, B. Smit, M.W. Deem, In Silico Discovery of High Deliverable Capacity Metal-Organic Frameworks, *J. Phys. Chem. C* 119 (1) (2015) 186–195, <https://doi.org/10.1021/jp5123486>.
- [23] S.P. Collins, T.D. Daff, S.S. Piotrkowski, T.K. Woo, Materials Design by Evolutionary Optimization of Functional Groups in Metal-Organic Frameworks, *Sci. Adv.* 2 (11) (2016) 1–7, <https://doi.org/10.1126/sciadv.1600954>.
- [24] X. Zhang, K. Zhang, Y. Lee, Machine Learning Enabled Tailor-Made Design of Application-Specific Metal-Organic Frameworks, *ACS Appl. Mater. Interfaces* 12 (1) (2020) 734–743, <https://doi.org/10.1021/acsmami.9b17867>.
- [25] X. Zhang, K. Zhang, H. Yoo, Y. Lee, Machine Learning-Driven Discovery of Metal-Organic Frameworks for Efficient CO₂ Capture in Humid Condition, *ACS Sustain. Chem. Eng.* 9 (7) (2021) 2872–2879, <https://doi.org/10.1021/acssuschemeng.0c08806>.
- [26] J.A.M. Nijssen, M.H.M. Winands, Enhancements for Multi-Player Monte-Carlo Tree Search, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer, Berlin, Heidelberg, 2011, 6515 LNCS, pp 238–249, https://doi.org/10.1007/978-3-642-17928-0_22.
- [27] S.A. Mirsoleimani, A. Plaat, J. van den Herik, J. Vermaseren, Structured Parallel Programming for Monte Carlo Tree Search. arXiv 2017.
- [28] G. Carlsson, Topology and data, *Bull. New. Ser. Am. Math. Soc.* 46 (2009) 255–308.
- [29] D. Weininger, SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules, *J. Chem. Inf. Model.* 28 (1) (1988) 31–36, <https://doi.org/10.1021/ci00057a005>.
- [30] C.B. Browne, E. Powley, D. Whitehouse, S.M. Lucas, P.I. Cowling, P. Rohlfschagen, S. Tavener, D. Perez, S. Samothrakis, S. Colton, A Survey of Monte Carlo Tree Search Methods, *IEEE Trans. Comput. Intell. AI Games* 4 (1) (2012) 1–43, <https://doi.org/10.1109/TCIAIG.2012.2186810>.
- [31] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: A Free Tool to Discover Chemistry for Biology, *J. Chem. Inf. Model.* 52 (7) (2012) 1757–1768, <https://doi.org/10.1021/ci3001277>.
- [32] Landrum, G. RDKit: Open-source cheminformatics <http://rdkit.org/>.
- [33] G.M. Chaslot, J.B. Winands, M.H.M. Van Den, H.J. Herik, Parallel Monte-Carlo Tree Search, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer, Berlin, Heidelberg, 2008, 5131 LNCS, pp 60–71, https://doi.org/10.1007/978-3-540-87608-3_6.
- [34] M.C. Corvo, J. Sardinha, T. Casimiro, G. Marin, M. Seferin, S. Einloft, S.C. Menezes, J. Dupont, E.J. Cabrita, A Rational Approach to CO₂ Capture by Imidazolium Ionic Liquids: Tuning CO₂ Solubility by Cation Alkyl Branching, *ChemSusChem* 8 (11) (2015) 1935–1946, <https://doi.org/10.1002/cssc.201500104>.
- [35] J. Ranke, S. Stolte, R. Störmann, J. Aming, B. Jastorff, Design of Sustainable Chemical Products - The Example of Ionic Liquids, *Chem. Rev.* 107 (6) (2007) 2183–2206, <https://doi.org/10.1021/cr050942s>.
- [36] Y. Zhao, R. Gani, R.M. Afzal, X. Zhang, S. Zhang, Ionic Liquids for Absorption and Separation of Gases: An Extensive Database and a Systematic Screening Method, *AIChE J.* 63 (4) (2017) 1353–1367, <https://doi.org/10.1002/aic.15618>.
- [37] D. Valencia-Marquez, A. Flores-Tlacuahuac, R. Vasquez-Medrano, An Optimization Approach for CO₂ Capture Using Ionic Liquids, *J. Clean. Prod.* 168 (2017) 1652–1667, <https://doi.org/10.1016/j.jclepro.2016.11.064>.
- [38] S.T. Lin, S.I. Sandler, A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model, *Ind. Eng. Chem. Res.* 41 (5) (2002) 899–913, <https://doi.org/10.1021/ie001047w>.
- [39] S. Wang, S.I. Sandler, C.C. Chen, Refinement of COSMO-SAC and the Applications, *Ind. Eng. Chem. Res.* 46 (22) (2007) 7275–7288, <https://doi.org/10.1021/ie070465z>.
- [40] I.H. Bell, E. Mickoleit, C.M. Hsieh, S.T. Lin, J. Vrabec, C. Breitkopf, A. Jäger, A Benchmark Open-Source Implementation of COSMO-SAC, *J. Chem. Theory Comput.* 16 (4) (2020) 2635–2646, <https://doi.org/10.1021/acs.jctc.9b01016>.
- [41] 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. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, D.J.F. Gaussian, 09 Revision A.02., Gaussian Inc., Wallingford CT, 2016.
- [42] S. Tang, G.A. Baker, H. Zhao, Ether- and Alcohol-Functionalized Task-Specific Ionic Liquids: Attractive Properties and Applications, *Chem. Soc. Rev.* 41 (10) (2012) 4030–4066, <https://doi.org/10.1039/c2cs15362a>.
- [43] Y. Zhou, X. Xu, Z. Wang, S. Gong, H. Chen, Z. Yu, J. Kiefer, The Effect of Introducing an Ether Group into an Imidazolium-Based Ionic Liquid in Binary Mixtures with DMSO, *Phys. Chem. Chem. Phys.* 22 (27) (2020) 15734–15742, <https://doi.org/10.1039/d0cp01568g>.
- [44] I. Mejía, K. Stanley, R. Canales, J.F. Brennecke, On the High-Pressure Solubilities of Carbon Dioxide in Several Ionic Liquids, *J. Chem. Eng. Data* 58 (9) (2013) 2642–2653, <https://doi.org/10.1021/je400542b>.
- [45] J.L. Anthony, J.L. Anderson, E.J. Maginn, J.F. Brennecke, Anion Effects on Gas Solubility in Ionic Liquids, *J. Phys. Chem. B* 109 (13) (2005) 6366–6374, <https://doi.org/10.1021/jp046404l>.
- [46] P.Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, G. Carlsson, Extracting Insights from the Shape of Complex Data Using Topology, *Sci. Rep.* 3 (1) (2013) 1236, <https://doi.org/10.1038/srep01236>.
- [47] Segaran, T. Programming Collective Intelligence: Building Smart Web 2.0 Applications, O'Reilly Media, Sebastopol, CA, 2007.
- [48] W. Gao, C.W. Coley, The Synthesizability of Molecules Proposed by Generative Models, *J. Chem. Inf. Model.* 60 (12) (2020) 5714–5723, <https://doi.org/10.1021/acs.jcim.0c00174>.